E-ISSN : 2581-0588
P-ISSN : 2301-7988

INSTITUT SAINS DAN BISNIS
ATMA LUHUR

JURNAL SISFOKOM

(Sistem Informasi dan Komputer)

VOLUME 14 - NO. 02 - MEY 2025

P-ISSN: 2301-7988
E-ISSN: 2581-0588

JURNAL SISFOKOM

(SISTEM INFORMASI DAN KOMPUTER)

Jurnal Sisfokom, an acronym for Journal of Information Systems and Computers, is a scholarly publication resulting from a collaborative effort between the academic community of ISB Atma Luhur and various higher education institutions across Indonesia. This journal serves as a vital platform for disseminating scientific articles from researchers, academics, and practitioners in the field of information technology. With a specific focus on information systems and computer science, Jurnal Sisfokom consistently presents high-quality papers, published four times a year in January, May, July, and November, ensuring a continuous and relevant flow of knowledge.

EDITORIAL BOARD

P-ISSN: 2301-7988
E-ISSN: 2581-0588

P-ISSN: 2301-7988
E-ISSN: 2581-0588

SECRETARIAT

LPPM Institut Sains dan Bisnis Atma Luhur
Jl. Jend. Sudirman, Selindung Baru, Pangkalpinang
Kepulauan Bangka Belitung - Indonesia
Telp. (0717) 433 506 Fax. (0717) 433 506
e-mail : sisfokom@atmaluhur.ac.id - lppm@atmaluhur.ac.id

P-ISSN: 2301-7988
E-ISSN: 2581-0588

EDITOR’S FOREWORD

Jurnal Sisfokom (Information Systems and Computers) stands as a distinguished scholarly publication, meticulously managed and proudly issued by the LPPM ISB Atma Luhur Pangkalpinang.

This particular edition, Volume 14 Number 02 – Mey 2025, exemplifies a robust collaborative endeavor, uniting the academic prowess of ISB Atma Luhur with a diverse array of esteemed universities throughout Indonesia.

The editorial board extends its profound gratitude for the invaluable participation and unwavering cooperation of our dedicated lecturers. Their significant contributions have been pivotal, enabling the timely and successful publication of Jurnal Sisfokom (Information Systems and Computers) Volume 14 Number 02 – Mey 2025, precisely in accordance with our meticulously laid plans.

Additionally, the editorial board extends its profound appreciation to the distinguished experts, both internal and external to ISB Atma Luhur, whose invaluable contributions in thoroughly assessing and meticulously refining the submitted manuscripts have been absolutely critical to the quality of this publication.

On this occasion, the editorial board cordially invites and extends the widest possible opportunity to all researchers, fellow lecturers, and discerning scholars/practitioners in both Information Systems and Informatics Engineering to contribute and publish their research findings through this esteemed journal.

Ultimately, the editorial board sincerely hopes that the scholarly articles published within this journal will yield substantial benefits for the entire academic community at ISB Atma Luhur, and profoundly contribute to the broader advancement of science and information technology.

Editorial,

P-ISSN: 2301-7988
E-ISSN: 2581-0588

TABLE OF CONTENTS

P-ISSN: 2301-7988
E-ISSN: 2581-0588

P-ISSN: 2301-7988
E-ISSN: 2581-0588

SIMPELMAS: an Integrated Information System for Research and Community Service using a Prototype Development Approach

Syafi'ul Muzid[1]*, Alif Catur Murti[2], Wiwit Agus Triyanto[3]
Department of Information System, Faculty of Engineering [1], [3]
Department of Informatics Engineering, Faculty of Engineering [2]
Universitas Muria Kudus
Kudus, Indoensia
syafiul.muzid@umk.ac.id [1], alif.catur@umk.ac.id [2] at.wiwit@umk.ac.id [3]

***Abstract*— The Institute for Research and Community Service (LPPM) plays a strategic role in coordinating academic research and community engagement activities. However, fragmented data management continues to hinder performance evaluations and strategic decision-making in many universities. This study aims to develop SIMPELMAS (Research and Community Service Management Information System), an integrated platform designed to streamline the management of human resources, research, and community service data. Using a prototype-based development methodology, SIMPELMAS was implemented and tested in Universitas Muria Kudus. The system achieved high success rates in various aspects: over 95% in functionality, 99–100% in security, and 80–85% in user satisfaction. Key features such as proposal submission, fund monitoring, and final reporting functioned optimally. Integration testing confirmed effective data synchronization, while user feedback highlighted the need for improvements in user experience, particularly for students and new users. This study contributes to the digital transformation of higher education by providing a replicable model for academic information systems that support real-time monitoring, transparency, and data-driven governance. While the system has met key eligibility standards, further enhancements in user interface and mobile responsiveness are recommended to ensure broader usability and adoption.**

Keywords—Management Information System, Research, Community Service, Data Integration, Higher Education, Digital Transformation

I. INTRODUCTION

The Research and Community Service Institute (LPPM) plays a pivotal role in coordinating and evaluating research and community engagement in higher education institutions. In today's era of digital transformation, effective and integrated information systems are crucial for managing complex academic data, ensuring transparency, and improving institutional performance. Despite the growing need for data-driven governance, many universities continue to manage their research, human resource, and community service data separately. This fragmented management approach hinders efficient performance monitoring and informed decision-making. Recent reports from the Ministry of Education (2023) indicate that more than 65% of Indonesian universities lack integrated systems for academic data governance, resulting in reporting delays and underutilized research outputs. Previous studies have explored the role of information systems in higher education [1] [2], but limited attention has been given to systems that simultaneously integrate human resource, research, and community service data. This presents a critical research gap, especially in the context of performance-based evaluations and output-oriented funding models increasingly adopted by academic institutions [3] [4].To address this gap, this study focuses on the development of SIMPELMAS (Sistem Informasi Manajemen Penelitian dan Pengabdian kepada Masyarakat), an integrated information system aimed at streamlining academic data management and supporting strategic planning. SIMPELMAS was developed using a prototype methodology that enables iterative improvements through user feedback and testing. The novelty of this research lies in its integrated design, real-time monitoring features, and comprehensive approach to data centralization, which contrasts with most existing systems that focus solely on administrative or siloed operations. By offering a holistic solution for managing academic performance, the proposed system aims to support universities in achieving better governance, improved service delivery, and stronger contributions to society and industry. This paper is structured as follows: Section 2 describes the methodology used in system development. Section 3 presents the results of system implementation and testing. Section 4 provides a detailed discussion of the findings and their implications. Section 5 concludes the study and offers recommendations for future research.

The existence of an integrated management information system is a pressing need to ensure accuracy in reporting and academic performance evaluation. The Research and Community Service Management Information System (SIMPELMAS) is expected to serve as a solution for enhancing efficiency in research and community service data management, as well as improving HR quality in supporting outcome-based education [5]. With a comprehensive system in place, outputs such as scientific publications, books, teaching materials, and community service results can be systematically

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2339, Copyright ©2025
Submitted : April 25, 2025, Revised : May 10, 2025, Accepted : May 15, 2025, Published : May 26, 2025

documented, supporting the achievement of the Tri Dharma of Higher Education [6]. As challenges in managing fragmented data increase, many universities in Indonesia have started implementing integrated information systems; however, these remain limited to administrative applications without supporting in-depth data-dr[7]iven evaluation. Recent studies indicate that an effective management information system can expedite evaluation processes, enhance accountability, and optimize the use of human resources and research [8][5]. Artificial intelligence and data analytics technologies are also beginning to be utilized to improve transparency and accuracy in academic data management [9][10]. However, the implementation of these technologies within the context of Indonesian universities, particularly in integrating HR and research data, remains limited. Therefore, this research contributes significantly to developing an integrated system that can improve academic data and community service management.

The implementation of an integrated management information system also enables academic stakeholders to monitor research and community service progress more effectively [11]. With integration, faculty members can easily access and report their research findings, while universities can more accurately map academic achievements [12]. This system also has the potential to enhance administrative efficiency and reduce redundancies in documentation processes [7]. Beyond administrative aspects, an integrated system supports data-driven research strategies, enabling universities to identify research trends, optimize resources, and strengthen collaborations between academia, industry, and government [13]. Digitalization in research and community service management is also believed to improve institutional transparency and accountability, as well as support the attainment of national and international accreditation standards [14]. As the need for accurate and real-time data in higher education management increases, the use of artificial intelligence and data analytics technology is becoming a growing trend in academic information systems [9]. These technologies allow for deeper analysis of faculty performance and the effectiveness of research and community service activities [10]. Consequently, data-driven decision-making can be more accurate and strategic in supporting the improvement of higher education quality [15].

There is a significant gap between the need for an integrated information system and the reality of its implementation in many universities. Although many data-driven information systems are beginning to be adopted, most remain focused on administrative data management and do not support strategic data-driven decision-making. Furthermore, many systems have yet to fully leverage the potential of analytics and artificial intelligence in supporting research and community service performance evaluation. This study focuses on developing a more comprehensive system that integrates HR data, research, and community service into a single platform to provide a more accurate and holistic overview while supporting the achievement of the Tri Dharma of Higher Education. Given the various benefits offered, this study aims to develop and implement the integration of HR data with research and community service data through SIMPELMAS. This approach is expected to enhance the quality and effectiveness of faculty performance management and strengthen academic contributions to society and industry [7]. Additionally, this research will evaluate the impact of system integration on the effectiveness of reporting and decision-making in higher education institutions [16].

II. Method

This research employed a prototype-based system development methodology to design and implement the SIMPELMAS platform. The methodological framework comprises three main phases: data collection, system development, and validation. 1. Data Collection The data collection process was conducted using both primary and secondary sources: - Primary data were obtained through direct observation of LPPM operations, interviews with faculty members, and usability testing involving 10 lecturers and 5 administrative staff. These interactions provided insights into the user requirements, operational workflows, and system expectations. - Secondary data included institutional documents, academic reports, and relevant literature. A review of recent studies (published within the last five years) helped situate the system within current academic discourse and informed best practices in system design. 2. System Development Approach The development followed the Prototype model, which supports iterative design and feedback-driven improvement. The process included:- Requirement Analysis: Identification of functional and non-functional requirements through stakeholder meetings and process mapping.- Rapid Design: Creation of system architecture and initial user interface mock-ups using UML diagrams (Use Case, Activity, and Class Diagrams).- Prototype Construction: Initial version of SIMPELMAS was built incorporating key features like proposal submission, progress tracking, and member verification.- User Evaluation: Users tested the prototype in a controlled environment, focusing on functionality, usability, and responsiveness. Feedback was systematically recorded. - Refinement and Iteration: Based on feedback, interface improvements and functional adjustments were made to optimize user experience. - Final Implementation: The refined prototype was deployed in the production environment of Universitas Muria Kudus. 3. Validation and Testing To ensure reliability and performance, several types of testing were conducted:- Functionality Testing: Assessed whether each feature operated as intended.- Integration Testing: Verified data flow between modules.- Security Testing: Checked data encryption, authentication mechanisms, and SQL injection vulnerability.- Responsiveness Testing: Ensured system compatibility with multiple devices and network conditions.- User Acceptance Testing: Measured satisfaction and usability, with success rates recorded at 85% for lecturers and 80% for students. This comprehensive methodology ensures that the developed system not only aligns with the operational context of LPPM but also reflects academic standards for integrated digital platforms in higher education.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2339, Copyright ©2025
Submitted : April 25, 2025, Revised : May 10, 2025, Accepted : May 15, 2025, Published : May 26, 2025

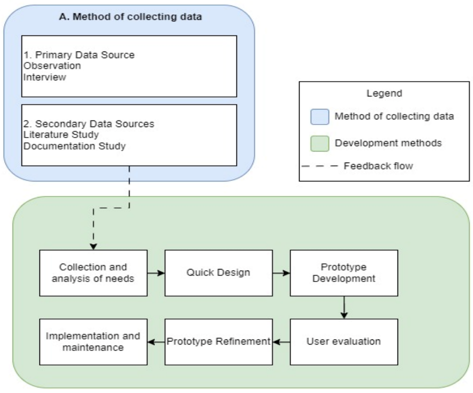


Fig. 1. Research framework

A. Data Collection Methods

The data collection approach in this study was designed hierarchically with two categories of sources:

A.1 Primary Data Sources

Primary data is obtained through direct interaction with the phenomenon under study, ensuring high validity and credibility for the research [17]. The methods for acquiring primary data include:

Observation: Direct observation of business processes at Universitas Muria Kudus, particularly regarding thesis submission and thesis outcomes within the study program. According to [18], this approach enables researchers to document phenomena objectively and gain a comprehensive understanding of the operational context.

Interviews: Data collection through dialogical interactions with relevant stakeholders allows for an in-depth exploration of perspectives, needs, and challenges faced in the existing system [19].

A.2 Secondary Data Sources

Secondary data complements and enriches primary findings, providing a theoretical and referential foundation for the research [20]. The collection methods include:

Literature Review: Exploration of scientific literature related to system analysis and design, guidance systems, and thesis outcomes. As stated by [21], this approach strengthens the conceptual foundation of the research and positions it within a broader scientific context.

Document Study: A systematic examination of documents from various sources, including online materials, textbooks, and other literature, to enhance analytical and comparative perspectives [22].

B. System Development Method (Prototype)

The implementation of the Prototype methodology follows an iterative-incremental approach in system development, integrating six interrelated stages [23]:

B.1 Requirement Collection and Analysis

The initiation phase involves a comprehensive identification of user requirements through collaborative dialogues between developers and stakeholders. As explained by [24], this process aims to articulate functional and non-functional expectations in detail, ensuring that the developed system aligns with the organizational needs of LPPM.

B.2 Rapid Design

Based on the requirement analysis, developers construct a conceptual design framework encompassing system architecture, process flows, and user interfaces. According to [25], although preliminary, this design serves as an essential blueprint for the development of the initial prototype.

B.3 Prototype Development

The conceptual design is implemented into a functional prototype that materializes the system's core features. As noted by [26], this prototype serves as a tangible representation of the proposed concept and forms the basis for user evaluation.

B.4 User Evaluation

This critical phase involves testing the prototype with potential users within relevant operational contexts. [27] highlights that the feedback collected covers aspects of functionality, usability, and the system's effectiveness in addressing identified needs.

B.5 Prototype Refinement

Based on insights from user evaluations, the prototype undergoes modifications and refinements through an iterative process. [28] suggests that this refinement may involve feature revisions, interface improvements, or workflow restructuring to optimize user experience and system functionality.

B.6 Implementation and Maintenance

The culmination of the development process sees the final prototype transformed into an operational system deployed in a production environment. According to [29], this phase also includes continuous maintenance strategies to ensure the system's sustainability and adaptability to institutional changes.

C. Methodological Integration and Information Flow

The schematic representation illustrates a bidirectional flow of information, where outputs from the data collection methods serve as fundamental inputs for the system development process. Findings from observations, interviews, and literature reviews inform each stage of development, from requirement analysis to prototype evaluation [25]. The iterative nature of the Prototype methodology, represented by feedback loops, underscores an adaptive development philosophy wherein the system undergoes continuous refinement based on empirical insights. As emphasized by [30], this approach ensures that the final system not only meets technical specifications but is also aligned with the operational context and user expectations at LPPM Universitas Muria Kudus. This comprehensive methodology represents a holistic approach that integrates academic rigor with operational pragmatism, ensuring the development of a system that is both responsive to the specific needs of the institution and grounded in scientific principles [18].

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2339, Copyright ©2025
Submitted : April 25, 2025, Revised : May 10, 2025, Accepted : May 15, 2025, Published : May 26, 2025

III. Results And Discussion

The SIMPELMAS platform was successfully developed and implemented at Universitas Muria Kudus as an integrated system that centralizes research and community service data. This section presents the key findings from system testing and discusses their implications.

1. System Features and User Interface SIMPELMAS includes a dashboard with intuitive navigation, modules for proposal submission, progress tracking, and final reporting. Key features like membership verification and fund monitoring were also integrated. Figures 2 to 5 illustrate these functionalities. User feedback indicated that the system's layout was easy to understand, though performance on mobile devices with poor internet connection could be improved.

2. Functionality and Integration Testing As shown in Table 1 to Table 4, system components performed with high success rates: - Research module: average success rate of 97.2%. - Community service module: success rate of 96.7%. - General functions (login, file upload/download): 97.6%. - System integration: 98% effectiveness in data flow and feature synchronization.

3. Security and Responsiveness Security tests demonstrated high reliability: - Login and authentication: 99%. - SQL injection protection: 98%. - Data encryption and session timeout: 100% and 97%, respectively. Responsive testing across desktop, tablet, and mobile platforms averaged 98% success. However, the system's speed dropped by 4–5% under slow connections, requiring further optimization.

4. User Acceptance and Usability User satisfaction varied across roles:- Lecturers: 85% satisfaction. - Students: 80% satisfaction.- Admin LPPM and reviewers: over 88%. These scores highlight the need to enhance the user experience, particularly for students and first-time users, through UI/UX refinements and simplified navigation.

5. Comparison with Existing Systems Compared to existing academic platforms, SIMPELMAS provides more comprehensive integration of research and community service management. Most existing systems focus only on administrative data without real-time analytics. The centralization of HR data and research outputs in SIMPELMAS allows more accurate tracking and evaluation, enhancing institutional accountability.

6. Practical Implications The use of SIMPELMAS enables LPPM and faculty to manage research and community service more transparently. Real-time reporting and automated verification reduce manual workload, allowing academic staff to focus on quality improvement. The platform also supports strategic planning through data analytics and reporting features.

7. Limitations Despite high system performance, limitations include: - Suboptimal user experience for mobile and rural users. - The need for training sessions for less tech-savvy users. These insights will guide further development and ensure broader adoption across departments. The integration of academic functions through SIMPELMAS demonstrates a practical model for other institutions seeking digital transformation in higher education.

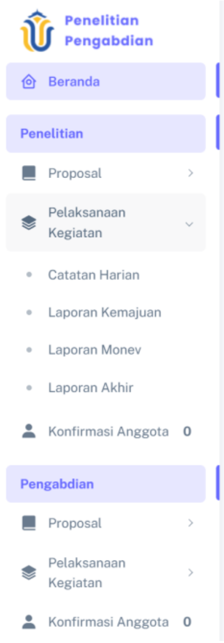


Fig. 2. Menu Features on the SIMPELMAS System

Before displaying the main page of SIMPELMAS, it is important to first understand the main role of this system in supporting academic and research management in higher education environments. SIMPELMAS is designed as an integrated platform that combines human resource data, research, and community service in one comprehensive system. With a data-based approach, this system not only facilitates the recording and reporting of academic activities, but also increases the effectiveness of decision-making and transparency in managing lecturer performance. Through an intuitive interface and complete features, SIMPELMAS provides more structured and efficient access for all users, both academics and other stakeholders. On Figure 3, users will be presented with various main menus designed to facilitate navigation and access to important information related to research, community service, and academic performance evaluation. This view provides a comprehensive overview of the main features of SIMPELMAS and how this system contributes to improving collaboration and academic achievement in higher education.

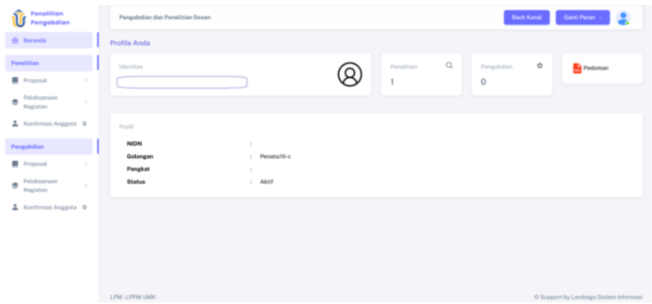


Fig. 3. Dashboard SIMPELMAS

In order to support the improvement of the quality and

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2339, Copyright ©2025
Submitted : April 25, 2025, Revised : May 10, 2025, Accepted : May 15, 2025, Published : May 26, 2025

efficiency of research management in higher education, SIMPELMAS provides the Addition of Research Proposals feature. This feature is designed to facilitate lecturers and researchers in submitting research proposals systematically, structured, and well-documented. Through this page, users can input important information related to research, such as title, field of study, research team, and funding sources. With a data-based system, every proposal submitted will be stored in an integrated manner, allowing the verification and evaluation process to run more transparently and efficiently. In Fig. 4, the Addition of Research Proposals page interface is displayed, which makes it easier for users to compile and manage research proposals before they are submitted for further selection and approval processes.

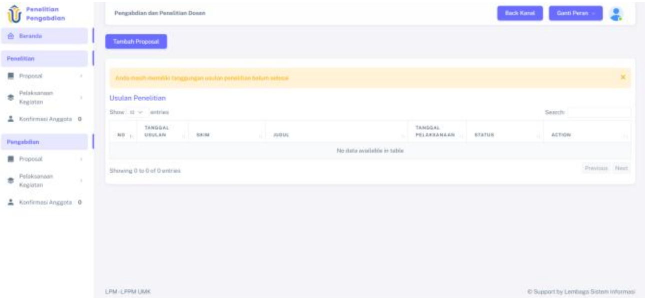


Fig.4. Propose Research Page

The process of managing research and community service in SIMPELMAS is designed to ensure the accuracy and validity of data integrated into the system. One important feature in this platform is the Member Confirmation Page, which functions as a verification step to ensure that each member in a research or community service activity has been registered correctly. Through this page, users can confirm membership by reviewing the list of members that have been entered, making changes if necessary, and approving the team structure involved in the project. This feature not only increases transparency in the management of academic teams but also helps in more accurate administration and reporting processes. Fig. 5 shows the interface of the member confirmation page, which makes it easy for users to ensure the validity of the data before proceeding to the next stage in the SIMPELMAS system.

Fig.5. Confirmation Page

Testing on the system focuses on the functionality of each menu and the integration between the menus. In this system testing table, I divide it into 7 main categories:

1. *Research Menu Testing* - Testing all submenus under the Research menu (Home, Research, Proposal, Activity Implementation, Daily Notes, Progress Report, Monitoring and Evaluation Report, Final Report, and Member Confirmation)

TABLE 1. RESEARCH MENU TESTING

No	Submenu	Tested Function	Expected Result	Success Rate (%)	Evaluator
1.1	Dashboard	Dashboard display	Dashboard appears correctly & updated	100%	LPPM Admin
1.2	Research	Research list display	Research data appears accurately	95%	Research Coordinator
1.3	Proposal Submission	New proposal submission	Proposal saved correctly	98%	Research Lecturer
1.4	Proposal Revision	Proposal revision	Revised data saved properly	97%	Lecturer & Reviewer
1.5	Activity Execution	Progress documentation	Progress data saved completely	96%	Research Lecturer
1.6	Daily Notes	Activity recording	Daily notes stored completely	94%	Research Team
1.7	Progress Report	Upload progress report	File uploaded without error	99%	Research Lecturer
1.8	Fund Monitoring	Fund evaluation	Monetary data is accurate & aligned	96%	Monitoring Team
1.9	Final Report	Final report upload	Final report saved correctly	98%	Research Coordinator
1.10	Member Verification	Member confirmation	All members verified	100%	Lead Researcher

2. *Testing the Community Service Menu* - Testing the submenus under Community Service (Proposal, Activity Implementation, and Member Confirmation).

TABLE II. COMMUNITY SERVICE MENU TESTING

No	Submenu	Tested Function	Expected Result	Success Rate (%)	Evaluator
2.1	Proposal Submission	Proposal submission	Proposal saved correctly	97%	Community Lecturer
2.2	Proposal Revision	Proposal revision	Revised data saved without error	96%	Lecturer & Reviewer
2.3	Activity Execution	Activity recording	Activity information saved completely	95%	Community Service Team
2.4	Member Verification	Member confirmation	All members verified	100%	Community Leader

3. *General Functional Testing* - Testing basic functions such as login, notifications, navigation, and file management that apply to the entire system.

TABLE III. GENERAL FUNCTIONAL TESTING

No	Component	Tested Function	Success Rate (%)	Evaluator
3.1	Login/Authentication	System access	99%	IT Team
3.2	Access Security	Login validation	98%	IT Security Team
3.3	Menu Navigation	Page-to-page navigation	97%	UX Specialist
3.4	File Upload	Document upload	95%	IT Team
3.5	File Download	File retrieval/download	99%	IT Team

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2339, Copyright ©2025
Submitted : April 25, 2025, Revised : May 10, 2025, Accepted : May 15, 2025, Published : May 26, 2025

4. *Menu Integration Testing* - Testing how different menus interact with each other and share data correctly.

TABLE IV. MENU INTEGRATION TESTING

No	Component	Test Scenario	Success Rate (%)	Evaluator
4.1	Proposal Confirmation	Member notification	100%	LPPM Admin
4.2	Proposal Execution	Change approval	98%	Research Coordinator
4.3	Progress Report	Progress synchronization	97%	System Administrator
4.4	Data Security	Encryption and protection	99%	IT Security Team

5. *Responsive and Adaptive Testing* - Ensuring the system functions properly across a variety of devices and network conditions.

TABLE V. RESPONSIVE AND ADAPTIVE TESTING

No	Aspect	Test Device	Success Rate (%)	Evaluator
5.1	Layout	Desktop (1920×1080)	99%	UI Designer
5.2	Layout	Tablet (768×1024)	98%	UI Designer
5.3	Responsive Layout	Mobile (375×667)	97%	UI Designer
5.4	Network Performance	Slow connection	96%	IT Team

6. *Menu Security Testing* - Tests specific security aspects related to menu access and data manipulation.

TABLE VI. MENU SECURITY TESTING

No	Security Component	Tested Function	Success Rate (%)	Evaluator
6.1	Authentication	Login validation	99%	IT Security Team
6.2	Input Validation	SQL Injection check	98%	IT Security Team
6.3	Data Encryption	HTTPS & protection	100%	IT Security Team
6.4	Session Timeout	Automatic logout	97%	IT Security Team

7. *User Acceptance Testing* - Evaluating user satisfaction from different groups of system users.

TABLE VII. USER ACCEPTANCE TESTING

No	Test Group	Parameter	Success Rate (%)	Evaluator
7.1	Lecturers	Usability Testing	85%	UX Researcher
7.2	Students	Usability Testing	80%	UX Researcher
7.3	LPPM Admin	Task Completion	90%	Head of LPPM
7.4	Reviewer	Focus Group	88%	Reviewer Coordinator

For each test item, I have defined a specific test scenario, expected results, success criteria, and the responsible party as evaluator.

IV. CONCLUSION

This study developed SIMPELMAS, an integrated information system designed to manage research and community service data at Universitas Muria Kudus. Built using a prototype-based development methodology, the system was evaluated through a comprehensive series of functionality, security, integration, and usability tests. The results show that SIMPELMAS effectively centralizes academic data and enhances institutional efficiency. With functionality success rates above 95%, security performance reaching 99–100%, and user satisfaction levels exceeding 85%, the system demonstrates strong reliability and user acceptance. It significantly simplifies proposal submission, monitoring, and reporting processes while improving transparency and enabling data-driven decision-making. This research contributes to the digital transformation of academic management by offering a replicable model of integrated platform development. It addresses a gap in current systems that often operate in silos and lack real-time analytics capabilities. Future research is encouraged to incorporate artificial intelligence for predictive analytics, performance dashboards, and automation features. Additionally, enhancing user interface design and ensuring accessibility in low-bandwidth environments will be vital for broader adoption. Overall, SIMPELMAS serves as a strategic digital innovation that strengthens governance, academic performance, and institutional collaboration in higher education.

REFERENCES

[1] S. Sarah, S. Sugiarto, and M. Ahmad, “Digital-Based Academic Management Service : A Case Study in Open University,” *Jurnal Paedagogy*, vol. 10, p. 910, Jul. 2023, doi: 10.33394/jp.v10i3.8374.

[2] C. McInnis, “The impact of technology on faculty performance and its evaluation,” *New Directions for Institutional Research*, vol. 2002, pp. 53–62, Jun. 2002, doi: 10.1002/ir.46.

[3] M. I. Ekarini, A. Rahayu, D. Disman, and L. A. Wibowo, “Implementation of Digital Transformation and Government Enterprise Architecture in Improving the Performance of Integrated Social Services,” *Jurnal Manajemen Pelayanan Publik*, vol. 8, no. 2, pp. 742–761, 2024, doi: 10.24198/jmpp.v8i2.54095.

[4] C. Jim and H.-C. Chang, “The current state of data governance in higher education,” vol. 55, pp. 198–206, Feb. 2018, doi: 10.1002/pra2.2018.14505501022.

[5] C. Jim and H.-C. Chang, “The current state of data governance in higher education,” vol. 55, pp. 198–206, Feb. 2018, doi: 10.1002/pra2.2018.14505501022.

[6] D. Zustiyantoro, “The Proceedings of the English Language Teaching, Literature, and Translation (ELTLT) Data Integration and Lecturer Expertise: Its Relevance to the Academic Reputation of the Faculty of Languages and Arts UNNES,” vol. 11, no. 1, pp. 2580–7528, 2022.

[7] Daryono, “The Trends in Educational Management Research and Its Implications to Higher Education: A Bibliometric Analysis,” *Ilomata International Journal of Social Science*, vol. 4, pp. 350–365, Jul. 2023, doi: 10.52728/ijss.v4i3.776.

[8] M. I. Ekarini, A. Rahayu, D. Disman, and L. A. Wibowo, “Implementation of Digital Transformation and Government Enterprise Architecture in Improving the Performance of Integrated Social Services,” *Jurnal Manajemen Pelayanan Publik*, vol. 8, no. 2, pp. 742–761, 2024, doi: 10.24198/jmpp.v8i2.54095.

[9] J. Sithumini, A. Sanjuka, P. Ranawaka, H. Hasaranga, T. Samarakkody,

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2339, Copyright ©2025
Submitted : April 25, 2025, Revised : May 10, 2025, Accepted : May 15, 2025, Published : May 26, 2025

and G. Pathirana, *Systematic Review: The Role of Data Analytics in Enhancing Academic Performance Classroom interaction, Learning Analytics in Higher Education*. 2024.

[10] Y. Widodo, R. Korwa, and R. Nuraini, "Artificial Intelligence Based Decision Support System for Education Management in Higher Education," *al-fikrah: Jurnal Manajemen Pendidikan*, vol. 11, p. 352, Dec. 2023, doi: 10.31958/jaf.v11i2.12120.

[11] X. Wei, "Research on the Optimization of University Performance Management Information System in the Big Data Era," 2021, pp. 628–635. doi: 10.1007/978-3-030-69999-4_86.

[12] A.-R. Bologa, R. Bologa, G. Sabau, and M. Muntean, "Integrated information systems in higher education," *WSEAS Transactions on Computers*, vol. 7, May 2008.

[13] X. Zhu, S. Ge, and N. Wang, "Digital transformation: A systematic literature review," *Computers & Industrial Engineering*, vol. 162, p. 107774, 2021, doi: https://doi.org/10.1016/j.cie.2021.107774.

[14] I. Solin, R. Nugrahani, and N. Kusumoretno, "A REVIEW AND CASE STUDIES OF INTEGRATED MANAGEMENT SYSTEM IMPLEMENTATION IN THE ELECTRIC POWER TRANSMISSION BUSINESS," *Jurnal Kelola Jurnal Ilmu Sosial*, vol. 6, p. 50, Sep. 2023, doi: 10.54783/jk.v6i2.731.

[15] B. George and O. Wooden, "Managing the Strategic Transformation of Higher Education through Artificial Intelligence," *Administrative Sciences*, vol. 13, no. 9, 2023, doi: 10.3390/admsci13090196.

[16] D. Lee, S. Kim, and S. H. Cha, "Evaluating the effectiveness of research centers and institutes in universities: Disciplines and life cycle stages," *KEDI Journal of Educational Policy*, vol. 11, no. 1, pp. 119–137, 2014.

[17] H. Taherdoost, "Data Collection Methods and Tools for Research; A Step-by-Step Guide to Choose Data Collection Technique for Academic and Business Research Projects Hamed Taherdoost. Data Collection Methods and Tools for Research; A Step-by-Step Guide to Choose Data Coll," *International Journal of Academic Research in Management (IJARM)*, vol. 2021, no. 1, pp. 10–38, 2021.

[18] N. Carter, D. Bryant-Lukosius, A. DiCenso, J. Blythe, and A. J. Neville, "The use of triangulation in qualitative research," *Oncology nursing forum*, vol. 41, no. 5, p. 545—547, Sep. 2014, doi: 10.1188/14.onf.545-547.

[19] R. K. Yin, *Case Study Research and Applications: Design and Methods*. SAGE Publications, 2017.

[20] M. Saunders, P. Lewis, and A. Thornhill, *Research Methods for Business Students*. Financial Times/Prentice Hall, 2007.

[21] D. Mourtzis, J. Angelopoulos, and N. Panopoulos, "A Literature Review of the Challenges and Opportunities of the Transition from Industry 4.0 to Society 5.0," 2022. doi: 10.3390/en15176276.

[22] G. Bowen, "Document Analysis as a Qualitative Research Method," *Qualitative Research Journal*, vol. 9, pp. 27–40, Aug. 2009, doi: 10.3316/QRJ0902027.

[23] R. S. Pressman, *Software Engineering: A Practitioner's Approach*. in McGraw-Hill higher education. Boston, 2005.

[24] F. Davis and P. Warshaw, "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management Science*, vol. 35, pp. 982–1003, Aug. 1989, doi: 10.1287/mnsc.35.8.982.

[25] J. S. Valacich and J. F. George, *Modern Systems Analysis and Design*. Pearson Education, Incorporated, 2024.

[26] K. C. Laudon and C. G. Traver, *E-Commerce 2023: Business, Technology, Society, Global Edition*. Pearson Education, 2023.

[27] J. Lewis and J. Sauro, "USABILITY AND USER EXPERIENCE: DESIGN AND EVALUATION," 2021, pp. 972–1015. doi: 10.1002/9781119636113.ch38.

[28] S. Davis, *Software Engineering: The Year 11 Course*. Parramatta Education Centre, 2024.

[29] R. Stair and G. Reynolds, *Principles of Information Systems*. Cengage Learning, 2016.

[30] Nirsal *et al.*, *MODERN DATABASE MANAGEMENT*. 2024.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2339, Copyright ©2025
Submitted : April 25, 2025, Revised : May 10, 2025, Accepted : May 15, 2025, Published : May 26, 2025

Implementation of Elementary School Student Attendance Information System Based on Android using AppSheet

Reni Haerani[1]*, Putri Ayu Permata Devi[2], Penny Hendriyati[3], Ahmad Sofan Ansor[4]
Information System Study Program, Bina Insani University[1]*, Digital Business of Polytechnic PGRI Banten[2], Information Systems, Insan Unggul College of Computer Science Technology[3],
Public Sector Human Resource Management, Polytechnic PGRI Banten[4]
renihaerani@binainsani.ac.id[1], ayudevie8682@gmail.com[2],pennyhendriyatiinsanunggul@gmail.com[3],
sofanansor65@gmail.com[4]

***Abstract*— This study aims to implement an Android-based elementary school student attendance information system using the AppSheet platform. This system is designed to replace the manual attendance method that is still widely used, making it easier for teachers to record student attendance and provide real-time attendance reports. AppSheet was chosen because of its ability to create Android-based applications without the need for complex programming skills. This system has key features such as attendance recording, cloud data storage, and integrated report access. The study results show that implementing this attendance information system can increase the efficiency of the attendance recording process by 89% compared to the manual method. In addition, attendance reports can be accessed by the school quickly and accurately. This system also received positive responses from teachers and administrative staff because of its ease of use. This system also improves the efficiency of attendance data management and makes the communication process between schools and parents more effective. Thus, the Android-based attendance information system using AppSheet provides a practical solution relevant to current technological developments, supporting digital transformation for managing student attendance data in elementary schools.**

Keywords—Students Attendance, Information System, Elementary School, Android, AppSheet, Efficiency.

I. Introduction

The advancement of information technology has brought significant changes in various aspects of life, including in the field of education. In today's digital era, school administration management, especially related to student attendance, requires a more modern and efficient approach. The manual attendance recording system, still widely used in elementary schools, often faces various obstacles, such as the risk of data loss, human error, and lack of accessibility and time efficiency.

Implementing a technology-based information system is a relevant solution to overcome these obstacles. One platform that allows application development without requiring complex programming skills is AppSheet. With AppSheet, school administrators can develop Android-based attendance applications that can be used by teachers and school staff to record, manage, and monitor student attendance in real-time and can be accessed anywhere [1].

This Android-based attendance system is designed to provide features such as daily attendance recording, automatic data recap, notification to parents, and integration with mobile devices. The use of this technology is expected to improve the efficiency of school administration processes, reduce teachers' workloads, and provide better transparency to schools and parents. Digitalization of attendance increases the transparency and accuracy of student attendance recording [2]. Using web-based or mobile applications in various educational institutions has improved attendance data management [3].

Related research shows that digital technology positively impacts student attendance in several ways. According to research [4] on the use of mobile technology in student attendance systems, the application developed uses QR codes as a medium for recording attendance. Although this system is effective, its implementation requires additional devices, such as QR code scanners, which can be an obstacle in schools with limited resources. A study focuses on using cloud technology to store and manage student attendance data [5]. Cloud-based systems provide flexible access to data for various stakeholders, including teachers and parents. However, the study also highlights challenges related to data security and the cost of subscribing to cloud services.

Study [6] explains how to implement a student attendance information system using an Android-based application. Although the Android-based AppSheet has limitations in complex design and function customization, the survey results show that AppSheet allows for fast and easy system development and is highly regarded as an application media expert in terms of functionality.

Therefore, this study aims to describe the process of developing an Android-based attendance system using AppSheet, analyze the benefits achieved, and evaluate the system's effectiveness in supporting student attendance management in elementary schools. The results are expected to provide practical guidance to other elementary schools that want to apply similar techniques in attendance management.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2351, Copyright ©2025
Submitted : May 1, 2025, Revised : May 8, 2025, Accepted : May 9, 2025, Published : May 26, 2025

II. Literature Review

A. Attendance Information System

A time and attendance information system is an application or software for recording and managing attendance data [7]. It replaces manual methods that are prone to errors and can be used in various situations, including work and educational environments.

B. Android

Android is an operating system for Linux-based mobile devices, including an operating system, middleware, and applications. It is also a Linux-based operating system for touchscreen mobile phones and tablet computers [8]. However, along with its development, Android has become a platform that innovates quickly. This is inseparable from the main developer behind it: Google. Google bought Android and created its platform. The Android platform comprises a Linux-based operating system, a graphical user interface (GUI), a web browser, and downloadable end-user applications. Developers can work freely to create the best and most open applications for use by a wide variety of devices [9].

C. AppSheet

AppSheet is a no-code platform that allows users to build data-driven applications without programming knowledge. However, to enable digital presence on AppSheet, AppSheet offers the following benefits: Ease of use, cloud integration, and customization, which allows you to customize its features according to your needs. Manage attendance data, reports and automatic notifications. AppSheet works by connecting spreadsheets to applications. This application can also be used on mobile phones both online and offline [10].

D. Spreadsheet

A spreadsheet is a computer program that allows users to organize, store, and analyze data in tables. It also stores, displays, and processes data in rows and columns. [11].

III. Research Method

Data analysis techniques are stages of the research process that process collected data to answer existing questions. The analysis method for building a student attendance information system is based on the prototype system development method. The research stages will be explained in the figure below

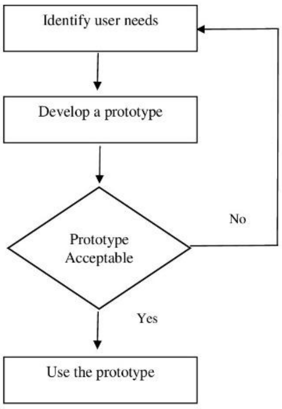


Fig 1. Research Stages

The prototyping method is a very fast technique for repeating the interaction process to develop and test new application behaviour models so that they can be used properly [12]. Prototypes can also overcome the problem of misunderstanding between users and analysts, namely that users cannot identify each other. Prototyping is a widely used system development technique. This technique also provides an opportunity for developers and users to interact with each other during the creation process, allowing developers to model the software created easily [13]. The process stages applied in this study utilize a prototype model diagram through five processes: communication, quick plan, quick design, prototype construction, and delivery and feedback, as explained in Figure 1.

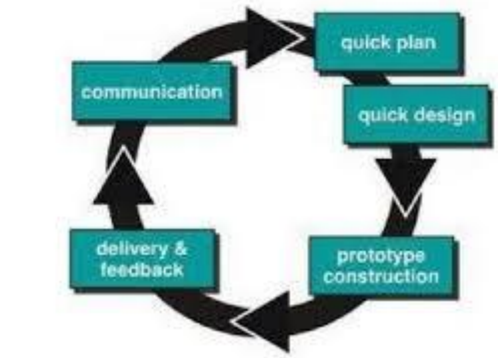


Fig 2. Prototype Model

The processes can be explained as follows:

a. ***Communication*:** During this phase, the developer and client review and establish general goals, desired requirements and an overview of the parts that are needed next.
b. ***Quick Plan*:** in this phase, the design is carried out quickly, revealing all known aspects of the software. This design becomes the basis for making a prototype.
c. ***Modelling Quick Design*:** This phase focuses on representing aspects of the software that are visible to the customer/user. It usually involves making a

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2351, Copyright ©2025
Submitted : May 1, 2025, Revised : May 8, 2025, Accepted : May 9, 2025, Published : May 26, 2025

prototype.

d. ***Construction of Prototype*:** creating a framework or designing a prototype of the software that is made.

e. ***Delivery & Feedback*:** The prototype created by the developer is distributed to the user/client for evaluation. The customer then provides feedback that is used to change the requirements for the software that is made.

IV. RESULT AND DISCUSSION

A. *System Implementation*

The Android-based time attendance system was developed using AppSheet, a low-code platform that enables rapid application development. The application allows teachers and school officials to record student attendance in real time and provides stakeholders access to reports.

User Interface Prototype

1. *Login Page Design*

Fig 3. Login Page

2. *Student Form Page*

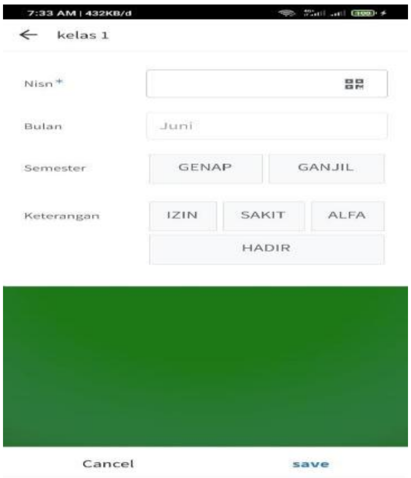


Fig 4. Student Form

3. *Student Data Page*

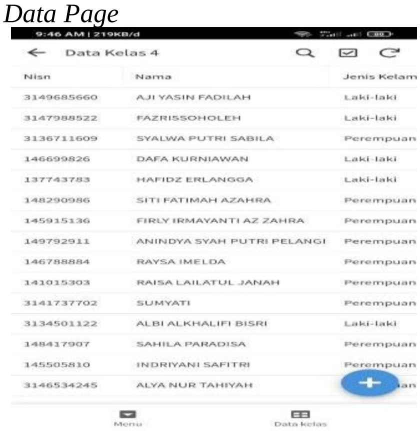


Fig 5. Student Data Page

4. *Student Data Report Page*

The data report page is used to print a report of all student data.

Fig 6. Student Report Page

Implementing the attendance system using AppSheet yielded significant results that impacted the school's operations and stakeholders, including teachers, parents, and administrators. Below is an outline of the key outcomes:

1. ***Enhanced Efficiency in Attendance Tracking***

   Time-Saving: Teachers could mark attendance within seconds using the app instead of manual record-keeping, which was more time-consuming.

   Automation of Reports: Daily and monthly attendance reports were automatically generated and distributed, eliminating the need for manual compilation.

2. ***Improved Communication***

   Real-Time Notifications: Parents received instant updates about their child's attendance status, fostering better communication between schools and families.

   Transparent Data Access: Administrators could access attendance data in real-time, allowing for immediate interventions in cases of frequent absenteeism.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2351, Copyright ©2025
Submitted : May 1, 2025, Revised : May 8, 2025, Accepted : May 9, 2025, Published : May 26, 2025

3. ***Reduction in Errors***
   Accurate Data Entry: QR code scanning or digital forms reduced errors caused by illegible handwriting or manual miscalculations.
   Data Validation: AppSheet’s built-in validation rules ensured incorrect or incomplete data could not be submitted.

4. ***Accessibility and User Adoption***
   Ease of Use: The intuitive interface of the AppSheet app enabled even non-tech-savvy users to navigate the system effortlessly.
   Broad Device Compatibility: Since the system was Android-based, it was accessible to a majority of users, with an option for iOS or web-based access.
5. ***Enhanced Monitoring and Analytics***
   Attendance Trends: Administrators could easily track attendance patterns across grades, identifying students with chronic absenteeism.
   Improved Decision-Making: Data-driven insights helped school management implement targeted strategies to address attendance issues.
6. ***Cost and Resource Optimization***
   Paperless System: The digital system reduced dependency on paper records, contributing to environmental sustainability and cost savings.
   Low Development Costs: AppSheet's no-code nature eliminated the need for hiring specialized developers, reducing overall implementation costs.
7. ***Positive Feedback from Stakeholders***
   Teachers: Reported satisfaction due to reduced administrative workload.
   Parents: Appreciated the transparency and real-time notifications, which helped them stay informed about their child’s school activities.
   Administrators: Valued the system’s ability to provide instant access to attendance data and reports.

Implementing an Elementary School Student Attendance Information System based on Android using AppSheet proved successful. It streamlined attendance management, reduced errors, and enhanced stakeholder communication. The results demonstrated that such systems could significantly modernize and improve school administration, paving the way for future enhancements and broader adoption.

B. System Testing

The next stage is the system evaluation stage. Where testing is carried out using black box testing and usability testing. Black box testing is a technique that analyzes software functionality by comparing input and output values [14]. As explained in Table I, the results of black box testing are intended to determine whether the design of this information system is acceptable.

TABLE I. BLACK BOX TEST RESULTS

No	Module	Test	Expected Result	Conclusion
1	Login	Enter the correct username and password Entering the wrong username and password	Enter the main page Login failed; return to home page	Valid
2	Manage User Menu	Complete the user management process by adding, editing or deleting the required users.	The process of adding, editing or deleting the desired users has been completed.	Valid
3	Add Document Menu	Add transactional data related to archived documents.	Transactions are successfully entered, and the amount increases in the Add Document menu.	Valid
4	Upload Document Menu	The process of downloading the desired document is in progress.	The transaction successfully uploaded the desired document.	Valid
5	Document Download Menu	Download the desired document.	The transaction successfully downloaded the desired document.	Valid
6	Edit Document Menu	Make changes to documents.	Transaction successfully made the required document changes.	Valid
7	Delete Documents Menu	Delete documents.	Transaction successfully deleted the desired document.	Valid

Usability testing was conducted to determine the opinions of respondents who acted as users regarding the monitoring system, which was developed and easy to use. This test involved 40 respondents whose opinions were assessed on learning ability, effectiveness, memory, error tolerance, and satisfaction. Users rated their experiences using a Likert scale of 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), and 5 (strongly agree). The survey results are shown in Table 2.

TABLE II. PERCENTAGE OF USABILITY TESTING QUESTIONNAIRE ANSWERS

Question	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Number of Respondents	Percentage
P1	15	20	5	-	-	40	87%
P2	22	10	8	-	-	40	85%
P3	21	16	3	-	-	40	85%
P4	23	15	2	-	-	40	93%
P5	34	6	-	-	-	40	96%

Based on the test results in Table II, the calculation results for the percentage of questionnaire answers from 40 respondents to 5 questions were 87% for the question all functions and menus of this android app work fine, 85% for questions this android app is easy to understand and easy to use, 85% for questions this android app has an attractive appearance, 93% regarding the fourth question that this android app helps in

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2351, Copyright ©2025
Submitted : May 1, 2025, Revised : May 8, 2025, Accepted : May 9, 2025, Published : May 26, 2025

data processing and 96% this android app can speed up admin work.

V. Conclusion

Implementing an Android-based student attendance information system in elementary schools using AppSheet has provided an effective and efficient solution to the attendance management process. This system is useful for teachers and schools to record, monitor, and manage student attendance data in real time. AppSheet offers several important benefits compared to manual attendance methods: The attendance process is now faster and less prone to errors. Attendance data can be accessed directly from Android devices, improving communication between teachers, schools, and parents. The digital system makes attendance information more accurate and transparent, increasing trust between schools and parents. Overall, this AppSheet-based attendance system shows great potential to improve operational efficiency and the quality of student attendance management at the elementary school level. With further development, this system can be expanded to include additional features, such as automatic reporting and integration with other academic systems, thus supporting digital transformation in Education.

References

[1] Z. Zulhidayati, R. Okra, H. A. Musril, and S. Derta, "Perancangan Sistem Absensi Online Berbasis Mobile untuk Guru dan Pegawai menggunakan Appsheet," *Juwara J. Wawasan dan Aksara*, vol. 4, no. 1, pp. 23–32, 2024, doi: 10.58740/juwara.v4i1.83.

[2] M. P. Rad, "IMPLEMENTATION OF A WEB-BASED STUDENT AND TEACHER ATTENDANCE SYSTEM WITH QR CODE INTEGRATION USING THE RAD IMPLEMENTASI SISTEM PRESENSI SISWA DAN GURU BERBASIS WEB DENGAN INTEGRASI QR CODE," 2025.

[3] M. Alda, M. Juarsyah, A. Nugraha, and L. R. Alfachry, "Aplikasi Absensi Mahasiswa Kerja Praktik Menggunakan QR Code Berbasis Android," *J. Manaj. Inform.*, vol. 14, no. 1, pp. 27–41, 2024, doi: 10.34010/jamika.v14i1.11775.

[4] D. Hamdani, A. P. W. Wibowo, and H. Heryono, "Perancangan Sistem Presensi Online dengan QR Code Menggunakan Metode Prototyping," *J. Teknol. dan Inf.*, vol. 14, no. 1, pp. 62–73, 2024, doi: 10.34010/jati.v14i1.11844.

[5] R. Kurniawan, T. H. Budianto, and W. Yandi, "Rancang Bangun Aplikasi Presensi Dosen dan Mahasiswa Berbasis Android dan Cloud Server," *J. Ecotipe (Electronic, Control. Telecommun. Information, Power Eng.*, vol. 9, no. 1, pp. 97–102, 2022, doi: 10.33019/jurnalecotipe.v9i1.2971.

[6] A. M. Rizki, D. L. Setiawan, P. Teknologi, and U. M. Kuningan, "MENGGUNAKAN APLIKASI APPSHEET BERBASIS ANDROID," vol. 8, no. 5, pp. 10704–10712, 2024.

[7] M. K. Ikhwanudin, Sopingi, and Agustina srirahyu, "Pemodelan Sistem Absensi Karyawan Di PT Egref Telematika Menggunakan Tekonologi QR Dan GPS," *JEKIN - J. Tek. Inform.*, vol. 4, no. 3, pp. 600–609, 2024, doi: 10.58794/jekin.v4i3.868.

[8] H. Gusdevi, S. Kuswayati, M. Iqbal, M. F. Abu Bakar, N. Novianti, and R. Ramadan, "Pengujian White-Box Pada Aplikasi Debt Manager Berbasis Android," *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 4, no. 1, pp. 11–22, 2022, doi: 10.53580/naratif.v4i1.147.

[9] I. Irmawati, M. Olivya, and T. Indrabulan, "Perancangan Dan Implementasi Dashboard Rekapitulasi Kompensasi Sebagai Media Monitoring Ketidakhadiran Mahasiswa," *Semin. Nas. Has. Penelit. Pengabdi. Kpd. Masy.*, vol. 7, no. 1, pp. 135–140, 2022.

[10] I. Handayani, H. Kusumahati, and A. Nurul Badriah, "Alpiah Nurul Badriah Title of manuscript is short and clear," vol. 7, no. 2, p. 177, 2017.

[11] M. Irsan, F. T. S. B, and A. Husain, "Implementasi Aplikasi Pandas (Phyton) Dalam Mengelola Data Excel Sebagai Media Persiapan Pelaporan Nilai Raport Siswa," *J. Pengabdi. Masy. Bangsa*, vol. 2, no. 4, pp. 1243–1249, 2024, doi: 10.59837/jpmba.v2i4.977.

[12] Wilda Syahfitri, "Penerapan QR Code Dengan Foto Diri dan Lokasi Pada Absensi Karyawan Berbasis Android," *J. Komput. Teknol. Inf. dan Sist. Inf.*, vol. 2, no. 2, pp. 339–349, 2023, doi: 10.62712/juktisi.v2i2.79.

[13] I. Nur Azis, M. Indra, and M. Khoirusofi, "Perancangan Sistem Informasi E-Arsip Pada CV Asli Satia Persada Berbasis Web Menggunakan Metode Prototype," *Biner J. Ilmu Komput. , Tek. dan Multimed.*, vol. 1, no. 2, pp. 320–331, 2023.

[14] D. Irawan, Y. Saputra, H. Hamuda, T. Komputer, and U. M. Karanganyar, "Analisis Blackbox Testing Dalam Pengembangan," 2024.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2351, Copyright ©2025
Submitted : May 1, 2025, Revised : May 8, 2025, Accepted : May 9, 2025, Published : May 26, 2025

Optimizing Gated Recurrent Unit (GRU) for Gold Price Prediction: Hyperparameter Tuning and Model Evaluation on Historical XAU/USD Data

Abdul Faqih[1]*, Muhammad Jauhar Vikri[2], Ita Aristia Sa'ida[3]
Department of Informatics Engineering[1], [2], [3]
University of Nahdlatul Ulama Sunan Giri
Bojonegoro, Indonesia
abdulfaqih7777@gmail.com [1], vikri@unugiri.ac.id [2], itaaristia@unugiri.ac.id [3]

Abstract*— This study investigates the use of a Gated Recurrent Unit (GRU) model with a four-layer architecture for daily gold price closing prediction, motivated by the model's ability to effectively capture temporal dependencies in time series data. Gold price forecasting is highly challenging due to its volatility and external factors, making it an important area of research for investors and financial analysts. By systematically optimizing hyperparameters through 72 combinations of epochs, batch size, GRU layer units, and dropout rates, the study identifies the optimal configuration (100 epochs, batch size of 16, 256 units, dropout rate 0.1) based on MSE performance on validation data. The best model achieved MAE of 25.76, MSE of 954.97, and RMSE of 30.90, after inverse transformation on test data. These results highlight the potential of the GRU model in accurately forecasting gold prices, with implications for financial decision-making . However, the prediction error suggests that further improvements could be made by incorporating external factors or exploring advanced model architectures.

***Keywords*— Gated Recurrent Unit (GRU), Gold Price Prediction, Hyperparameter Optimization, Time Series**

I. Introduction

Investment activities are an integral part of the modern economy, where individuals and institutions allocate capital with the aim of generating returns or preserving asset value in the future. A variety of asset classes are available to investors, ranging from stocks, bonds, real estate, to commodities, each with distinct risk profiles and potential returns. Among the many investment options available, individuals tend to choose investments that offer higher returns [1]. One attractive option is gold, which is not only valued as a raw material for jewelry and technology, but also widely recognized as a vital investment instrument. Gold serves as a store of value *(safe haven)* especially during times of global economic turmoil, as well as a hedge against inflation and currency devaluation [2].

Price fluctuations in XAU/USD are driven by factors such as changes in currency exchange rates, interest rate policies (especially those of central banks like the U.S. Federal Reserve), inflation rates, and geopolitical tensions [3]. The dynamics of physical gold supply and demand also contribute to these fluctuations, making it crucial for investors to accurately predict price movements [4]. The complex and interrelated nature of these factors makes forecasting gold prices a challenging yet valuable task.

This study aims to address the challenge of predicting daily gold prices by implementing the Gated Recurrent Unit (GRU) model, which is well-suited for capturing the temporal dependencies in financial time series data. By evaluating various hyperparameter configurations, this research seeks to improve the accuracy of gold price predictions and provide a more reliable tool for investors.

To support investment decision-making amid gold price volatility, various studies have been conducted to develop accurate prediction methods. [5]This study compares the performance of models based on Recurrent Neural Networks (RNN), including LSTM and GRU, in forecasting economic and financial data in Indonesia, such as the IHSG, export value, and GDP. The results of this study indicate that the GRU model performs best overall and is more stable than RNN and LSTM on the data. On average, GRU recorded the smallest MAPE values on IHSG (Index Harga Saham Gabungan) data (0.3695%), export data (7.36%), and GDP data (1.77%). However, this study does not specifically focus on gold price prediction. The study also suggests increasing the number of scenarios with other combinations of hyperparameters and using model search techniques [6].

Specifically implements and compares GRU, Bi-GRU, LSTM, and Bi-LSTM for global gold price prediction. This study uses historical gold price data from *Yahoo Finance* and explores various optimization techniques, *batch sizes,* and *time steps*. The results of this study indicate that the Bi-GRU model with *Adam optimization*, a *batch size* of 8, and a *time step* of 20 provides the best performance for global gold price prediction, with an MSE value of 4.1153, an RMSE value of 2.0286, an MAE value of 1.5881, and a MAPE value of 0.8857%. Although relevant to the topic of gold price prediction using GRU and its variations, this study did not use the best hyperparameter selection [7].

A separate study evaluated the efficacy of various deep learning algorithms, including *Artificial Neural Network* (ANN), *Convolutional Neural Network* (CNN), *Recurrent Neural Network* (RNN), and *Long Short-Term Memory* (LSTM), in predicting gold prices utilizing a dataset sourced from Kaggle. The study determined that the CNN model

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2352, Copyright ©2025
Submitted : May 1, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

utilizing *Adam optimization* and the MSE loss function exhibited optimal performance in predicting gold prices within the used dataset. The CNN model achieved the smallest MAE value of 0.004848717761305338, the smallest MSE value of 4.3451079619612133e-05, and the smallest RMSE value of 0.006591743291392053. Although comparing RNN and LSTM, this study did not explore the performance of GRU in depth and recommends the use of a larger and more recent dataset as well as exploration of other models such as Transformer for further research.

Based on this literature review, there is a research gap focusing on gold price prediction in Indonesia using the GRU model. Although the GRU model has shown good performance in forecasting economic time series data in Indonesia and global gold price prediction, its in-depth application for specific gold price prediction in Indonesia with adequate *hyperparameter* exploration still requires further research. Therefore, This study seeks to develop a predictive model for gold closing prices (XAU/USD) utilizing the Gated Recurrent Unit (GRU) architecture, while systematically assessing the effects of diverse *hyperparameter* combinations: *epochs* (50, 100, 150), *batch sizes* (16, 32, 64), *neuron counts* (50, 100, 128, 256 units), *dropout rates* (0.1, 0.0), and a *learning rate* of 0.0001 on model performance. The performance of each combination will be measured and compared using the *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), and *Root Mean Squared Error* (RMSE) metrics on the test data to identify the most optimal configuration.

II. METHODOLOGY

The research process was carried out in several stages, starting with the collection of gold price data (XAU/USD) from TradingView, a popular financial data platform. Preprocessing steps were then performed, including normalization using the *Min-Max Scaler* to scale the data within the [0,1] range. A time series sequence was created with a timestep of 30, representing the past 30 days of price data for each prediction. Next, the dataset was split into training (80%) and testing (20%) sets, ensuring that the data followed a temporal order without shuffling, to maintain the integrity of time series forecasting. The GRU model with a four-layer hidden architecture was then applied, utilizing the Adam optimizer and MSE loss function. Various *hyperparameter* values were tested, including epochs (50, 100, 150), *batch sizes* (16, 32, 64), number of neurons (50, 100, 128, 256 units), *dropout* (0.1, 0.0), and *learning rate* (0.0001).

The choice of these hyperparameters was based on preliminary experiments and existing literature that indicated these values would provide a good balance between training time and model performance. For instance, the choice of batch size and number of epochs was influenced by previous research in financial time series forecasting, which suggested these values as optimal for minimizing overfitting and improving generalization. Despite the benefits of automated *hyperparameter* optimization methods like Grid Search or Random Search, they were not employed in this study due to time constraints and computational resource limitations. Instead, a more manual approach was taken to test a broad range of hyperparameters, ensuring that the model's performance could be thoroughly evaluated with reasonable computational effort.

Validation during training was carried out using a 20% split of the training data to monitor model performance and prevent overfitting through Early Stopping. Final evaluation was conducted on the test dataset, using the MAE, MSE, and RMSE metrics to compare the performance of different hyperparameter configurations. This research method is presented in the form of a flowchart in Fig. 1.

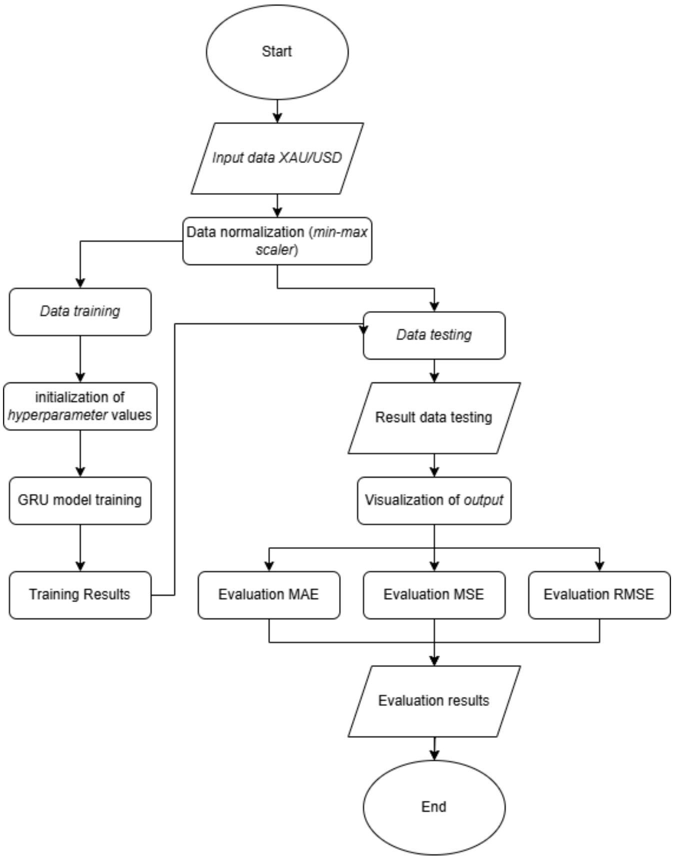


Fig 1. Research Method

A. Dataset

This analysis utilizes historical data on the price of gold traded against the US dollar (XAU/USD). This data was obtained from the TradingView platform (www.tradingview.com), a popular online platform that provides real-time charting and financial market analysis tools. The data includes time, opening price, high price, low price, and closing price. The data used spans from November 2006 to March 2025, with some of the data presented in Table 1. Utilizing large and diverse datasets is crucial in financial analysis as they enable more accurate predictions and better generalization of models. Integrating deep learning and big data algorithms significantly enhances the accuracy of financial risk behavior predictions, highlighting the importance of extensive datasets in financial forecasting [8].

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2352, Copyright ©2025
Submitted : May 1, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

TABLE I. XAU/USD DATASET FROM 2006 TO 2025

Time	Open	High	Low	Close
2006-09-21 21:00:00	584.2	592.7	584.2	589.65
2006-09-24 21:00:00	589.65	592	582.3	590.7
2006-09-25 21:00:00	590.7	594	586.1	591.8
2006-09-26 21:00:00	591.8	603.65	589.8	603
2006-09-27 21:00:00	603	607.1	600.4	601.1
2006-09-28 21:00:00	601.1	603.8	594.4	598.25
2006-10-01 21:00:00	598.25	604.15	594.75	596.7
2006-10-02 21:00:00	596.7	598.5	573.8	574.8
…	…	…	…	…
2025-03-02 22:00:00	2873.14	2895.21	2859	2892.985
2025-03-03 22:00:00	2892.7	2927.91	2881.98	2917.85
2025-03-04 22:00:00	2917.885	2929.98	2894.35	2919.265
2025-03-05 22:00:00	2918.685	2926.72	2891.15	2911.1
2025-03-06 22:00:00	2912.935	2930.54	2896.91	2909.55
2025-03-09 21:00:00	2912.41	2918.385	2896.71	2904.68

B. Data Normalization

Data normalization is an important technique in machine learning and time series analysis, especially when using gradient-based models such as GRU. Normalization aims to minimize errors, ensure data is scaled and easy to process in models. In this case study, we use the min-max scaling technique with an interval of [0,1]. The formula for normalization is shown in equation (1). Reference [9] data normalization techniques, such as *Min-Max scaling* and *Z-normalization*, have been explored to enhance model performance. The results of normalization with *min-max scaler* are shown in Fig. 2.

$$X_{scaled} = \frac{X - X_{min}}{X_{min} - X_{min}} \quad (1)$$

Explanation :
X_{scaled} = value of the data after scaling.
X = original value of the data.
X_{min} = minimum value of the feature in the dataset.
X_{max} = maximum value of the feature in the dataset.

The results of data normalization are shown in the following Fig 2:

Fig 2. Data after data normalization process

C. Data Splitting

Maintaining temporal order during data splitting is crucial for time series forecasting models to prevent data leakage and ensure realistic performance evaluation, as highlighted [8]. The constructed dataset (x and y) is partitioned into training and testing subsets, with an allocation of 80% for training and 20% for testing. The division is performed sequentially *(shuffle=False)* to maintain the time sequence, in accordance with the characteristics of time series data. The data division visualization is presented in Fig. 3.

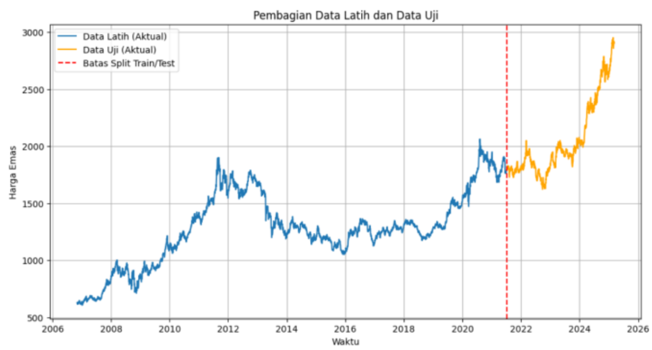


Fig 3. Training and Testing Data Split

D. GRU Model

The *Gated Recurrent Unit* (GRU) is a *Recurrent Neural Network* (RNN) architecture developed to mitigate the vanishing gradient issue commonly faced by conventional RNNs, particularly when processing sequential data with long-term dependencies [6].GRU can be considered an evolution of simpler and more efficient RNNs, with the ability to learn and retain information from data sequences over longer periods. The primary function of GRU in neural networks is to process sequential data by retaining relevant information from previous time steps while filtering out irrelevant or outdated information. This allows the network to understand the temporal context in the data and make more accurate predictions.

The GRU structure simplifies the recurrent unit by introducing two main gates, the *update gate* and the *reset gate*. The *Update gate* (z_t) controls the proportion of information from the previous hidden state (h_{t-1}) that is passed on to the current hidden state (h_t). Mathematically, the update gate is

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2352, Copyright ©2025
Submitted : May 1, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

calculated using the following formula:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (2)$$

Explanation:
σ = sigmoid function
x_t = current input
h_{t-1} = previous hidden state
W_z and U_z = weight matrices
b_z = bias for the update gate.

The *reset gate* (r_t) determines how much of the previous hidden state is ignored in the calculation of the current hidden state, calculated using the formula:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (3)$$

Explanation:
σ = sigmoid function
x_t = current input
h_{t-1} = previous hidden state
W_r and U_r = weight matrices
b_r = bias for the reset gate.

The internal architecture of the GRU entails computing the candidate *hidden state* ($\tilde{h}_t$), which is derived from the current input and the preceding *hidden state*, adjusted by the *reset gate*. The candidate concealed state is calculated using the subsequent formula:

$$\tilde{h}_t = tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \quad (4)$$

Explanation:
$tanh$ = hyperbolic tangent function
r_t = reset gate
h_{t-1} = previous hidden state
W_h and U_h = weight matrices
b_h = bias for the candidate hidden state
$\odot$ = denotes element-wise multiplication.

An illustration of GRU model hidden state computation is shown in Fig. 4.

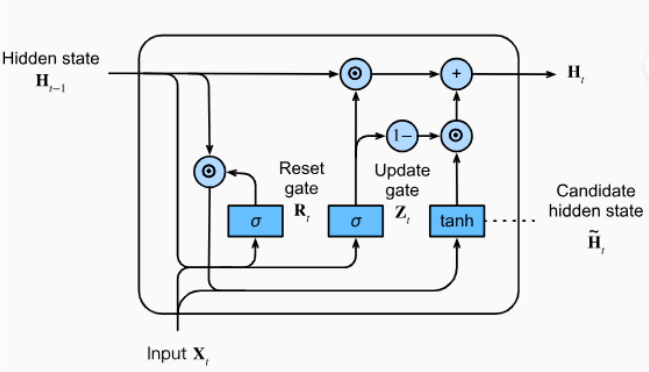


Fig 4. GRU Model Hidden State Computation

At this stage, the optimal GRU model architecture is built. The model is configured as a sequential model, meaning layers are added sequentially. The prepared GRU model is trained using the training data. The model specifications are as follows:

1. Optimizer: Adam
2. Hidden layers: 4
3. Number of Neurons: 50, 100, 128, 256 units
4. Timestep: 30
5. Epochs: 50, 100, 150 (with early stopping)
6. Batch Size: 16, 32, 64
7. Dropout: 0.1, 0.0
8. Learning rate: 0.0001

E. *Testing Process*

The testing process is carried out after the computation process on the training data. The trained model is then implemented using the testing data to obtain the prediction results.

F. *Output Visualization*

The visualization of the GRU model's prediction results is performed to compare the actual gold price movement with the gold price predicted by the GRU model. A line plot is used to represent these two time series. The X-axis of the graph represents time, taken from the *'time'* column in the original dataset, for the testing data segment. The Y-axis represents the gold price in its original scale. The output visualization is presented in Fig. 5 as follows:

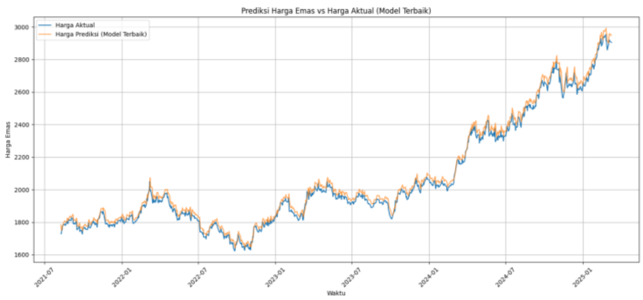


Fig 5. Visualization of Actual vs Predicted Prices

G. *Evaluation*

To measure the performance of the GRU model in predicting XAU/USD prices, the following evaluation metrics will be used:

1. *Mean Absolute Error* (MAE), This metric computes the average of the absolute prediction errors. MAE offers a summary of the mean prediction error expressed in the original price units. The formula for *Mean Absolute Error* (MAE) is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|yi - \hat{y}i| \quad (5)$$

Where:
n is the total number of predictions.
y_i is the actual XAU/USD price at time i.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2352, Copyright ©2025
Submitted : May 1, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

$\hat{y}i$ is the predicted XAU/USD price by the model at time *i*.

MAE is readily interpretable as its outcome is expressed in the same units as the original data. Nevertheless, MAE does not assign greater significance to substantial errors, rendering it less responsive to outliers in comparison to MSE and RMSE. [10]

2. *Mean Squared Error* (MSE), computes the average of the squared prediction errors. MSE exhibits more sensitivity to substantial errors than MAE due to the squaring of errors. The formula for Mean Squared Error (MSE) is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2 \quad (6)$$

Where:
n, $y_i, \hat{y}i$ have the same meanings as in the MAE formula.

MSE penalizes large errors more than MAE, which is beneficial if the model needs to avoid large errors. However, the MSE result is in squared units of the original data, making interpretation more difficult [11]

3. *Root Mean Squared Error* (RMSE) is the square root of *Mean Squared Error* (MSE). RMSE provides the error value in the same units as the original data, facilitating interpretation in contrast to MSE. Furthermore, as it represents the square root of the Mean Squared Error, the *Root Mean Squared Error* (RMSE) is particularly sensitive to significant errors. The formula for RMSE is:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2} \quad (7)$$

Where:
MSE = calculated as in the previous MSE formula.

RMSE penalizes large errors more heavily and is more sensitive to outliers compared to MAE. This metric is commonly used for comparing the performance of different models[10]

III. Discussion And Results

The historical gold data (XAU/USD) obtained from TradingView, spanning from November 21, 2006, to March 9, 2025, serves as the dataset for this study. The original dataset consists of 4782 data points (rows). After performing preprocessing steps as outlined in Section 3 (Methodology), including data normalization using the *Min-Max Scaler* and the formation of time series data sequences with a timestep of 30, the number of data samples ready for training and testing the model is 4752 samples (i.e., the original data minus the lookback).

The GRU model used in this study has a Sequential architecture consisting of four hidden layers, each being a GRU layer with 50, 100, 128, and 256 units. The first GRU layer receives input with dimensions (30, 1), reflecting the 30 timesteps and 1 feature (normalized closing price). The first three GRU layers are configured with *return_sequences=True*, so that each layer produces a full output sequence, which then becomes the input for the subsequent GRU layer. The fourth GRU layer uses return_*sequences=False* because it is the last GRU layer before the output Dense layer. Its output is averaged into a single vector of size 50, representing the temporal features of the 30-step input sequence. The output layer is a single Dense layer that predicts a single price value (the next normalized price).

The model is compiled with the *Adam optimizer*, known for its efficiency in handling sparse or noisy data, and uses Mean Squared Error (MSE) as the loss function, which is commonly used for regression problems. A summary of the model architecture and the number of trainable parameters is shown in the table 2.

TABLE 1. GRU Model Architecture

Layer (type)	Output Shape	Param
gru (GRU)	(None, 30, 50)	7.950
gru_1 (GRU)	(None, 30, 50)	15.300
gru_2 (GRU)	(None, 30, 50)	15.300
gru_3 (GRU)	(None, 50)	15.300
dense (Dense)	(None, 1)	51

To identify the most effective hyperparameter configuration, the GRU model was trained nine times, each with different combinations of *epochs* (50, 100, 150) and *batch sizes* (16, 32, 64). During the training process, the model's performance on a small portion of the training data (set aside as a validation split, accounting for 20% of the training data) was monitored. The training loss and validation loss were recorded at each epoch.

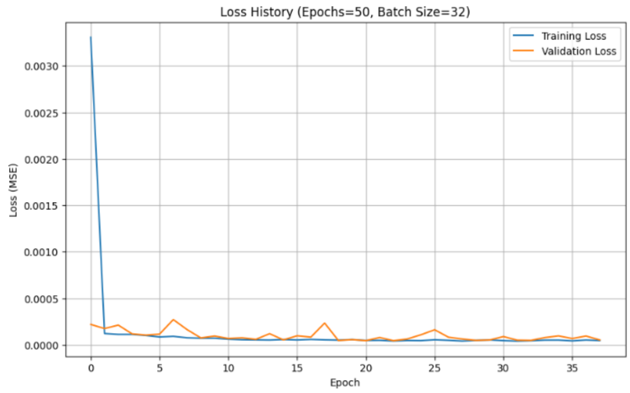


Fig 6. An Example of a Loss History Plot

Fig. 6 presents several examples of *loss history plots* during training. Ideally, both the training loss and validation loss should decrease as the epochs progress, indicating that the model is learning from the data. However, if the validation loss begins to increase while the training loss continues to decrease, this is an indication of overfitting—the model memorizes the training data but loses its ability to generalize to unseen data.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2352, Copyright ©2025
Submitted : May 1, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

In this context, the role of Early Stopping becomes crucial. Early Stopping is a regularization technique that monitors performance metrics (in this case, validation loss) on the validation set. If the monitored metric does not show improvement (or even worsens) over a specified number of consecutive epochs (*patience = 15*), the training process is automatically halted, and the best model weights from the previous epoch are restored. This prevents the model from training too long and overfitting, thus improving its generalization ability on new data (test data). The *"Epochs Completed"* column in Table 3 reflects the actual number of *epochs* completed by the training before Early Stopping was activated, which is often less than the nominal *epoch* count (50, 100, 150) specified.

TABLE 2. 72 hyperparameter Combination Experiment

Epoch	Batch Size	MAE	MSE	RMSE	Training Epochs Completed
50	16	0.006941	0.000089	0.009451	30
50	32	0.007080	0.000096	0.009809	50
50	64	0.008334	0.000134	0.011584	50
100	16	0.008420	0.000140	0.011827	28
…	…	…	…	…	…
100	32	0.008514	0.000143	0.011949	44
100	64	0.007486	0.000109	0.010457	81
150	16	0.008577	0.000148	0.012154	31
150	32	0.007239	0.000101	0.010043	43
150	64	0.006961	0.000091	0.009555	64

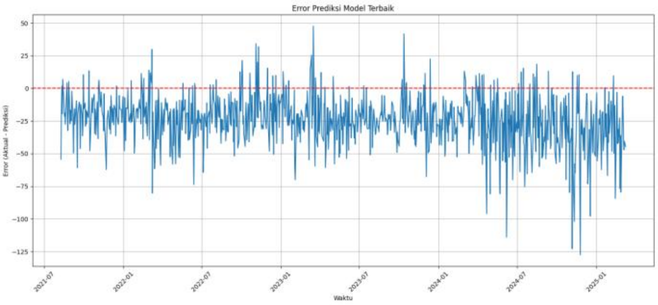


Fig 7. Prediction Model Error Graph

Based on the results from 72 experiments, the *hyperparameter* combination that resulted in the lowest normalized MSE value on the internal validation set (monitored by Early Stopping and confirmed by final evaluation on the test set) was: *Epochs*=100 (although it was stopped earlier by Early Stopping at *epoch* 83), *Batch Size*=16, *Units*=256, and *Dropout Rate*=0.1. The model, retrained with this optimal configuration, was then evaluated on the test dataset. Fig. 7 presents a graph of testing data errors with the best hyperparameter values. After reversing the prediction results to the original price scale through inverse transformation using the same scaler, the following performance metrics were obtained: MAE of 25.761967, MSE of 954.970235, and RMSE of 30.902593. Fig. 8 visualization comparing the actual and predicted prices on the test data demonstrates the ability of this optimal model to capture the general trend of gold price movement.

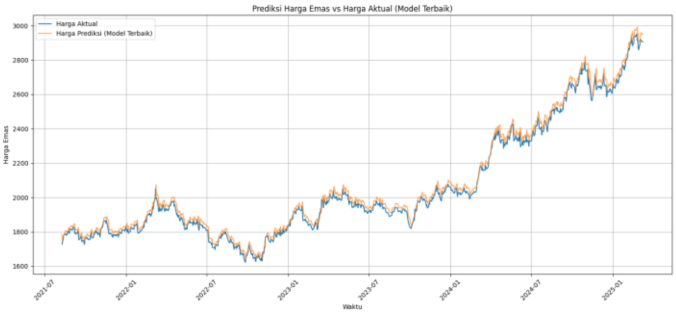


Fig 8. Visualization of Actual vs Predicted Prices

To validate the relative effectiveness of the GRU model, a comparison was made with a simple *Recurrent Neural Network* (RNN) model as the baseline model. Using the MAE metric, the optimized GRU model MAE = 0.017721 (normalized) was compared with the performance of the RNN model. The RNN model produced an MAE of 0.006556 (normalized) on the same test data. This comparison shows that the GRU model demonstrates superior performance compared to RNN, indicating that the gate mechanism in GRU (update gate and reset gate) is more effective in handling long-term dependencies and gradient flow in gold price time series data compared to the simpler RNN architecture. The following is a graph comparing MAE between GRU and RNN in Fig. 9.

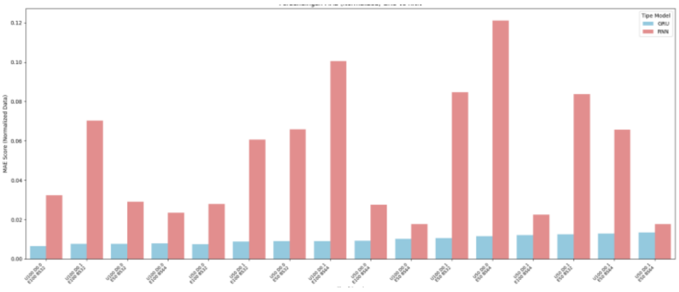


Fig 9. Comparison of MAE GRU and RNN

Although the GRU model shows promising results, the presence of prediction errors indicates that there are factors that cannot be fully modeled. Some potential causes of prediction errors include:

1. *Intrinsic Data Volatility*: Gold prices, like other financial assets, are inherently volatile and influenced by complex and often unpredictable factors. Periods of high volatility, during which prices can fluctuate significantly in a short period of time, pose a particular challenge for time series prediction models. Models may struggle to fully capture sudden spikes or drops in prices if such patterns are not sufficiently represented in the training data.
2. *External Economic Factors*: The models developed in this study only use historical price data as input. However, gold prices are strongly influenced by various external macroeconomic and geopolitical factors that are not explicitly included as features in the models. Factors such as central bank monetary policy (e.g., interest rate changes), exchange rates

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2352, Copyright ©2025
Submitted : May 1, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

(particularly the US Dollar), inflation rates, global political uncertainty, industrial demand, and market sentiment can cause price movements that cannot be predicted based solely on historical data. The absence of these exogenous variables in the model may limit its predictive accuracy, especially when these external factors undergo significant changes.

Overall, these findings suggest that the GRU model is a valid and promising approach for gold price forecasting. However, to enhance the accuracy and robustness of the model in the future, considering the integration of relevant external features and exploring more advanced model architectures or hybrid approaches could be beneficial research directions. A more in-depth analysis of error characteristics, as discussed, is also important to guide further model development iterations.

IV. CONCLUSION

This study successfully implemented and evaluated a *Gated Recurrent Unit* (GRU) model with a four-layer architecture for the task of daily gold closing price prediction. Through a systematic hyperparameter optimization process, testing 72 different combinations of epochs, batch size, units per GRU layer, and dropout rate, it was found that the configuration with 100 nominal epochs (stopped early at epoch 83 by Early Stopping), batch size of 16, 256 units, and dropout rate of 0.1 resulted in the best performance based on the MSE metric on the validation data. Evaluation on the test data showed that this optimal model was able to capture the main patterns and trends in the historical gold price data, achieving an RMSE value of around 30.90 after the scale was reversed. These results indicate that the GRU model, when properly configured and trained using techniques such as Early Stopping to prevent overfitting, is a promising approach for forecasting complex financial time series such as gold prices. However, the presence of prediction errors suggests that further improvements are still possible, such as through the addition of external features or exploration of more advanced model architectures

ACKNOWLEDGMENT

We would like to express our deepest gratitude to the faculty and staff at *Universitas Nahdlatul Ulama Sunan Giri* for their continuous support and guidance throughout this research. Their valuable insights and feedback have been instrumental in the development and completion of this study. We also extend our sincere thanks to TradingView for providing the historical gold price data, which formed the foundation of this research.

REFERENCES

[1] M. Fauzi, M. Jauhar Vikri, and S. Wahyudhi, ‘Sistem PendukungcKeputusan Pemilihan Jenis Investasi MenggunakanxMetode Analytical Hierarchy Process (AHP)’.

[2] A. F. Yuliana and R. Robiyanto, ‘PERAN EMAS SEBAGAI SAFE HAVEN BAGI SAHAM PERTAMBANGAN DI INDONESIA PADA PERIODE PANDEMI COVID-19’, *Jurnal Ilmiah Bisnis dan Ekonomi Asia*, vol. 15, no. 1, pp. 1–11, Feb. 2021, doi: 10.32815/jibeka.v15i1.217.

[3] S. Novianto and H. Akbar Wibowo, ‘The Implementation of Data Mining for Predicting XAU/USD Price Trends in the Forex Market on MetaTrader 5 using Naïve Bayes Method’, *Intelmatics*, vol. 3, no. 2, pp. 85–90, Sep. 2023, doi: 10.25105/itm.v3i2.17199.

[4] R. Mattera, G. Athanasopoulos, and R. Hyndman, ‘Improving out-of-sample forecasts of stock price indexes with forecast reconciliation and clustering’, *Quant Finance*, vol. 24, no. 11, pp. 1641–1667, Nov. 2024, doi: 10.1080/14697688.2024.2412687.

[5] C. Alkahfi, A. Kurnia, and A. Saefuddin, ‘Perbandingan Kinerja Model Berbasis RNN pada Peramalan Data Ekonomi dan Keuangan Indonesia’, *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1235–1243, Jul. 2024, doi: 10.57152/malcom.v4i4.1415.

[6] A. I. Putri, Y. Syarif, N. R. Aisyi, and N. Waeyusoh, ‘Implementation of Gated Recurrent Unit, Long Short-Term Memory and Derivatives for Gold Price Prediction’, vol. 2, no. 2, pp. 68–80, 2025, doi: 10.57152/predatecs.v2i2.1609.

[7] M. F. Julianto, M. Iqbal, W. F. Hidayat, and Y. Malau, ‘PERBANDINGAN PENERAPAN ALGORITMA DEEP LEARNING DALAM PREDIKSI HARGA EMAS’, *INTI Nusa Mandiri*, vol. 19, no. 1, pp. 71–76, Aug. 2024, doi: 10.33480/inti.v19i1.5559.

[8] A. Jamarani, S. Haddadi, R. Sarvizadeh, M. Haghi Kashani, M. Akbari, and S. Moradi, ‘Big data and predictive analytics: A sytematic review of applications’, *Artif Intell Rev*, vol. 57, no. 7, Jul. 2024, doi: 10.1007/s10462-024-10811-5.

[9] R. Chaudhuri, S. Deb, and H. Das, ‘Noble Approach on Sensor Fused Bio Intelligent Path Optimisation and Single Stage Obstacle Recognition in Customized Mobile Agent’, *Procedia Comput Sci*, vol. 218, pp. 778–787, Jan. 2023, doi: 10.1016/J.PROCS.2023.01.058.

[10] David Andrés, ‘Error Metrics for Time Series Forecasting - ML Pills’. Accessed: May 01, 2025. [Online]. Available: https://mlpills.dev/time-series/error-metrics-for-time-series-forecasting/

[11] R. J. Hyndman and A. B. Koehler, ‘Another look at measures of forecast accuracy’, *Int J Forecast*, vol. 22, no. 4, pp. 679–688, Oct. 2006, doi: 10.1016/J.IJFORECAST.2006.03.001.

Sentiment Classification of Public Perception on LHKPN Using SVM and Naive Bayes

Ahmad Rijal Hermawan[1], Isa Faqihuddin Hanif [2]*
Department of Industrial Technology and Informatics, Informatics Engineering[1]
Department of Industrial Technology and Informatics, Information Systems and Technology[2]
University of Muhammadiyah Prof. DR. HAMKA
Jakarta, Indonesia
rijalhermawanahmad@gmail.com[1], isa@uhamka.ac.id[2]

***Abstract*— The public's perception of the State Officials' Wealth Report (LHKPN) serves as a vital measure of confidence in the government's commitment to transparency and efforts to combat corruption.This research seeks to examine public sentiment as reflected on the social media platform X. A dataset comprising 1,200 tweets was gathered and processed through various text mining methods, such as case folding, data cleaning, tokenization, normalization, stemming, stopword elimination, and TF-IDF vectorization. The tweets were then manually annotated into two sentiment categories: positive and negative, with 77.3% of tweets labeled as positive and 22.7% as negative. Sentiment classification was conducted using two machine learning algorithms: Support Vector Machine (SVM) and Naive Bayes. The Naive Bayes algorithm recorded an accuracy of 86.66%, with a precision of 0.93, a recall score of 0.88, and an F1-score of 0.87. Conversely, the SVM model with a linear kernel demonstrated superior performance, achieving an accuracy rate of 93.33%, along with a precision of 0.93, recall of 0.98, and an F1-score of 0.95. To uncover frequently occurring topics, WordCloud visualizations were generated. These revealed that positive tweets often included words such as 'lapor' and 'transparan', while negative ones were more likely to contain terms like 'bohong' and 'korupsi'. These findings indicate that public sentiment toward the LHKPN initiative is largely favorable, despite persistent concerns surrounding integrity and trustworthiness in asset reporting. This study highlights the effectiveness of sentiment analysis in gauging public opinion and informing future policy improvements.**

Keywords*— *Sentiment Analysis; Naive Bayes; Support Vector Machine; LHKPN; Social Media

I. INTRODUCTION

Corruption continues to pose a significant challenge to governance, particularly in developing nations such as Indonesia. In response to this persistent problem, the Indonesian government formed the Corruption Eradication Commission (KPK), granting it extensive powers to conduct investigations and bring corruption cases to justice. [1]. One of the primary transparency measures implemented by the KPK is the State Officials' Wealth Report (Laporan Harta Kekayaan Penyelenggara Negara or LHKPN), which functions as a preventive mechanism to identify inconsistencies in the asset declarations of public officials [2]. By fostering openness in asset disclosure, the LHKPN system enhances accountability and contributes to reinforcing public confidence in governance.

In recent times, the emergence of digital platforms has significantly reshaped the way citizens interact with political matters and policy discussions. Platforms like X (previously known as Twitter) have evolved into arenas where individuals openly share their views regarding governmental transparency and ethical governance. These online dialogues provide a rich source of data that can be leveraged through sentiment analysis to better understand public opinion[3].

Sentiment analysis facilitates the automated categorization of public opinions based on textual data. This method has seen extensive use across multiple fields, including the analysis of political developments, policy assessments, and evaluations of public services. For example,[4]employed Support Vector Machine (SVM) to evaluate public sentiment on the issue of lobster seed exports, achieving an accuracy rate of 84.21%. Similarly, [5] utilized SVM to analyze public responses during the 2019 presidential election, and [3]examined sentiment surrounding the KPK Bill, which predominantly reflected negative opinions. These findings underscore the reliability of machine learning techniques—particularly SVM—in interpreting public sentiment on political matters.

Although sentiment analysis plays a significant role in understanding public discourse, specific investigations into public sentiment regarding the LHKPN remain scarce. This research seeks to fill that gap by conducting an analysis of tweets related to the LHKPN using two machine learning algorithms: Support Vector Machine (SVM) and Naive Bayes. Through performance evaluation and comparison of these models, the study offers empirical evidence on public perceptions of asset disclosure by state officials. The results aim to contribute to enhancing policy communication, promoting greater transparency, and utilizing social media as a strategic instrument to measure public trust.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2341, Copyright ©2025
Submitted : April 25, 2025, Revised : May 3, 2025, Accepted : May 6, 2025, Published : May 26, 2025

II. LITERATURE REVIEW

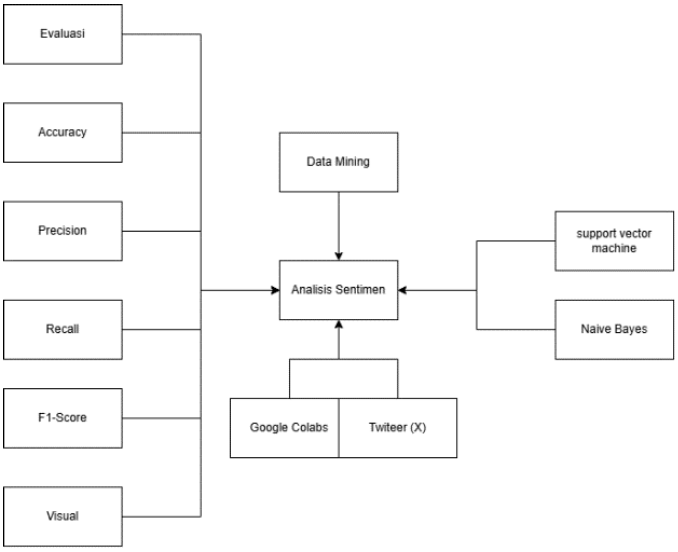


Figure. 1 Conceptual Framework

The conceptual framework outlines the process of data acquisition, preprocessing, sentiment labeling, classification, and evaluation used in this study.

A. Sentiment Analysis

Sentiment analysis is a method within Natural Language Processing (NLP) that focuses on detecting and categorizing the sentiments or emotional tones present in textual data. It is commonly employed to gain insights into public attitudes toward various entities, policies, or occurrences. When applied to social media, sentiment analysis enables the transformation of vast amounts of unstructured data into meaningful and actionable information [6].

B. Naive Bayes Classifier

Naive Naive Bayes is a classification technique grounded in Bayes' Theorem, operating under the strong assumption that each feature—such as individual words in a document—is independent of one another. Despite this assumption, the method has demonstrated high effectiveness in text analysis tasks, valued for its simplicity and strong performance in classification accuracy.[7].

C. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a machine learning technique known for its strong performance in handling high-dimensional datasets, such as textual data. It operates by identifying the optimal hyperplane that effectively separates data points into distinct classes. The advantage of SVM is its ability to handle irregular and complex data efficiently[8]

D. State Official Wealth Report (LHKPN)

The LHKPN is a wealth disclosure system administered by Indonesia's Corruption Eradication Commission (KPK), designed as a mechanism for public oversight to deter corruption. It enables citizens to evaluate the integrity of public officials by examining the transparency of their declared assets.[2]

E. Related Research

Various previous studies have effectively applied sentiment analysis to assess public opinion on governmental and political matters in Indonesia. These studies frequently employed machine learning algorithms such as Support Vector Machine (SVM) and Naive Bayes, recognized for their strong performance in text classification tasks. [4] Several studies have utilized Support Vector Machine (SVM) to classify sentiment on various political and governmental issues in Indonesia. For example, one study applied SVM to analyze public sentiment related to the controversy over lobster seed exports, achieving an accuracy rate of 84.21%, thereby demonstrating SVM's effectiveness in addressing issue-specific sentiment analysis on Indonesian Twitter data. Likewise, [5] implemented SVM to examine sentiments during the 2019 presidential election, attaining an accuracy of 91.5%, further affirming SVM's reliability in mining political opinions.[3]

also used SVM to explore public reactions to the revision of the KPK Law, identifying a predominant negative sentiment (60.9%). Their findings highlighted Twitter's potential as a real-time platform for gauging public responses to controversial legislative changes. In another study, [9] compared the performance of SVM and Naive Bayes in analyzing sentiment regarding the Jakarta gubernatorial election, with results showing that SVM surpassed Naive Bayes in both accuracy and precision.[10]

examined the performance of Naive Bayes Classifier (NBC), K-Nearest Neighbors (KNN), and SVM in evaluating public sentiment toward government performance. Among these algorithms, SVM delivered the highest performance, further validating its robustness in text-based sentiment analysis.While these studies have provided meaningful insights into sentiment analysis using machine learning in the context of elections, public policies, and services, none have specifically investigated public sentiment regarding state wealth transparency through the LHKPN system. This study addresses that gap by applying SVM and Naive Bayes to assess public opinion on LHKPN, thereby contributing a novel perspective. It expands the application of sentiment analysis into the domain of institutional transparency and preventive anti-corruption efforts—an area that remains underrepresented in current research.

III. RESEARCH METHODOLOGY

This study adopts a quantitative methodology by leveraging machine learning algorithms to categorize public sentiment toward the State Officials' Wealth Report (LHKPN) as expressed on the social media platform X. The method ological stages include data collection, preprocessing, annotation, feature extraction, classification, and evaluation.

1) Data Collection

A total of 1,200 tweets related to LHKPN were collected using the X API v2 (formerly Twitter API). Keywords such as "LHKPN", "lapor harta", and "transparansi pejabat" were used to retrieve relevant tweets. Data was gathered using the Tweepy library in Python, executed within the Google Colaboratory environment. In accordance with ethical research standards, all personally identifiable information was either removed or

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2341, Copyright ©2025
Submitted : April 25, 2025, Revised : May 3, 2025, Accepted : May 6, 2025, Published : May 26, 2025

anonymized during the preprocessing stage.

2) *Data Preprocessing*
3) *The tweets underwent multiple preprocessing steps to standardize the data:*
   1. *Case Folding*
      - Converting all characters in the text to lowercase to standardize the input data.
   2. *Cleansing*
      - Removing unnecessary elements such as punctuation marks, special characters, numbers, and hyperlinks to reduce noise in the dataset.
   3. *Tokenizing*
      - The text was segmented into individual words or tokens to enable more detailed analysis at the lexical level.
   4. *Normalization*
      - Converting informal or slang words into their formal equivalents, for example using a dictionary or predefined list.
   5. *Stemming*
      - Words were reduced to their root or base forms using an algorithm such as the Nazief-Adriani stemmer, which is specifically designed for the Indonesian language.
   6. *Stopword Removal*
      - Common words that do not carry significant sentiment, such as "dan", "yang", and "itu", were removed because they are irrelevant for sentiment classification.

F. *Data Annotation*

The dataset was manually labeled into two sentiment categories: positive and negative. Annotation was performed by three independent human annotators with background in data science and communication studies. Each tweet was assessed in terms of its tone, context, and underlying meaning. Any disagreements among annotators were addressed through discussion, and inter-annotator agreement was tracked to maintain consistency in the labeling process. The final distribution was **77.3%** positive (925 tweets) and **22.7%** negative (275 tweets).

Annotation Examples:

- *Positive: "Bagus sekali pejabat ini, sudah melaporkan kekayaannya secara transparan."*
- *Negative: "Pejabat kok tidak lapor harta, pasti ada yang disembunyikan."*

G. *Feature Extraction*

Text data was transformed into numerical features using the Term Frequency–Inverse Document Frequency (TF-IDF) method. This method quantifies a word's relevance within a document in relation to the entire corpus, making it well-suited for high-dimensional text classification tasks. The resulting TF-IDF matrix served as input for the machine learning algorithms.

H. *Data Splitting and Balancing*

The dataset was split into 80% for training and 20% for testing using a stratified method to preserve the distribution of sentiment classes. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was employed on the training data. This technique enhances the representation of the minority class by generating synthetic samples through interpolation between existing instances, thereby improving the model's capacity to learn from the limited negative class data.

I. Classification Algorithms

Two classification models were implemented and compared:

- *Naive Bayes (Multinomial NB): A probabilistic model assuming feature independence, commonly used in text classification due to its simplicity and speed.*
- *Support Vector Machine (SVM): Implemented with a linear kernel and regularization parameter C = 1.0, SVM was selected for its ability to handle high-dimensional and sparse data typical of textual inputs.*

J. Evaluation Metrics

To evaluate the effectiveness of the models, several performance metrics were utilized:

- *Accuracy: Measures the proportion of total predictions that the model got right, reflecting its overall reliability.*
- *Precision: Represents the ratio of correctly predicted positive cases to all instances that were labeled as positive, highlighting the model's ability to avoid false positives.*
- *Recall: Indicates the proportion of actual positive instances that were successfully detected by the model, emphasizing its sensitivity to relevant data.*
- *F1-Score: Combines both precision and recall into a single score using their harmonic mean, making it particularly valuable when dealing with uneven class distributions.*
- *Confusion Matrix: A comprehensive table that categorizes prediction outcomes into true positives, true negatives, false positives, and false negatives. It provides an in-depth view of how the model performs across different classification outcomes, aiding in the identification of specific areas for improvement.*

IV. Results And Discussion

The first step in this research involved collecting tweets related to the State Officials' Wealth Report (LHKPN) using the X API v2. This was conducted using Python and the Tweepy library within Google Colaboratory. A total of 1,200 tweets were gathered based on keywords such as "LHKPN", "lapor harta", and "transparansi pejabat".

```
# Crawl Data

filename = 'Data Skripsi LHKPN Rijal 2.csv'
search_keyword = 'LHKPN KPK'
limit = 1000

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}
```

Figure 2 Crawling Data

Figure 2 shows the Python script implementation used to perform tweet crawling, including authentication process

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2341, Copyright ©2025
Submitted : April 25, 2025, Revised : May 3, 2025, Accepted : May 6, 2025, Published : May 26, 2025

and query formulation.

the raw tweets were reviewed. These included various expressions of public opinion ranging from support for transparency to criticism of corruption.

	Text
921	50 Menteri dan Wakil Menteri Kabinet Merah Putih Prabowo-Gibran Belum Sampaikan LHKPN ke KPK https://t.co/4kFI3SZ2EU
321	@M45Broo_ Gak ada yg ribut terkait LHKPN andika coba kalo dr kim plus koar2 KPK suruh usut
101	@gea_asa @FarhanAtjeh @prastow Sgt bs utk @KPK_RI menjalankan pembuktian terbalik atas LHKPN pejabat2. Hny sj kembali kpd good will KPK itu sndr dan landasan UU/Aturan Pembuktian Terbalik- nya itu ada/blm ada. Kl blm ada aturannya mrk pejabat2 itu bs berk
920	KPK meminta Menteri dan Wakil Menteri Kabinet Merah Putih untuk segera melaporkan harta kekayaan dengan mengisi LHKPN. Batas waktu yang diminta KPK yakni tiga bulan setelah dilantik atau pada Januari 2025. https://t.co/NHtuO4G4o1
58	@intinyadeh LHKPN ini kayaknya lucu deh coba kita cek setiap pejabat berapa aja. Pasti jadi ladang meme banyak banget Lagian KPK ini percaya ga sama beginian? Mayoritas rakyat aja ga percaya eh
790	Tambah berat Ya Min @Pupufafa1 @Gerindra pak @prabowo .. Kau lapor kan itu LHKPN ke KPK Taim Miftah
948	@TOMSETOM_DR berani ga KPK cek LHKPN ne AL ?
969	@dhemit_is_back @KPK_RI @DitjenPajakRI Ini kelemahan lhkpn masih sebatas self asesmen harus ada pembuktian terbalik
410	Polisi harus masuk selidiki dugaan penggunaan jam kw itu. KPK juga harus masuk selidiki jam itu yg konon tak masuk LHKPN
1079	Eng ing eng.. KPK Turun Tangan Dalami Asal Usul Jam Tangan Mewah Milik Dirdik Jampidsus Kejagung Saya lihat dulu ya (LHKPN Qohar) kata Deputi Bidang Pencegahan dan Monitoring @KPK_RI Pahala Nainggolan https://t.co/euJwTp4hmN

Figure 3 Random data crawling results

Figure 3 displays a screenshot of the raw dataset structure, which includes tweet ID, timestamp, tweet text, and user information. All personal identifiers were anonymized for ethical compliance.

The tweets underwent multiple preprocessing steps to clean and standardize the data for sentiment classification.

No	Text Original	Text Processed
1	LHKPN itu cuma formalitas.. ngisinya asal asalan dan KPK nya malas ngecek..	lhkpn formalitas ngisinya asal kpk nya malas ngecek
2	Buntut Kasus Aniaya Dokter Koas KPK Sebut Proses Analisis LHKPN Dedy Mandarsyah Berlangsung 1 Pekan https://t.co/5zCVXyHlct	buntut aniaya dokter koas kpk proses analisis lhkpn dedy mandarsyah pekan
3	Mari kita terus viralkan agar @KPK_RI tak perlu menunggu dua minggu untuk melakukan penyelidikan LHKPN dan penyidik jika telah memiliki dua bukti permulaan.	mari viralkan kpkri tunggu minggu lidi lhkpn sidik milik bukti mula
4	@KPK_RI apakah sudah terima LHKPN bapaknya lady ?	kpkri terima lhkpn bapak lady
5	KPK Ungkap Ada Pejabat Negara Tidak Jujur Isi LHKPN Sindir Jaksa Agung? https://t.co/B7yK4CVLYq	kpk jabat negara jujur isi lhkpn sindir jaksa agung

Figure 4 Data Processing Results

Figure 4 demonstrates a sample of tweets before and after preprocessing, showing the transformation from noisy text to a clean token list ready for vectorization.

Insert a screenshot of the code or interface used in Google Colab that demonstrates the data crawling process using the X API v2. This can be a snippet of Python code with function calls such as tweepy.Client() or search_recent_tweets() along with a visible terminal output showing total tweets collected., conducted via Google Colaboratory to collect tweet data containing keywords related to the LHKPN. This initial step resulted in 1,200 tweets for analysis.

Results tweets were manually labeled into positive and negative sentiments. This involved human annotators assessing the tone and implication of each tweet based on contextual interpretation.

1	Text	label
2	lhkpn formalitas ngisinya asal kpk nya malas ngecek	negatif
3	buntut aniaya dokter koas kpk proses analisis lhkpn dedy mandarsyah pekan	positif
4	mari viralkan kpkri tunggu minggu lidi lhkpn sidik milik bukti mula	positif
5	kpkri terima lhkpn bapak lady	positif
6	kpk jabat negara jujur isi lhkpn sindir jaksa agung	negatif
7	kalo fair lapor lhkpn ga suruh nilai aset input harta lokasi kpk kerjasama surveyor finance nilai aset	positif
8	opposisi m udah aneh sampe puluh m kpk lhkpn ya gak tindak lhkpn rakyat yg ngecek ricek	negatif
9	kpkri ada analisis investigasi thd lhkpn an jokowi gibrantweet nih biar kpkri sehat kuat jumat sepeda kk	positif
10	dhemitisback nih yg d mksd kpk jabat negara yg asaln isi lhkpn lg isitulis	negatif

Figure 5 Manual data labeling results

Figure 5 shows labeled data samples, where tweets are presented with corresponding sentiment labels. This figure is intended to demonstrate the outcome of manual sentiment annotation conducted during the labeling stage. Each tweet is assessed based on its linguistic tone and context, and then assigned either a 'Positive' or 'Negative' sentiment.

In the example shown in Figure 4, two tweets that express support or appreciation for transparency in public official reporting are labeled as Positive, while two tweets that express skepticism or criticism towards the LHKPN process are labeled as Negative. This stage is crucial for ensuring that the supervised learning model is trained with correct and contextually appropriate data.

Each tweet is assigned either a positive or negative sentiment, with a total of 925 positive and 275 negative entries. the data revealed an imbalance with significantly more positive tweets.

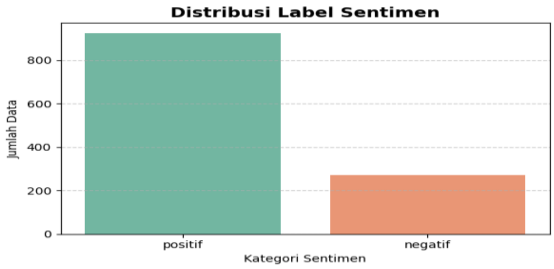


Figure 6 Distribution of Sentiment Data Labeling

Figure 6 presents a bar chart that visualizes the total count of positive and negative sentiments observed in the dataset. This chart serves to give a quick and intuitive overview of the sentiment distribution resulting from the manual annotation process.

As depicted, the positive sentiment category significantly outnumbers the negative one, with 925 tweets classified as positive and 275 as negative. The dominance of positive sentiment reflects the general tone of public discourse on Twitter regarding the LHKPN. Such visual representation supports further analysis and model training by confirming the class imbalance, which is subsequently addressed using techniques like SMOTE.

To prepare the data for machine learning, the tweet texts were transformed into numerical vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) method.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2341, Copyright ©2025
Submitted : April 25, 2025, Revised : May 3, 2025, Accepted : May 6, 2025, Published : May 26, 2025

```
       agung  analisis    aniaya      asal     bapak     bukti    buntut  \
0  0.000000  0.000000  0.000000  0.430786  0.000000  0.000000  0.000000
1  0.000000  0.321926  0.339942  0.000000  0.000000  0.000000  0.333278
2  0.000000  0.000000  0.000000  0.000000  0.000000  0.294949  0.000000
3  0.000000  0.000000  0.000000  0.000000  0.588212  0.000000  0.000000
4  0.441058  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000

      dedy    dokter  formalitas  ...    ngecek  ngisinya       nya     pekan
0  0.00000  0.000000    0.368098  ...  0.403046  0.458526  0.251883  0.000000
1  0.30438  0.339942    0.000000  ...  0.000000  0.000000  0.000000  0.379154
2  0.00000  0.000000    0.000000  ...  0.000000  0.000000  0.000000  0.000000
3  0.00000  0.000000    0.000000  ...  0.000000  0.000000  0.000000  0.000000
4  0.00000  0.000000    0.000000  ...  0.000000  0.000000  0.000000  0.000000

     proses     sidik    sindir    terima    tunggu  viralkan
0  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
1  0.317003  0.000000  0.000000  0.000000  0.000000  0.000000
2  0.000000  0.322952  0.000000  0.000000  0.284954  0.383177
3  0.000000  0.000000  0.000000  0.460069  0.000000  0.000000
4  0.000000  0.000000  0.566548  0.000000  0.000000  0.000000

[5 rows x 37 columns]
```

Figure 7 Text Vaccination Results

Figure 7 illustrates a portion of the TF-IDF matrix, showing token importance across sample documents. This figure is designed to provide insight into how unstructured tweet text is quantitatively represented for machine learning processes.

To convert text data into a numerical format compatible with machine learning algorithms, this research implemented the Term Frequency–Inverse Document Frequency (TF-IDF) technique. This approach assigns a weight to each term based on how often it appears in a particular document (term frequency) and how rare it is across all documents in the dataset (inverse document frequency). Words that are frequently used in an individual tweet but rarely appear in the overall corpus are given greater importance, making them more influential in the classification process.

In contrast to Count Vectorization, the TF-IDF method demonstrates superior capability in identifying significant terms while effectively reducing the influence of frequently occurring, non-informative words. In this study, the preprocessing steps combined with the TF-IDF transformation generated 2,311 distinct features, which served as input for the classification algorithms.

This method was chosen due to its computational efficiency and strong performance in prior sentiment analysis studies. The resulting TF-IDF matrix provided a high-dimensional, sparse representation of tweet content, allowing SVM and Naive Bayes to learn sentiment-related patterns effectively.

was divided into training and testing sets in an 80:20 ratio to evaluate model performance effectively.

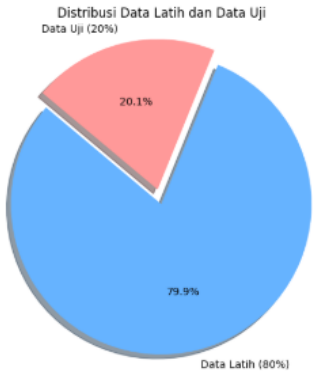


Figure 8 Data splitting results

Figure 8 displays the proportion and count of training and test data points per sentiment category. This figure aims to explain how the dataset was divided to ensure robust evaluation of model performance.

An 80:20 data split is a widely adopted approach in machine learning, enabling the model to train on the bulk of the dataset while assessing its ability to generalize using the remaining unseen portion. The accompanying visualization illustrates how the data is distributed across classes in both sets, providing clarity and transparency during the model preparation stage.

To address the issue of imbalanced data, this study employed the Synthetic Minority Over-sampling Technique (SMOTE) on the training dataset. By generating artificial samples for the underrepresented class—specifically, those labeled with negative sentiment—SMOTE helped balance the dataset. This enhancement enabled the classification models to better capture and recognize sentiment trends across both majority and minority classes, ultimately improving predictive performance.

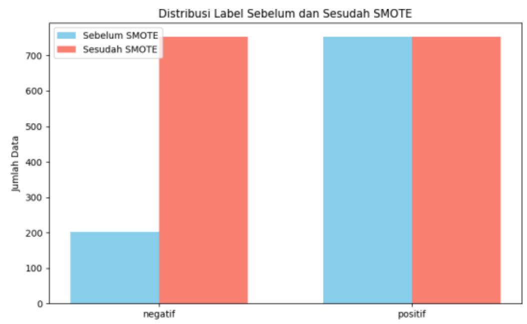


Figure 9 Data Smote Method

Figure 9 provides a visual explanation of the SMOTE process and how synthetic samples are generated. This technique is applied to balance the dataset by oversampling the minority class.

The illustration demonstrates how SMOTE operates conceptually by generating synthetic samples along the lines connecting minority class samples with their nearest neighbors. By introducing these artificial instances, the dataset becomes more balanced, which in turn improves the model's classification accuracy—especially in correctly identifying samples belonging to the minority class, such as those expressing negative sentiment.

results showed that 77.3% of the data were classified as positive sentiment and 22.7% as negative.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2341, Copyright ©2025
Submitted : April 25, 2025, Revised : May 3, 2025, Accepted : May 6, 2025, Published : May 26, 2025

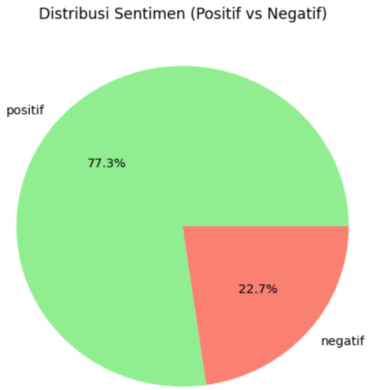


Figure 10 Sentiment Analysis Results

Figure 10 presents the overall sentiment analysis result. This pie chart is a summarization of the sentiment classification, showing the proportion of tweets categorized as positive and negative after the model predictions.

The dominance of positive sentiment in the chart reflects the overall tone of public perception on Twitter toward the LHKPN. The visual also validates the earlier manual annotation distribution and supports the interpretation that the public tends to support transparency initiatives.

TABLE I. MODEL PERFORMANCE EVALUATION REPORT

	Precision	Recall	F1-Score	Support
Negative	0.73	0.84	0.78	69
Positive	0.93	0.88	0.90	171
Accuracy			0.87	240
Macro Avg	0.83	0.86	0.84	240
Weighted Avg	0.87	0.87	0.87	240

TABLE I displays the classification report for the Naive Bayes algorithm, highlighting key performance indicators such as accuracy, precision, recall, and F1-score. This report provides a detailed overview of how well the model performs across various sentiment categories. The presented metrics reflect the model's capability to distinguish between positive and negative sentiments. Accuracy indicates the proportion of correct predictions overall, while precision and recall delve into how effectively the model identifies each sentiment type. The F1-score, which combines precision and recall into a single measure, proves especially useful in scenarios with imbalanced sentiment distributions.

TABLE II. SVM MODEL PERFORMANCE EVALUATION REPORT

	Precision	Recall	F1-Score	Support
Negative	0.95	0.81	0.88	69
Positive	0.93	0.98	0.95	171
Accuracy			0.93	240
Macro Avg	0.94	0.90	0.91	240
Weighted Avg	0.93	0.93	0.93	240

TABLE II Displays the classification report for the SVM model, which demonstrated superior performance compared to Naive Bayes across all evaluation metrics. This figure provides a comprehensive summary of the SVM model's effectiveness in performing the sentiment classification task.

The higher values in precision, recall, and F1-score compared to the Naive Bayes model suggest that SVM is more robust, especially in identifying minority class instances. This supports the selection of SVM as the preferred model for this particular sentiment analysis case.

- *Naive Bayes*: Accuracy 86.66%, precision 0.93, recall 0.88, f1-score 0.87
- *SVM*: Accuracy 93.33%, precision 0.93, recall 0.98, f1-score 0.95

TABLE III. CONFUSION MATRIX NAIVE BAYES

	Actual: Positive	Actual: Negative
Predicted: Positive	TP: 150	FP: 11
Predicted: Negative	FN: 21	TN: 58

Table III The Naive Bayes algorithm attained an accuracy rate of 86.66%, but it exhibited a noticeable tendency to incorrectly classify several positive tweets. According to the confusion matrix, it successfully identified 150 tweets as true positives and 58 as true negatives. Nevertheless, 21 positive tweets were mistakenly categorized as negative (false negatives), and 11 negative tweets were classified as positive (false positives). Despite achieving high precision in detecting positive sentiment, the substantial number of false negatives led to a lower recall, highlighting the model's limitations in accurately capturing negative sentiment.

TABLE IV. CONFUSION MATRIX SVM

	Actual: Positive	Actual: Negative
Predicted: Positive	TP: 168	FP: 13
Predicted: Negative	FN: 3	TN: 56

On the other hand, the Support Vector Machine (SVM) model delivered better overall results, reaching an accuracy of 93.33%. As reflected in the confusion matrix, the model accurately recognized 168 positive tweets (true positives) and 56 negative tweets (true negatives), with only 3 positive tweets misclassified as negative (false negatives) and 13 negative tweets incorrectly labeled as positive (false positives). The minimal false negative rate suggests that the model possesses a high recall, indicating its effectiveness in identifying genuine positive sentiments. While a few misclassifications occurred in the negative class, the results overall affirm that SVM is more consistent and dependable than Naive Bayes for sentiment classification within this dataset.

Figure 11 Positive Sentiment WordCloud

Figure 15 illustrates the WordCloud for positive sentiment

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2341, Copyright ©2025
Submitted : April 25, 2025, Revised : May 3, 2025, Accepted : May 6, 2025, Published : May 26, 2025

tweets. This visualization highlights the most frequently occurring keywords in tweets labeled as positive.

The larger the word appears in the WordCloud, the more frequently it occurs in the dataset. Words such as "lapor," "transparan," and "harta" are indicative of public approval, often linked with praise for transparent reporting practices or admiration for public figures who disclose their wealth properly. The WordCloud serves as a visual summary of public expressions of trust and appreciation toward the LHKPN process. Prominent keywords consist of terms like “lapor,” “transparan,” and “harta,” along with mentions of notable individuals such as ministers or celebrities known for advocating transparency.

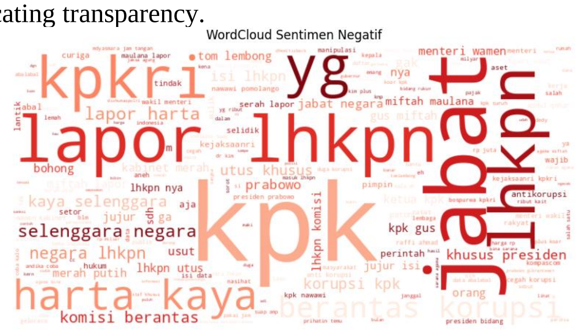


Figure 12 Negatif Sentiment WordCloud

Figure 16 shows the WordCloud for negative sentiment tweets. This figure emphasizes the dominant words used in tweets expressing dissatisfaction, skepticism, or criticism toward the LHKPN.

Prominent terms like "bohong," "korupsi," and "tidaklapor" suggest public concerns about dishonesty and lack of compliance by some officials. The negative WordCloud helps identify key issues in public discourse and reflects areas where transparency efforts may still be perceived as insufficient. Keywords like “bohong,” “korupsi,” and “tidaklapor” reflect public skepticism regarding the integrity of certain officials in disclosing their assets.

Dominant keywords:

- *Positive*: "lapor" (report), "transparan" (transparent)
- *Negative*: "bohong" (lie), "korupsi" (corruption)

SVM shown better performance due to its capability to handle high-dimensional data with complex distributions. These findings are consistent with previous studies.

V. CONCLUSION

This research effectively categorized public sentiment regarding the State Officials’ Wealth Report (LHKPN) shared on the social media platform X by utilizing Support Vector Machine (SVM) and Naive Bayes algorithms. A total of 1,200 tweets were gathered and manually annotated as either positive or negative, with 77.3% identified as positive and 22.7% as negative. The sentiment analysis process involved thorough text preprocessing and TF-IDF feature extraction, followed by an 80:20 split between training and testing datasets.

The evaluation revealed that the SVM model surpassed Naive Bayes across all perforsmance indicators, achieving 93.33% accuracy, 0.93 precision, 0.98 recall, and an F1-score of 0.95. Although Naive Bayes proved to be a fast and straightforward approach, its accuracy reached only 86.66%, and it struggled with identifying negative sentiment effectively. The analysis was further reinforced by WordCloud visualizations, which highlighted commonly used terms in each sentiment category—such as lapor (report), transparan (transparent), and jujur (honest) in positive tweets, and bohong (lie), korupsi (corruption), and tidaklapor (not reporting) in negative ones.

The findings indicate that public perception of the LHKPN initiative is largely positive, reflecting strong support for transparent disclosure of state officials’ assets. These insights can provide valuable guidance for policymakers, particularly the Corruption Eradication Commission (KPK), in evaluating public confidence in asset reporting systems and enhancing their outreach and communication efforts. Moreover, the existence of critical viewpoints underscores the public’s call for stronger verification mechanisms and stricter enforcement of asset declaration policies.

Looking ahead, this research is limited to binary sentiment classification using traditional machine learning algorithms. Future research is recommended to investigate more advanced approaches, such as deep learning techniques like Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT), which are capable of capturing deeper semantic meaning and contextual nuances in sentiment analysis. Additionally, incorporating multi-class sentiment classification (e.g., positive, negative, neutral, sarcastic), topic modeling, and temporal trend tracking could offer richer insights into public opinion and deliver more adaptive, real-time feedback to support transparency initiatives in governance.

ACKNOWLEDGMENT

The author gratefully acknowledges the blessings and guidance of Allah SWT, which have been instrumental in the successful completion of this research. Heartfelt appreciation is extended to the academic advisor for their insightful guidance, valuable suggestions, and consistent encouragement throughout the research journey. The author also wishes to thank family and friends for their unwavering support and motivation. Gratitude is further extended to all individuals and organizations involved in the data gathering process and in facilitating this study. It is hoped that the outcomes of this research will make a meaningful contribution to the fields of sentiment analysis and studies on public transparency.

REFERENCES

[1] R. Y. Oly Viana Agustine, Erlina Maria Christin Sinaga, “Politik Hukum Penguatan Kewenangan Komisi Pemberantasan Korupsi dalam Sistem Ketatanegaraan Legal Politics of the Strengthening of Authority in the Constitutional System,” *Konstitusi*, vol. 16, no. 2, pp. 314–338, 2019.

[2] U. M. Sosiawan, “Peran Komisi Pemberantasan Korupsi (KPK) Dalam Pencegahan dan Pemberantasan Korupsi,” *J. Penelit. Huk. Jure*, vol. 19, no. 4, p. 517, 2019, doi: 10.30641/dejure.2019.v19.517-538.

[3] R. Nooraeni, H. D. Sariyanti, A. F. F. Iskandar, S. F. Munawwaroh, S. Pertiwi, and Y. Ronaldias, “Analisis Sentimen Data Twitter Mengenai Isu RUU KPK Dengan Metode Support Vector Machine (SVM),” *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 1, pp. 55–60, 2020, doi:

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2341, Copyright ©2025
Submitted : April 25, 2025, Revised : May 3, 2025, Accepted : May 6, 2025, Published : May 26, 2025

10.31294/p.v22i1.6869.

[4] B. Pamungkas, A. Syaifuddin, and M. Muslimin, "Analisis Sentimen Twitter MenggunakanMetode Support Vector Machine (SVM) padaKasus Benih Lobster 2020," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 3, no. 2, pp. 10–20, 2021.

[5] O. Zoellanda A.Tane, K. Muslim Lhaksmana, and F. Nhita, "Analisis Sentimen pada Twitter Tentang Calon Presiden 2019 Menggunakan Metode SVM (Support Vector Machine)," *eProceedings Eng.*, vol. 6, no. 2, pp. 9716–9725, 2019.

[6] B. W. Sari and F. F. Haranto, "Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, 2019, doi: 10.33480/pilar.v15i2.699.

[7] A. Rosadi *et al.*, "Analisis Sentimen Berdasarkanan Opini Pengguna pada Media Twitter Terhadap BPJS Menggunakan Metode Lexicon Based dan Naïve Bayes Classi er Twitter Text Mining," vol. 20, pp. 39–52, 2021.

[8] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.

[9] E. Putra Nuansa, "Analisis Sentimen Pengguna Twitter Terhadap Pemilihan Gubernur Dki Jakarta Dengan Metode Naïve Bayesian Classification Dan Support Vector Machine," *Inst. Teknol. Sepuluh Nop. Surabaya*, pp. 1–101, 2017.

[10] P. Simposium, N. Multidisiplin, and U. M. Tangerang, "Analisis Sentimen Kinerja Pemerintahan Menggunakan Algoritma," vol. 4, pp. 114–121, 2022.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2341, Copyright ©2025
Submitted : April 25, 2025, Revised : May 3, 2025, Accepted : May 6, 2025, Published : May 26, 2025

The Effect of SMOTE and Optuna Hyperparameter Optimization on TabNet Performance for Heart Disease Classification

Danang Wijayanto[1]*, Robert Marco[2], Acihmah Sidauruk[3], Mulia Sulistiyono[4]
Department of Computer Science [1], [2], [3],[4]
University of AMIKOM
Yogyakarta, Indonesia
danangwijayanto507@gmail.com[1], robertmarco@amikom.ac.id[2], acihmah@amikom.ac.id[3] ,
muliasulistiyono@amikom.ac.id[4]

***Abstract*— Heart disease is a medical condition affecting the cardiovascular system, disrupting blood circulation and reducing cardiac function efficiency, which can lead to severe health complications. Early diagnosis of heart disease has become increasingly crucial as delayed detection can significantly impact patient outcomes and survival rates. While numerous studies have explored various approaches for heart disease classification, challenges related to data imbalance and improper parameter settings remain persistent issues that affect model performance. This research evaluated the effectiveness of combining TabNet with SMOTE and optuna hyperparameter optimization for heart disease classification. We conducted four experimental scenarios using a heart disease dataset with 303 instances: baseline TabNet, baseline TabNet with SMOTE, TabNet with Optuna, and TabNet with both SMOTE and Optuna. Results demonstrated that applying SMOTE alone to TabNet decreased model performance (accuracy from 85.24% to 77.04%, AUC from 0.89 to 0.83). However, when combining SMOTE with Optuna hyperparameter optimization, we achieved optimal performance with 90.16% accuracy, 93.33% precision, 87.50% recall, 90.32% F1-score, and 0.93 AUC. This represented a significant improvement over other configurations and several previous classification approaches. The integration of SMOTE with Optuna optimization provided an effective framework for heart disease classification that outperformed traditional methods particularly in discriminative capability as evidenced by the superior AUC score.**

***Keywords*— TabNet, SMOTE, Optuna, Classification, Heart Disease**

I. INTRODUCTION

Global cardiovascular disease (CVD) data from 2015 recorded 422.7 million cases and 17.92 million deaths. Ischemic heart disease emerged as the primary cause of CVD-related health loss worldwide, with stroke being the second most common cause. While high-income and some middle-income countries showed declining age-standardized CVD death rates between 1990-2015, mortality patterns shifted from women to men in regions with higher social development indices [1].

Within the context of heart disease research, Artificial Intelligence technology made substantial advances across multiple domains, especially in healthcare applications. As a specialized field within Artificial Intelligence, machine learning showed remarkable utility in various health-related cases, particularly in heart disease classification systems. Machine learning encompassed the development of computer systems that could autonomously improve their capabilities through accumulated experience [2].

Numerous previous studies extensively explored heart disease classification utilizing machine learning and deep learning methodologies. Nevertheless, challenges pertaining to data imbalance persisted, and the performance of machine learning models, especially deep learning architectures, depended heavily on appropriate hyperparameter configurations, which were difficult to determine manually [3].

Research conducted by Yogianto et al. [4] demonstrated the implementation of the K-Nearest Neighbors (KNN) algorithm in heart disease classification, yielding an accuracy rate of 64.03%. Masruriyah et al. [5] employed the SMOTE technique to address class imbalance issues and conducted a comparative analysis of multiple algorithms, producing varying classification accuracies: C4.5 achieved 70%, Random Forest 87%, K-Nearest Neighbors 86%, and Logistic Regression 73%.

The TabNet architecture, introduced by Arik and Pfister [6], offered several theoretical advantages for heart disease classification that addressed the limitations of previous approaches. Unlike conventional neural networks that processed all features simultaneously, TabNet employed sequential attention mechanisms that systematically identified and prioritized significant features throughout each decision-making phase. This approach was particularly suited to medical diagnostics, where certain features carried varying importance for different patient profiles.

To address the identified research gaps, this study proposed a comprehensive approach that combined TabNet with SMOTE for handling class imbalance and Optuna for hyperparameter optimization. This integration specifically targeted the dual challenges that limited previous heart disease classification models: data imbalance and suboptimal parameter selection

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2348, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

[7],[8],[9]. By systematically evaluating different combinations of these techniques, we aimed to determine their individual and combined effects on classification performance.

II. METHODOLOGY

In this study, we proposed a heart disease classification methodology using TabNet model, as illustrated in Figure 1. Our approach aimed to evaluate different combinatios of techniques using TabNet model, which offered excellent interpretability capabilities and efficiently handled tabular data.

The methodology followed a systematic workflow comprising four main stages: heart disease dataset, pre-processing, model implementation with experimental scenarios, and evaluation. We evaluated our approach using multiple performance metrics including accuracy, precision, recall, F1-score, and area under the ROC curve, and compared the results across different experimental scenarios to determine the individual and combined effects of SMOTE and Optuna Optimization.

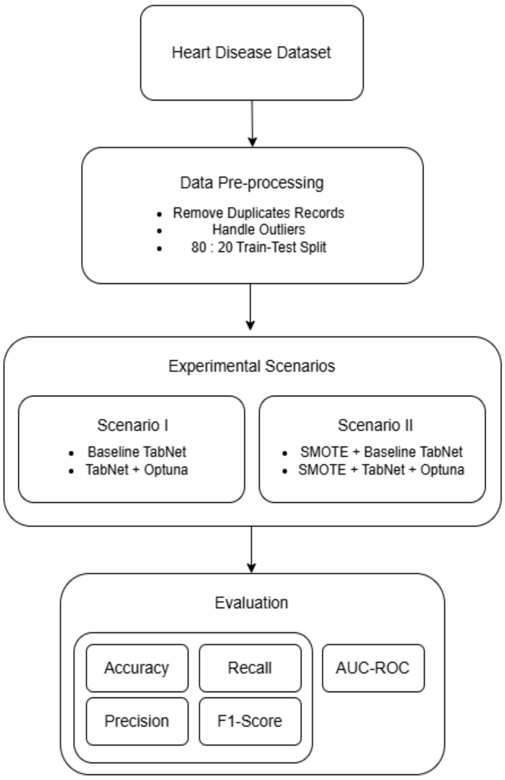


Figure 1. Methodology Research

A. Heart Disease Dataset

The dataset used in this research was obtained from a public dataset avalilable on Kaggle.com (https://www.kaggle.com/datasets/yasserh/heart-disease-dataset) which is derived from the UCI Heart Disease dataset, a widely used benchmark in medical classification research. The dataset contained a total of 303 data points comprising 13 features and 1 target variable. The target variable contained two values: 1 indicating the presence of heart disease and 0 indicating normal condition. The detailed description of each feature and data type is presented in Table I.

TABLE I. DATASET DESCRIPTION

No	Feature Name	Description	Data Type
1	Age	Patient's age in years	Numeric
2	Sex	Gender of patient (1:male; 2:Female)	Categorial
3	Cp	Type of chest pain experienced (0: asymptomatic, 1: atypical angina, 2 : non-anginal pain, 3: typical angina	Numeric
4	Tresbps	Resting blood pressure in mmHg	Numeric
5	Chol	Serum cholesterol level in mg/dl	Numeric
6	Fbs	Fasting blood sugar > 120 mg/dl (1:true, 0:false)	Categorial
7	Restecg	Resting electrocardiogram	Numeric
8	Thalach	Maximum heart rate achieved	Numeric
9	Exang	Exercise-induced angina (1:yes, 0:no)	Categorial
10	Oldpeak	ST depression induced by exercise relative to rest	Numeric
11	Slope	Slope of peak exercise ST segment	Numeric
12	Ca	Number of major vessels colored by fluoroscopy	Numeric
13	Thal	Thalassemia type	Numeric
14	Target	Heart disease diagnosis (1: heart disease, 0:normal)	Categorial

B. Pre-Processing

The preprocessing stage consisted of several steps to ensure the quality of data used in this research. Handling duplicate data aimed to identify and manage duplicated data entries. This was crucial for improving data quality before use and reducing false positives in the results [10].

Handling outliers focused on detecting and managing values that fall significantly outside the dataset's normal range. These anomalous values can substantially bias statistical calculations, particularly affecting mean values through under or overestimation. Thus, addressing outliers through modification or value substitution was essential before conducting data analysis [11]. We split the dataset into training and testing sets with an 80:20 ratio, which was a standard practice widely adopted in previous heart disease classification studies [7], [9], [19].

C. Experimental Scenarios

This research divided the experiments into two main scenarios to obtain more comprehensive evaluation results and enable more detailed comparative analysis. The details of both experimental scenarios are presented in Table II.

TABLE II. SCENARIOS

Scenario	Method
I	TabNet
	TabNet + Optuna
II	SMOTE + TabNet
	SMOTE + TabNet + Optuna

In both scenarios, we implemented two variants of TabNet: a baseline TabNet model and an optimized TabNet model using

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2348, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Optuna for hyperparameter tuning. The first scenario used standard data splitting, while the second scenario incorporated SMOTE to address class imbalance issues. Sampling techniques such as SMOTE were only applied to the training dataset, not to validation or test sets, ensuring model evaluation occurred on data distributions that truly represented the actual problem domain, thus avoiding bias in performance assessment [12].

D. TabNet

TabNet is a deep learning algorithm specifically designed to process tabular data by combining sequential attention and neural networks concepts. TabNet employed sequential attention to select feature subsets, enabling efficient learning of the most prominent features, and its architecture consisted of sequential multi-step processing, where each step contributed to the decision based on selected features [6].

According to paper [6], TabNet architecture consisted of several main components: feature transformer that converted input features into more meaningful representations, attentive transformer that determined feature masks for each decision step, and feature masking that implemented sparse feature selection. At each decision step, TabNet used a learnable mask to select the most important features with the formula:

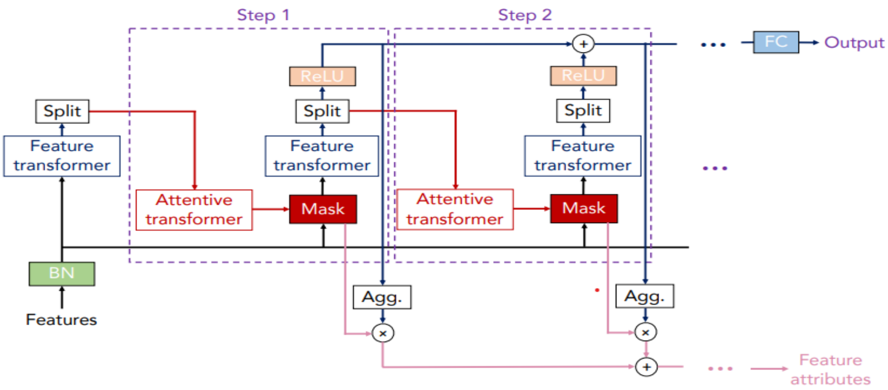


Figure 2. TabNet Architecture[6]

$$M[i] \cdot f \tag{1}$$

Where :
M[i] = Mask for step i
f = Input features

This mask was obtained using an attentive transformer with the formula:

$$M[i] = sparsemax(P[i-1] \cdot hi(a[i-1])) \tag{2}$$

Where:
M[i] = mask for step i
hi = trainable transformation function
a[i-1] = processed features from the previous step sparsemax = normalization that produces sparse weights

P[i] is the prior scale term indicating how much a feature has been previously used:

$$P[i] = \sum_{j=1}^{i} (\gamma - M[j]) \tag{3}$$

Where:
P[i] = prior scale at step i
γ = relaxation parameter (γ ≥ 1)
M[j] = mask from previous steps
P[0] = initialized as a 1B×D matrix

After features are selected, TabNet uses a feature transformer to process these features. The output of this process was divided into two parts:

$$[d[i], a[i]] = fi(M[i] \cdot f) \tag{4}$$

Where:
d[i] = output for the current decision step
a[i] = output information to be used in the next step
fi = the feature transformer function
M[i] = mask for step i
f = input feature

d[i] was the decision step output and a[i] was the information for the next step. To produce the final decision, TabNet aggregated the output from all decision steps using the formula:

$$dout = \sum_i ReLU(d[i]) \tag{5}$$

Where:
Dout = final decision output
ReLU = rectified Linear Unit activation function
d[i] = output from decision step i
Σi = summation over all steps

For interpretability, TabNet used an aggregate feature importance mask that was calculated by:

$$Magg - b, j = \sum_i \eta b[i] Mb, j[i] \,/\, normalization \tag{6}$$

Where:
Magg_b,j := Aggregated importance mask

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2348, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

nb[i] = Feature importance score at step i
Mb,j[i] = Mask for batch b and feature j at step i
Normalization= Normalization factor for value standardization

The contribution score ηb[i] was determined by:

$$\eta b[i] = \sum c \; ReLU(db, c[i]) \quad (7)$$

Where:
nb[i] = Feature importance score at step i
Σc = Summation over all classes
db,c[i] = Decision output for batch b and class c at step i
ReLU = Rectified Linear Unit activation function

E. Optuna

Optuna was a hyperparameter optimization framework developed by Akiba et al. in 2019. Optuna was designed with a "define-by-run" principle that allowed users to dynamically construct parameter search spaces, offering efficient implementation of search and pruning strategies, a flexible and versatile architecture for various purposes, and equipped with Tree-structured Parzen Estimators (TPE) in its optimization process which was useful for learning from previous optimization trials [13].

Through Optuna optimization, The parameter ranges were determined through preliminary experiments. We deliberately selected narrower, more focused ranges rather than broader exploration to maximize optimization efficiency given computational constraints are detailed in Table III.

TABLE III. TABNET OPTIMIZATION PARAMETERS

Parameter	Range Value
n_d	8 - 32
n_a	8 – 32
n_steps	5 – 8
n_independent	1 - 2
learning_rate	0.01 – 0.1
gamma	1.0 – 2.0
lambda_sparse	0.0001 – 0.01

F. Evaluation

In this study, the model evaluation was conducted using confusion matrix and AUC-ROC curve analysis. confusion matrix, as described by [14], was a fundamental evaluation tool in machine learning that displayed the relationship between predicted and actual classifications. It utilized a two-dimensional structure where one axis represented the true class labels while the other showed the model's predictions.

TABLE IV. CONFUSION MATRIX

		Actual	
		Positive	**Negative**
Prediction	**Positive**	TP	FP
	Negative	FN	TN

The structure of the binary classification confusion matrix implemented in this study consisted of four key components.

- True Positive (TP) : represented the number of correctly classified positive instances.
- True Negative (TN) : indicated the number of correctly classified negative instances.
- False Positive (FP) : also known as Type I error, represented negative instances incorrectly classified as positive
- False Negative (FN) : Type II error, indicated positive instances incorrectly classified as negative.

These components and their relationships are illustrated in Table IV. From the confusion matrix components, several key performance indicators can be calculated to evaluate the model's performance, including accuracy, precision, recall, and F1-score [15]. These evaluation metrics are calculated using the following formulas :

$$Accuracy = \frac{(\mathrm{TP+TN})}{(\mathrm{TP+FP+FN+TN})} \quad (8)$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (9)$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (10)$$

$$F1-score = \frac{2*\mathrm{presisi}*\mathrm{recall}}{\mathrm{presisi+recall}} \quad (11)$$

In addition to the confusion matrix, this study also utilized the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), a popular evaluation metric used to measure binary classification performance [16]. AUC-ROC was employed to analyze model performance in greater depth, particularly in identifying areas where the model struggled to separate positive and negative labels, which ultimately helped identify the classifier's decision boundary and potential AUC improvements.

III. RESULT AND DISCUSSION

In this study, we used a heart disease dataset containing 303 records with 13 features and 1 target variable. The features included patient characteristics and medical measurements such as age, sex, chest pain type, blood pressure, cholesterol, and other cardiac indicators, as shown in Figure 3. These features were selected based on their established clinical relevance to cardiac health assessment and diagnostic procedures in medical literature. The preprocessing phase began with duplicate detection, where one duplicate record was identified and removed, reducing the dataset to 302 records. This elimination of duplicates was an essential step to ensure the integrity of our analysis and prevent potential bias in model training and evaluation outcomes.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2348, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

Figure 3. Heart Disease Dataset

In the next step, we continued with outlier detection for the continuous features, which identified several outliers as shown in Figure 4. Considering the small dataset size, these outliers were replaced with their respective upper and lower bounds for each feature.

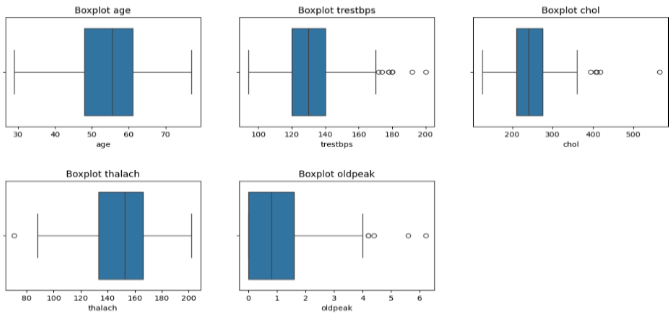


Figure 4. Outliers Data

After preprocessing, the dataset was duplicated for use in two scenarios. The data was split into training and testing sets with an 80:20 ratio. The data distribution was shown in Table V. For scenario II, SMOTE was applied to the training data to achieve balanced class distribution during model training to investigate potential performance improvements.

TABLE V. DISTRIBUTION OF HEART DISEASE DATASET

Scenario	Dataset	
	Training	**Testing**
I	241	61
II	264	61

The evaluation of TabNet performance on heart disease classification was conducted through two scenarios. Each scenario examined two model variations: baseline TabNet and TabNet with Optuna hyperparameter optimization. For both variations, we established initial parameters including a patience value of 20 for early stopping when no learning improvement was observed, batch and virtual_batch sizes of 32 to accommodate the small dataset, and n_trials of 20 in Optuna for the number of optimization attempts.

Figure 5 presents the training loss curves for all four experimental configurations across both scenarios. Statistical analysis was conducted on the final 20 epochs, chosen because at this stage all models had surpassed their initial rapid learning phase and entered more stable convergence patterns. providing a more reliable representation of each model's final learning characteristics. This analysis revealed significant differences in convergence patterns among the models.

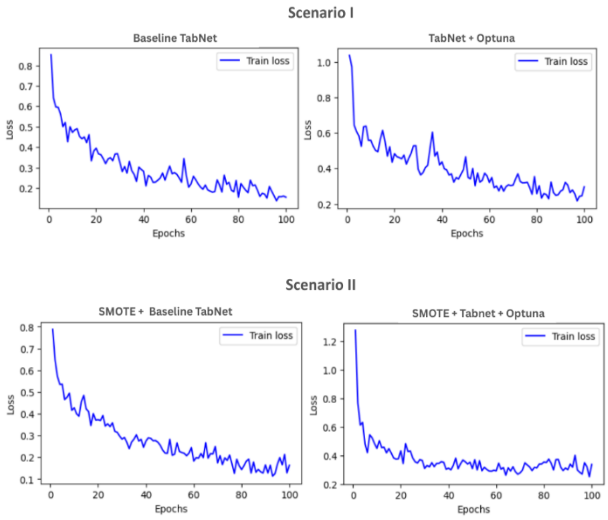


Figure. 5 Training Loss Over Epoch Scenario I and II

For Scenario I, the baseline TabNet demonstrated efficient early learning with a rapid decrease in training loss from 0.8 to 0.4 within the first 10 epochs, eventually reaching stable convergence around 0.2. When Optuna optimization was applied, the model showed higher initial loss (~1.0) but stabilized around 0.3 after epoch 60. Statistical comparison between these models revealed a significant difference in training behavior ($t = -9.74$, $p < 0.0001$), with baseline TabNet consistently maintaining lower loss values across the final 20 epochs. The variance analysis showed TabNet+Optuna exhibited slightly higher fluctuations ($\sigma^2 = 0.000863$) compared to baseline TabNet ($\sigma^2 = 0.000722$), supporting our observation that the optimized model explored a more diverse feature space.

For Scenario II with SMOTE application, the baseline TabNet achieved smoother convergence to approximately 0.15 by epoch 90. When SMOTE was combined with Optuna optimization, the model began with higher initial loss (~1.2) but stabilized between 0.3-0.4 after epoch 40. Statistical analysis of the final 20 epochs revealed an extremely significant difference between these models ($t = -15.83$, $p < 0.0001$), indicating SMOTE+TabNet consistently maintained lower training loss compared to SMOTE+TabNet+Optuna. The variance in SMOTE+TabNet+Optuna ($\sigma^2 = 0.001421$) was notably higher than SMOTE+TabNet ($\sigma^2 = 0.000757$), suggesting that Optuna optimization introduced beneficial regularization effects that prevented the model from minimizing training loss too aggressively.

Cross-scenario comparison revealed significant differences between baseline and SMOTE implementations ($t = 2.67$, $p = 0.015$), with SMOTE consistently resulting in lower training loss. Similarly, comparison between both Optuna-optimized models showed significant differences ($t = -5.93$, $p < 0.0001$), with SMOTE+TabNet+Optuna maintaining higher training loss. These statistical findings provided strong evidence that while SMOTE facilitated easier optimization of the loss function during training, Optuna's hyperparameter optimization introduced effective regularization effects that prevented overfitting to synthetic samples, explaining the superior test performance observed in subsequent evaluations despite higher training loss.

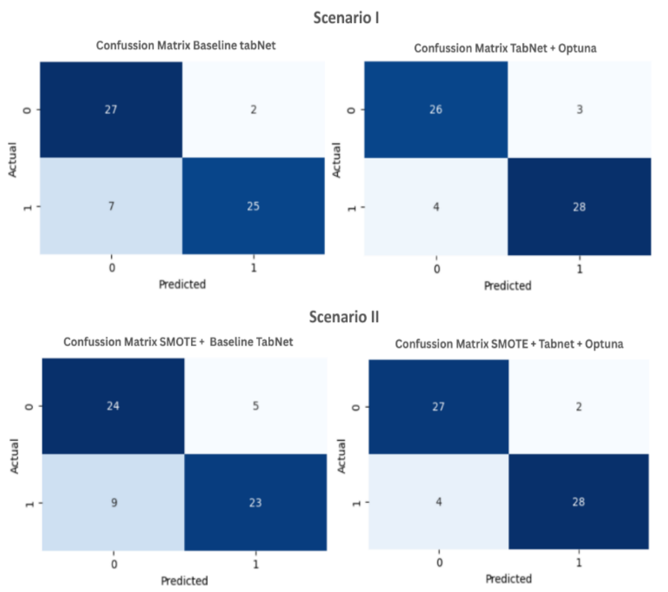


Figure 6. Confusion Matrix Scenario I and II

Figure 6 presents the confusion matrices for all four model configurations, providing a detailed view of classification performance. In Scenario I, the baseline TabNet correctly identified 27 negative cases (true negatives) and 25 positive cases (true positives), while misclassifying 2 negative cases as positive (false positives) and 7 positive cases as negative (false negatives). When Optuna optimization was applied to TabNet in Scenario I, the model demonstrated improved performance with 26 true negatives and 28 true positives. The optimization reduced misclassifications to 3 false positives and 4 false negatives.

In Scenario II, the application of SMOTE alone to the baseline TabNet unexpectedly decreased performance, with the model achieving 24 true negatives and 23 true positives, while showing increased misclassification rates with 5 false positives and 9 false negatives. This performance degradation could be attributed to several factors. First, the synthetic samples generated by SMOTE might have introduced noise in the feature space rather than meaningful patterns, given the relatively small original dataset size. Second, the TabNet's default parameters might not have been optimal for learning from the modified data distribution, causing the model to overfit to synthetic patterns that did not generalize well to the test set.

However, when SMOTE was combined with Optuna optimization, the model achieved the best overall performance with 27 true negatives and 28 true positives, while reducing misclassifications to 2 false positives and 4 false negatives. This demonstrated that while SMOTE alone might disrupt the original data distribution, Optuna's hyperparameter optimization effectively mitigated this issue by adapting the model architecture specifically to the characteristics of the balanced dataset.

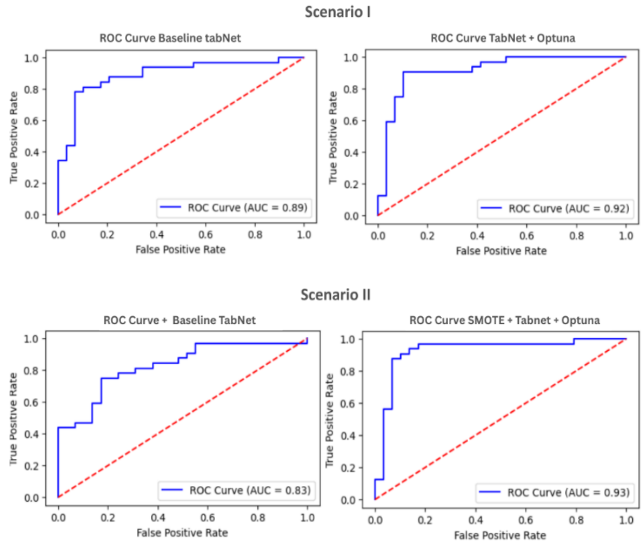


Figure 7. AUC-ROC for Scenario I and II

Figure 7 presents the ROC curves for all four model configurations, providing insights into their discriminative capabilities across different classification thresholds. In Scenario I, the baseline TabNet achieved an AUC score of 0.89, indicating strong overall classification ability. The Optuna-optimized version showed improved performance with an AUC of 0.92, demonstrated by a curve that rose more sharply at low false positive rates and maintained higher true positive rates throughout the threshold spectrum.

In Scenario II with SMOTE implementation, the baseline model showed a decreased performance with an AUC of 0.83, further confirming that class balancing alone negatively affected the model's discriminative ability. However, when combined with Optuna optimization, the model achieved the highest AUC of 0.93, characterized by a steep initial rise and consistently high true positive rates across different false positive rate thresholds.

TABLE VI. COMPARISON OF TABNET MODEL PERFORMANCE

Scenario	Method	Evaluation				
		Accuracy	Presisi	Recall	F1-score	AUC
I	TabNet	85.24%	92.59%	78.12%	84.74%	0,89
	TabNet + Optuna	88.52%	90.32%	87.50%	88.88%	0,92
II	SMOTE + TabNet	77.04%	82.14%	71.87%	76.66%	0,83
	SMOTE + tabnet + Optuna	90.16%	93.33%	87.50%	90.32%	0,93

The experimental results presented in Table VI show the evaluation metrics for all model variations across both scenarios. In Scenario I, the baseline TabNet achieved good performance with 85.24% accuracy, 92.59% precision, 78.12% recall, 84.74% F1-score, and 0.89 AUC. The Optuna-optimized

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2348, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

version showed improvement across all metrics, reaching 88.52% accuracy, 90.32% precision, 87.50% recall, 88.88% F1-score, and 0.92 AUC.

In Scenario II, the SMOTE-enhanced baseline TabNet initially showed decreased performance with 77.04% accuracy, 82.14% prerecision, 71.87% recall, 76.66% F1-score, and 0.83 AUC. However, when combined with Optuna optimization, the model achieved the best overall performance with 90.16% accuracy, 93.33% precision, 87.50% recall, 90.32% F1-score, and 0.93 AUC.

The results demonstrate that while SMOTE alone may reduce model performance, its combination with Optuna optimization leads to superior results across all evaluation metrics. The progression from baseline to optimized models in both scenarios highlights the significant impact of proper hyperparameter tuning, particularly when implementing class balancing techniques. A comparative visualization of these performance metrics across all model configurations can be seen in Figure 8. The optimal parameter configurations determined by Optuna that led to these improvements for both scenarios are presented in Table VII.

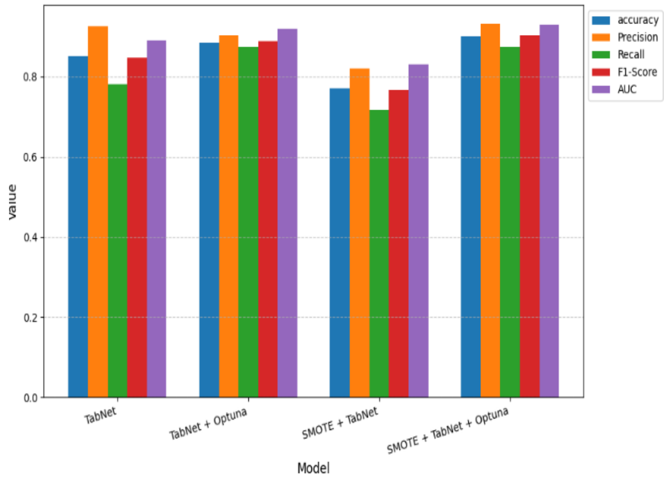


Figure 8. Comparative Visualization of Model Performance Metrics

TABLE VII. OPTIMIZED PARAMETERS FOR DIFFERENTS SCENARIOS

Parameter	Range Value	Scenario I	Scenario II
n_d	8 - 32	21	8
n_a	8 – 32	15	8
n_steps	5 – 8	8	8
n_independent	1 - 2	2	2
learning_rate	0.01 – 0.1	0.024196550894727036	0.02114274661219965
gamma	1.0 – 2.0	1.8305314575602554	1.0251532561037215
lambda_sparse	0.0001 – 0.01	0.0014497065638336094	0.00026367649497584590

TABLE VIII. COMPARATIVE ANALYSIS OF TABNET MODEL PERFORMANCE WITH RELATED RESEARCH

Author	Best Model	Evaluation				
		Accuracy	Precision	Recall	F1-Score	AUC
Hirwono et al. [17]	Naïve Bayes	86.64%	85.07%	89.36%	91.94%	-
Nawawi et al. [18]	Neural Network	84.52%	85.31%	98.85%	-	0.60
Firdaus et al. [19]	MLP	97.50%	97.55%	97.50%	97.48%	-
Baliani et al. [20]	Gradient Boosting	89.50%	-	-	-	-
Ratnasari et al. [21]	Naïve Bayes	84.67%	-	-	-	0.50
Proposed Method	**TabNet**	**90.16%**	**93.33%**	**87.50%**	**90.32%**	**0.93**

Compared to previous studies on heart disease classification Table VIII, our TabNet model with SMOTE and Optuna optimization demonstrated several significant advancements. While some previous approaches have achieved comparable accuracy, our integrated methodology addresses critical limitations in existing methods and offers distinct advantages for real-world clinical applications.

The Naïve Bayes model implemented by Hirwono et al. [17] achieved respectable accuracy (86.64%) but suffered from significant limitations in discriminative capability as evidenced by its unreported AUC values. Similarly, Ratnasari et al. [21] reported a comparable accuracy of 84.67% using Naïve Bayes, but their study revealed an extremely low AUC of only 0.50, effectively equivalent to random guessing in terms of ranking capability. These findings highlight a critical limitation in many previous studies: the overreliance on accuracy as the sole performance metric, which can be misleading in medical diagnostics where false negatives carry serious consequences.

Neural Network approaches, such as that employed by Nawawi et al. [18], showed particularly poor discriminative ability with an AUC of only 0.60 despite reasonable accuracy

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2348, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

(84.52%). This substantial performance gap compared to our approach underscores the limitations of conventional neural networks when handling tabular medical data without appropriate attention mechanisms and hyperparameter optimization. The attention mechanism in TabNet provides a critical advantage by focusing on the most relevant features for each individual case, unlike traditional neural networks which process all features equally. Research by Firdaus et al. [19] reported high accuracy (97.55%) using MLP, but this result was achieved using a 90:10 train-test split ratio, which can artificially inflate performance metrics compared to our more robust 80:20 split. Their extreme split ratio likely led to overly optimistic results with limited test samples, whereas our approach with a larger test set provides a more realistic assessment of generalization capability. Additionally, their study lacked comprehensive evaluation across diverse metrics beyond accuracy, particularly AUC, which our research demonstrates is crucial for clinical applications

The Gradient Boosting approach by Baliani et al. [20] achieved 89.5% accuracy using manual parameter tuning with fixed incremental values for learning rate and estimators, whereas our approach leveraged Optuna's Bayesian optimization to systematically explore the parameter space, achieving superior performance (90.16% accuracy). This difference highlights the advantage of our automated optimization strategy over predefined parameter testing, enabling discovery of optimal configurations that manual experimentation likely missed.

IV. CONCLUSION

The implementation of SMOTE technique alone in the context of heart disease classification did not necessarily lead to improved model performance. This was evidenced by the decrease in accuracy from 85.24% to 77.04% in the baseline TabNet implementation. Statistical analysis of the training loss patterns ($p < 0.0001$) revealed that applying SMOTE without appropriate parameter adjustments actually caused the model to overfit to synthetic samples rather than learning generalizable patterns from the data. This phenomenon was particularly pronounced in our relatively small dataset, where synthetic samples generated by SMOTE failed to adequately capture the complexity of real patient data.

However, when SMOTE was combined with parameter optimization using Optuna, the model achieved its best overall performance with 90.16% accuracy, 93.33% precision, 87.50% recall, 90.32% F1-score, and an AUC of 0.93. This significant improvement across all metrics demonstrated the synergistic effect of combining appropriate data balancing techniques with hyperparameter optimization, where Optuna effectively counteracted potential overfitting by fine-tuning regularization parameters.

The key contributions of this research included:

- Identification of the potential negative impact of SMOTE when applied in isolation
- Demonstration of Optuna's effectiveness in optimizing TabNet parameters
- Achievement of state-of-the-art discriminative capability with an AUC of 0.93, representing a substantial improvement over previous approaches

These findings have important implications for clinical applications, where improved classification accuracy and reliability could support earlier and more accurate diagnosis of heart disease, potentially improving patient outcomes through timely interventions. The sequential attention mechanism of TabNet, when properly optimized, provides an interpretable model that could help clinicians understand the factors contributing to a particular diagnosis.

Future research could extend this work by investigating the application of different optimization techniques to TabNet model, and evaluation on larger datasets to further validate the generalizability of our approach. Additionally, exploring the interpretability aspects of the optimized TabNet model could provide valuable insights for medical practitioners in understanding the factors contributing to heart disease classification.

REFERENCES

[1] G.A. Roth, C. Johnson, A. Abajobir, F. Abd-Allah, S.F. Abera, G. Abyu, et al., "Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015," J. Am. Coll. Cardiol., vol. 70, no. 1, pp. 1-25, 2017.

[2] M.I. Jordan and T.M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, vol. 349, no. 6245, pp. 255-260, 2015.

[3] S. Falkner, A. Klein, and F. Hutter, "Practical hyperparameter optimization for deep learning," in AutoML: Methods, Systems, Challenges, F. Hutter, L. Kotthoff, and J. Vanschoren, Eds., Cham: Springer, pp. 3-25, 2018.

[4] A. Homaidi and Z. Fatah, "Implementasi metode K-nearest neighbors (KNN) untuk klasifikasi penyakit jantung," G-Tech: Jurnal Teknologi Terapan, vol. 8, no. 3, pp. 1720-1728, 2024.

[5] A. Masruriyah, H. Novita, C. Sukmawati, A. Ramadhan, S. Arif, and B. Dermawan, "Pengukuran kinerja model klasifikasi dengan data oversampling pada algoritma supervised learning untuk penyakit jantung," Computer Science (CO-SCIENCE), vol. 4, no. 1, pp. 62-70, 2024.

[6] S.Ö. Arik and T. Pfister, "TabNet: Attentive interpretable tabular learning," in Proc. AAAI Conf. Artif. Intell., vol. 35, no. 8, pp. 6679-6687, May 2021.

[7] A.R. Raharja, A. Pramudianto, and Y. Muchsam, "Penerapan algoritma decision tree dalam klasifikasi data 'Framingham' untuk menunjukkan risiko seseorang terkena penyakit jantung dalam 10 tahun mendatang," Technol. J., vol. 1, no. 1, 2024.

[8] D. Nasien, et al., "Klasifikasi penyakit jantung menggunakan decision tree dan KNN menggunakan ekstraksi fitur PCA," JEKIN-Jurnal Teknik Informatika, vol. 4, no. 1, pp. 18-24, 2024.

[9] T. Indriyani, et al., "Metode decision tree C4.5 untuk klasifikasi penyakit jantung," Prosiding Seminar Nasional Sains dan Teknologi Terapan, no. 1, 2024.

[10] J.J. Tamilselvi and C.B. Gifta, "Handling duplicate data in data warehouse for data mining," Int. J. Comput. Appl., vol. 15, no. 4, pp. 7-15, 2011.

[11] S.K. Kwak and J.H. Kim, "Statistical data preparation: Management of missing values and outliers," Korean J. Anesthesiol., vol. 70, no. 4, pp. 407, 2017.

[12] J. Brownlee, Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning, Machine Learning Mastery, 2020.

[13] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 2623-2631, July 2019.

[14] C. Sammut and G. I. Webb, Eds., Encyclopedia of Machine Learning, Springer Science & Business Media, 2011.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2348, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

[15] S. Sathyanarayanan and B. R. Tantri, "Confusion matrix-based performance evaluation metrics," Afr. J. Biomed. Res., vol. 4023, pp. 4023-4031, 2024.

[16] A. Tafvizi, B. Avci, and M. Sundararajan, "Attributing AUC-ROC to analyze binary classifier performance," arXiv preprint arXiv:2205.11781, 2022.

[17] B. Hirwono, A. Hermawan, and D. Avianto, "Implementasi metode Naïve Bayes untuk klasifikasi penderita penyakit jantung," J. JTIK (Jurnal Teknol. Inf. Komun.), vol. 7, no. 3, pp. 450-457, 2023.

[18] H. M. Nawawi, J. J. Purnama, and A. B. Hikmah, "Komparasi algoritma neural network dan Naïve Bayes untuk memprediksi penyakit jantung," J. Pilar Nusa Mandiri, vol. 15, no. 2, pp. 189-194, 2019.

[19] R. Firdaus, D. Mualfah, and J. S. Hasanah, "Klasifikasi multi-class penyakit jantung dengan SMOTE dan Pearson's correlation menggunakan MLP," J. CoSciTech (Comput. Sci. Inf. Technol.), vol. 4, no. 1, pp. 262-271, 2023.

[20] M. D. I. Baliani, R. R. Huizen, and G. A. Pradipta, "Perbandingan performa data penyakit jantung menggunakan pendekatan klasifikasi boosting methods," in *Seminar Hasil Penelitian Informatika dan Komputer (SPINTER)*, Institut Teknologi dan Bisnis STIKOM Bali, 2024, pp. 894-899.

[21] A. J. Wahidin, A. E. Setiawan, and P. Bintoro, "Machine learning untuk klasifikasi penyakit jantung," *Aisyah J. Inf. Electr. Eng. (AJIEE)*, vol. 6, no. 1, pp. 145-150, 2024.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2348, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Clustering Snack Products Based on Nutrition Facts Using SOM and K-Means for Diabetic Dietary Recommendation

Maritza Adelia[1], Arum Handini Primandari[2]*
Department of Statistics, Faculty of Mathematics and Natural Science[1], [2]
Universitas Islam Indonesia
Yogyakarta, Indonesia
maritza.adelia@students.uii.ac.id[1] primandari.arum@uii.ac.id [2

***Abstract*— The number of diabetics in Indonesia continues to rise, with Type II Diabetes Mellitus (DM) dominating 90% of cases. One of the main contributors is the excessive consumption of snack products high in Sugar, Salt, and Fat (SSF), which increases health risks, particularly for diabetics. However, the current nutrition facts provided in the product package is not easy to understand. Creating label for the product can make an effective information to assist people on buying decision. This study aims to segment snack products based on their nutritional facts, particularly focusing on their SSF content, to identify products that are potentially high-risk for diabetics. In this study, data on the nutritional facts of snack products were analyzed. Utilizing a hexagonal Self-Organizing Map (SOM) topology with a 5 × 9 grid, the best clustering method identified was k-means. This method yielded two clusters, with a silhouette index of 0.44, a Dunn index of 0.09, and a connectivity index of 11.14. The first cluster comprises 165 products that have low levels of total fat, saturated fat, sugar, and salt. In contrast, the second cluster consists of 46 products with high total fat and saturated fat content, and this cluster is of particular concern due to its elevated levels of these unhealthy fats. The segmentation results can serve as a reference for more intuitive food labeling, potentially improving consumer awareness and aiding in dietary decision-making, particularly for diabetics.**

***Keywords*— Clustering SSF, nutrition facts, snack healthy label, SOM**

I. Introduction

Diabetes Mellitus (DM) is a serious problem in Indonesia and around the world. DM is a chronic condition characterized by the body's inability to produce adequate insulin or effectively utilize the insulin it produces, resulting in elevated blood glucose levels. According to IDF (International Diabetes Federation), number of people with DM in Indonesia reached 19.47 million in 2021 and is expected to increase over time. DM consists of 4 types, namely type I, type II, gestational DM, and other DM, but as reported by IDF, 90% of diabetics suffer from type II DM [1].

Type II DM is influenced by unhealthy lifestyle factors or triggered by other conditions such as high blood pressure or obesity. In addition, the consumption of packaged foods and beverages high in Sugar, Salt, and Fat (SSF, Indonesian: *Gula, Garam, Lemak*-GGL) is a risk factor [2]. According to the Individual Food Consumption Survey, approximately 77 million Indonesians have consumed SSFs above the daily limit, 53.1% of whom are adolescents aged 13-18 years [3]. This indicates a significant health risk that needs immediate attention.

To overcome this problem, the Food and Drug Administration (BPOM RI), through BPOM Regulation No. 26 of 2021, has required the inclusion of the nutrition facts on the label of processed packaged products and urges the public to always read the nutrition facts table correctly and carefully [4]. However, the low level of public awareness regarding the nutrition facts table has driven BPOM to introduce the Nutri-Level program. This program focuses on labeling the risk level of SSF content by indicating high and low SSF levels in packaged products [5]. To support this initiative, clustering packaged snack products based on their SSF content can be a powerful strategy. Through cluster analysis, it is possible to systematically group products with similar SSF content, identify high-risk products, and develop consistent labeling schemes. This not only helps policymakers implement targeted regulations but also increases public awareness of health risks associated with excessive SSF consumption. The Self-Organizing Map (SOM) method is one of the clustering techniques that enables the grouping of products based on the similarity of their SSF profiles.

First introduced by Professor Teuvo Kohonen in 1982, SOM is a technique for visualizing and clustering data according to its characteristics [6]. This method is able to cluster high dimensional data and is resistant to noise and outliers [7]. In the SOM method, there are output neurons that can be regrouped to simplify the clustering results and make them easier to understand. This method was chosen because it can cluster high-dimensional data and is resistant to noise and outliers [8]. The methods used in this research are hierarchical agglomerative methods in the form of complete linkage and average linkage and the k-means method because these three methods can be used in clustering output neurons in the SOM topology.

Previous research has explored the application of the SOM method for clustering various datasets. For instance, Hardika K.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2342, Copyright ©2025
Submitted : April 28, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

(2018) for clustering social and population data from 33 provinces in Indonesia, which are indicators of remote and disadvantaged areas, and 2 clusters were formed with SOM and k-means methods as advanced clustering methods. There is also research on grouping 38 packaged products based on nutrition facts by Husna et al. (2019) using k-means method and 2 clusters were formed. However, there has been no research specifically focusing on segmenting snack products using SOM with SSF content as the primary basis for clustering. This research aims to fill that gap by identifying high-risk snack product groups, which can further support effective public health strategies through product labeling and increased consumer awareness.

II. METHODOLOGY

A. Material and Data

The population in this study were food products included in the category 15.0 of snack products based on BPOM Regulation Number 13 of 2023. The category includes all types of savory or other flavored snacks: 15.1 snacks – potato, tubers, cereals, flour or starch (from tubers and nuts); 15.2 nut preparations, including coated nuts and nut mixtures (examples with dried fruit); and 15.3 fish-based snacks. A total of 211 samples were taken using purposive sampling technique because it was based on the retrieval criteria: products included in the category of snack products having nutrition facts tables in their packaging and are sold at Manna Kampus Godean, Toko Agung Grosir Yogyakarta, and KN Putra Toserba Magelang. Data were collected by photographing the nutrition facts table of products that fall under the snacks category using a mobile phone camera as shown in Fig 1. The variables used in this study were total fat, saturated fat, sugar, and salt. The following in Table 1 are operational definition of research variables.

TABLE I. OPERATIONAL DEFINITION OF RESEARCH VARIABLE

No.	Variable	Definition	Unit	Scale
1	Total Fat	All fatty acids in food and expressed as triglycerides	Gram	Ratio
2	Saturated Fat	All fatty acids without double bonds	Gram	Ratio
3	Sugar	The sum of all monosaccharides and disaccharides found in processed foods	Gram	Ratio
4	Salt	Amount of salt (sodium) listed as total sodium	Gram	Ratio

Fig. 1. Example of Sample Data Collection

B. Research Method

This research consists of several stages. The following in Fig 2 is a research flowchart.

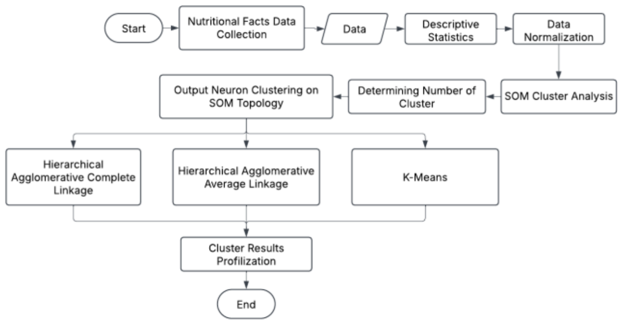


Fig. 2. Research Flowchart

1) Descriptive Statistics

Descriptive statistics is a stage of data analysis in which the data is described without the aim of making general conclusions or generalizations [9]. Descriptive statistical techniques used in this study are maximum, minimum, mean, and visualization in the form of boxplots.

2) Data Normalization

Data normalization is used to rescale the data so that the analysis results are more representative. This is because each variable in a data set often has a range of values that are very different. The method used is min-max normalization, where the data scale is changed to a range of 0 to 1. The following is the formula for min-max normalization [10].

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where x' is the normalized data, x_i is the actual data, x_{min} is the minimum value of data per variable, and x_{max} is the maximum value of data per variable.

3) Self-Organizing Map (SOM)

SOM represent a form of unsupervised Artificial Neural Network that functions to both reduce dimensionality and group similar data points into clusters according to their characteristics. The architecture of SOM consists of an input layer with input neurons and an output layer with output neurons. SOM itself uses competitive learning in its algorithm, which means that the output neurons compete to determine the closest distance to the input neuron until a winning neuron is obtained [11]. To determine the size of the grid or the number of output neurons to form, the researcher uses the Kohonen formula, which states that the maximum number of output neurons should be $5 \times \sqrt{N}$, where N is the number of observations in the data [12]. The number of iterations and the type of topology used must also be determined. The optimal number of iterations is reached when the map has reached a stable state, or as can be seen from the average distance to the nearest unit value that is stable at each iteration [13]. SOM also has two types of topology such as hexagonal, where each neurons has at most 6 neighbors and rectangular, where each neurons has at most 4 neighbors. but hexagonal topology is preferred because it allows better visualization of the overall

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2342, Copyright ©2025
Submitted : April 28, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

data structure [14].

Then, after the input vectors are successfully clustered in each output neuron, a partitioning of the output neurons formed by the vector weights on each output neuron is performed using another clustering method, such as the k-means method. This is done to make the boundaries between clusters clearer and to simplify the clustering results [8]. The steps for performing SOM clustering are as follows [15].

1. Initialize the weight vector between the input neuron and the output neuron with a random number from 0 to 1.
2. Calculate the distance between the input vector and the weight vector for each output neuron using Euclidean distance, and take the output neuron with the smallest distance value as the winning neuron. The formula for the Euclidean distance is as follows
$$d_j = \sum_{i=1}^{p}\left(w_{ij} - x_{ki}\right), k = 1, 2, \ldots, n \quad (2)$$
where w_{ij} is the weight vector with $i = 1,2,\ldots,p$ and $j = 1,2,\ldots,m$ where p is the number of variables and m is the number of output neurons and x_{ki} is the value of the k-th input vector in the i-th variable.
3. Use the following formula to update the weight vector of the winning neurons.
$$w_{ij}(new) = w_{ij}(old) + \alpha(x_{ki} - w_{ij})(old) \quad (3)$$
where α is the learning rate, which has a value of $0 \leq \alpha \leq 1$ and will decrease with the number of iterations performed.
4. For each input vector x, repeat steps two through three.
5. Update the learning rate (α) at the t-th iteration with $t = 1,2,\ldots,T$ with the following equation.
$$\alpha(1 + t) = \alpha(t)\left(1 - \frac{t}{T}\right) \quad (4)$$
where α is the learning rate, which has a value of $0 \leq \alpha \leq 1$ and will decrease with the number of iterations performed.
6. Until the maximum iteration is reached and the learning rate converges to zero, repeat steps four through five.
7. Group each observation object or input vector into the output neuron with the closest distance or the one with the smallest distance value.

4) Hierarchical Clustering

Hierarchical clustering is a clustering method that uses a hierarchical structure or level in the process. This method is divided into two types, namely divisive and agglomerative. Divisive means that objects are placed in one cluster and then divided into several clusters, while agglomerative means that adjacent objects are combined into separate clusters and then adjacent clusters are combined until all objects are included in one cluster [16]. In this study, the agglomerative method will be used in the form of complete linkage and average linkage.

a) Complete Linkage

Grouping in the complete linkage method is based on the greatest distance between objects in different clusters. The steps for the calculation are as follows [17].

1. Using the Euclidean distance size equation, calculate the distance matrix D between objects using the equation.
$$d_{i,j} = \sqrt{\sum_{k=1}^{p}\left(x_{ik} - x_{jk}\right)^2} \text{with } k = 1, 2, \ldots, n \quad (5)$$
where d_{ij} is the Euclidean distance between the i-th object and the j-th object, x_{ik} and x_{jk} are the values of the i-th and j-th objects in the k-th variable, and p is the number of variables observed.
2. Select the closest distance in the distance matrix $D = \{d_{ij}\}$, then combine the two closest objects, for example objects U and V form a cluster (UV).
3. Update the distance matrix D by calculating the distance between clusters (UV) and other objects using the following formula.
$$d_{(UV)W} = \max\left(d_{UW}, d_{VW}\right) \quad (6)$$
where, $d_{(UV)W}$ is the distance between cluster (UV) and object W, d_{UW} is the distance between object U and W, and d_{VW} is the distance between object V and W.
4. Repeat the third step until all objects are placed in one cluster.

b) Average Linkage

Grouping in the average linkage method is based on the average between objects in different clusters. The steps for the calculation are as follows [17].

1. Using the Euclidean distance size equation (5), calculate the distance matrix D between objects using the equation.
2. Select the closest distance in the distance matrix $D = \{d_{ij}\}$, then combine the two closest objects, for example objects U and V form a cluster (UV).
3. Update the distance matrix D by calculating the distance between clusters (UV) and other objects using the following formula.
$$d_{(UV)W} = \frac{d_{(UW)} + d_{(VW)}}{n_{(UV)}n_W} \quad (7)$$
where, $d_{(UV)W}$ is the distance between cluster (UV) and object W, $d_{(UW)}$ is the distance between objects U and W, $d_{(VW)}$ is the distance between objects V and W, $n_{(UV)}$ is the number of members in cluster (UV), and n_W is the number of members in cluster W.
4. Repeat the third step until all objects are placed in one cluster.

5) K-Means

The k-means method is a non-hierarchical method that aims to group objects into clusters based on their characteristics. The following are the calculation steps of the k-means method [18].

1. Determine the number of clusters or the value of k and randomly initialize the center of the cluster (centroid) as many as k.
2. Calculate the distance of each object to the centroid using the Euclidean distance equation (5) until the closest distance of each object to the centroid is found.
3. Assign the object to the cluster with the closest centroid.
4. Perform iterations from step 3. The new centroid value is calculated using the following equation.
$$C_k = {}^{1}/_{n_k} \sum d_i \quad (8)$$
Where n_k is the number of data in cluster k and d_i is the

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2342, Copyright ©2025
Submitted : April 28, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

sum of the distance values contained in each cluster.

5. Iterate until the centroid value and members of each cluster do not change. If the condition is not met, repeat from step two.

6) *Cluster Validation*

Cluster validation is a step to quantitatively and objectively evaluate the results of cluster analysis [19]. This research uses the silhouette index, Dunn index, and connectivity index methods.

Silhouette index used to evaluate the quality and strength of clusters, or how accurately an object is placed in a cluster [20]. The silhouette index has a value between 1 and -1, the closer the value is to 1, the more correct is the clustering structure produced, the closer the value is to -1, the more overlapping is the clustering structure produced. The following is the calculation formula [21].

$$s(i) = \frac{b(i)-a(i)}{\max\,(a(i),b(i))} \tag{9}$$

where $s(i)$ is the silhouette value for the i-th data, $a(i)$ is the average distance of the i-th object to all objects in the same cluster, and $b(i)$ is the minimum value of the average distance between the i-th object and objects in other clusters.

Dunn index is the ratio of the smallest distance between observations in different clusters to the largest distance between observations in the same cluster. Dunn index has a value of $0 \leq D \leq \infty$, and the higher the dunn index value, the better the resulting cluster. The following is the calculation formula [21].

$$D = \min_{j=i+1,\dots,n_c}\left(\frac{d(c_i,c_j)}{\max_{k=1,\dots,n_c}(diam(c_k))}\right) \tag{10}$$

where, D is the Dunn index value, $d(c_i, c_j)$ is the distance between clusters c_i and c_j, and $\max_{k=1,\dots,n_c}(diam(c_k))$ is the maximum distance between objects in one cluster with n_c being the total number of clusters.

Connectivity index evaluates the homogeneity of the cluster. The connectivity index has a value of $0 \leq C \leq \infty$, and the smaller the value, the better the resulting cluster. The following is the calculation formula [21].

$$Conn = \sum_{i=1}^{N}\sum_{j=1}^{L} x_{i,nn_{i(j)}} \tag{11}$$

where, $nn_{i(j)}$ is the nearest neighbor of the i-th object to the j-th object, N is the number of objects, and L is the number of clusters. $x_{i,nn_{i(j)}}$ is 0 if i and $nn_{i(j)}$ are in the same cluster and $1\,/\,j$ if they are in different clusters.

7) *Independent Samples t-Test*

An independent samples t-test is used to determine whether there are significant differences in the means of a variable when comparing two unrelated groups. The null hypothesis for this test is that there is no significant difference between the means of the two groups. The test statistics for the independent samples t-test are as follows [22].

$$t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \tag{12}$$

Where $\bar{x}_1$ is the group 1 mean, $\bar{x}_2$ is the group 2 mean, s_1^2 is the group 1 variance, s_2^2 is the group 2 variance, n_1 is the number of group 1 observations, and n_2 is the number of group 2 observations.

III. RESULT AND ANALYSIS

A. *Descriptive Statistics*

Descriptive statistics include mean, minimum, maximum, and boxplot visualization. The following are descriptive statistics of the nutrition facts data for snack products.

TABLE II. DESCRIPTIVE STATISTICS

Variable	Measure		
	Mean	**Minimum**	**Maximum**
Total Fat (g)	5.811	1	21
Saturated Fat (g)	2.3	0	6
Sugar (g)	1.818	0	15
Salt (g)	0.133	0.005	1.095

According to Table II, the highest total fat is 21 g contained in Aceh Fish Skin Salted Egg, the highest saturated fat is 6 g within Chitato Lite Onion Cream Sauce and Japota Spicy Lime, the the highest sugar content is 15 g contained in Sunbay Snack Spicy Crispy Squid, and the the highest salt content is 1,095 g within Maxicorn Roasted Corn. Moreover, based on the average value of each variable, none of them exceeded the daily intake limit according to the Ministry of Health. The maximum recommended daily intake for SSL is 67 g of fat, 50 g of sugar, and 5 g of salt. However, the total fat and sugar content has a maximum value of 21 g and 15 g, indicating a fairly high number for snack products when compared to the daily consumption limit.

The following are the results of boxplot visualization of nutrition facts data for snack products.

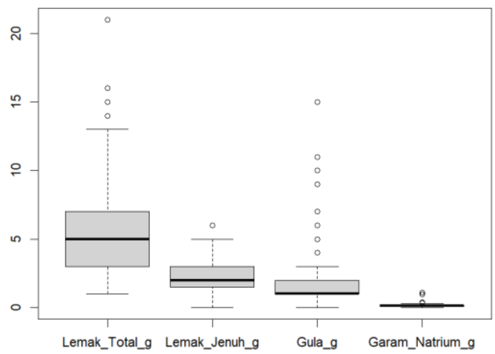


Fig. 3. Boxplot Visualization

According to the boxplot visualization in Fig 3, there are a total of 45 outliers. Since these outliers contain important information, the cluster analysis is performed using the SOM method which is insensitive to outliers.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2342, Copyright ©2025
Submitted : April 28, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

B. Data Normalization

Data normalization is essential because each variable in the dataset has a significant range. Standardizing the scale of the data is necessary to facilitate cluster analysis. The following presents the nutritional information of snack products after applying min-max normalization. The results, illustrated in Table III, demonstrate that the normalized values are uniformly distributed within a range of 0 to 1. Consequently, this data is suitable for further cluster analysis using SOM.

TABLE III. DATA AFTER NORMALIZED

No.	Product Name	Total Fat (g)	Saturated Fat (g)	Sugar (g)	Salt (g)
1	Nissin Sagu Keju	0.286	0.167	0.286	0.004
2	Happy Tos Corn Chips Merah	0.286	0.119	0.048	0.005
…	…	…	…	…	…
210	Aceh Fish Skin Salted Egg Spicy	1	0	0.048	0.018
211	Aceh Fish Skin Salted Egg	1	0	0.048	0.018

Based on Table III, the data is successfully normalized with a range between 0 and 1, so the data is ready for clustering analysis using SOM.

C. Cluster Analysis with SOM

In this analysis, a hexagonal topology was employed, and a total of 1500 iterations were conducted. To determine the optimal grid size and the number of output neurons, experiments were performed utilizing the silhouette index as a validation method. Below are the results of these experiments.

TABLE IV. RESULTS OF SOM TOPOLOGY GRID SIZE EXPERIMENT

No.	*Grid*	*Silhouette Index*	Number of Empty Output Neurons
1.	1 × 7	0.37	0
2.	1 × 8	0.36	0
…	…	…	…
27.	5 × 9	0.51	0
…	…	…	…
36.	8 × 9	0.48	11
37.	9 × 9	0.55	17

Based on the experimental results presented in Table IV, the optimal grid size for SOM analysis consists of a grid with up to 45 output neurons, yielding a silhouette index value of 0.51. This value indicates that the grid size is adequate for analysis. Although there are other grid sizes with higher silhouette index values, the chosen grid size is preferred because it avoids empty output neurons, which can lead to overfitting. Therefore, a grid of this size will be used in the SOM analysis. Below is a graph illustrating the training progress of the resulting SOM model.

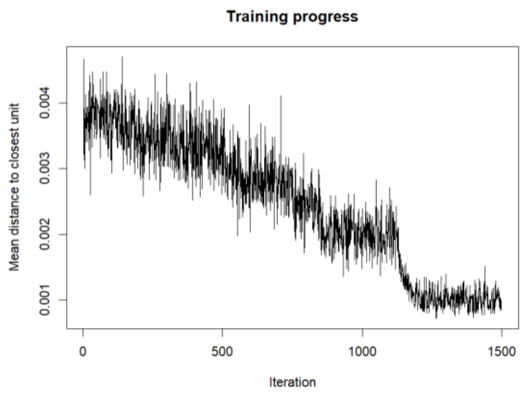


Fig. 4. SOM Model Training Progress Graph

The graph in Fig 4 shows that the average distance to the nearest output unit or neuron decreases as the number of iterations increases. After about 1100 iterations, the average distance to the nearest unit is less than 0.001 and remains stable or converges until the 1500th iteration. This indicates that the resulting cluster is good, because the smaller the average distance value to the nearest unit, the better the resulting cluster. After the iteration process, the resulting SOM topology with 45 output neurons is represented by a fan diagram as follows.

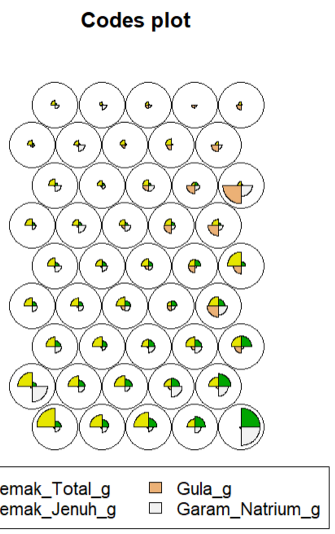


Fig. 5. Fan Diagram of SOM Analysis Results

The fan diagram in Fig 5 shows the distribution of each variable of the snack products in each output neuron. The larger the fan shape of a variable, the greater the content of that variable in the snack products included in the members of an output neuron. Below in Table V is a list of the snack product members in each output neuron.

TABLE V. LIST OF SNACK PRODUCTS IN SOM OUTPUT NEURONS

Output Neuron	Product Name
V1	Chitato Lite Saus Krim Bawang dan Japota Spicy Lime.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2342, Copyright ©2025
Submitted : April 28, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Output Neuron	Product Name
...	...
V5	Aceh Fish Skin Salted Egg Spicy dan Aceh Fish Skin Salted Egg.
...	...
V35	Oishi Caramel Popcorn, Oishi Chocolate Popcorn, Krizz Chocolate, etc.
...	...
V45	Oishi Pillows Ubi, Oishi Pillows Keju, Oishi Pillows Durian, etc.

According to Fig 5, there is an output neuron associated with products that have a high total fat and salt content. This neuron is identified as neuron V5, which has an average total fat content of 21 g and a salt content of 0.38 g. Additionally, there is another output neuron that indicates high sugar content; this neuron is designated as neuron V35, which has an average sugar content of 11 g.

To clarify the interpretation of the formed clusters, further analysis is conducted through clustering the output neurons. This clustering process is based on the vector weights of the 45 output neurons. Table VI presents the vector weights for each output neuron as established in the SOM topology.

TABLE VI. SOM OUTPUT NEURON VECTOR WEIGHT

Output Neuron	Total Fat (g)	Saturated Fat (g)	Sugar (g)	Salt (g)
V1	0.359629	0.285714	0.047619	0.00607
V2	0.425696	0.184134	0.044744	0.006705
...	...	...	...	...
V45	0.094989	0.047619	0.142857	0.001766
V41	0.144807	0.064383	4.41E-09	0.004938

D. Output Neuron Clustering

a) Complete Linkage

Before clustering output neurons using the complete linkage method, it is crucial to determine the optimal number of clusters. The results of cluster validation are shown below.

TABLE VII. CLUSTER VALIDATION OF COMPLETE LINKAGE METHOD

Method	Number of Cluster			
	2	3	4	5
Silhouette Index	0.53	0.43	0.38	0.30
Dunn Index	0.22	0.18	0.27	0.28
Connectivity Index	7.18	13.33	15.29	15.79

According to Table VII, the optimal number of clusters is determined to be 2, as it has the highest silhouette index and the lowest connectivity index. However, the Dunn Index indicates that the best results are achieved with 5 clusters. Despite this, we have decided to proceed with the clustering of 2 groups. The values of the all validation method indicate a good cluster structure, where objects in the same cluster have a high degree of similarity. Below are the results of clustering the output neurons of the SOM model, along with a visualization using a fan diagram.

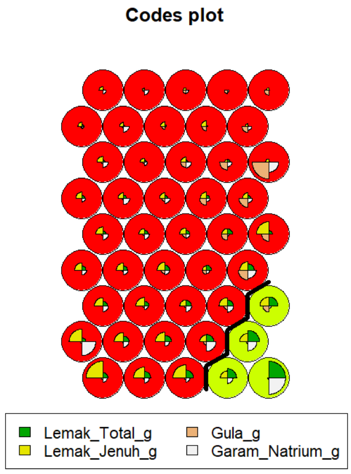


Fig. 6. Fan Diagram of Complete Linkage Method Results

Fig 6 shows that the output neurons are successfully grouped into 2 clusters, where the first cluster in red circles consists of 196 products with high sugar content and cluster 2 in yellow circles consists of 15 products with high total fat, saturated fat, and salt content. The clustering results show an extreme imbalance in the number of members in both clusters.

b) Average Linkage

The number of clusters associated with its measurement for average linkage is presented the following table.

TABLE VIII. CLUSTER VALIDATION OF AVERAGE LINKAGE METHOD

Method	Number of Cluster			
	2	3	4	5
Silhouette Index	0.65	0.46	0.37	0.31
Dunn Index	0.43	0.56	0.19	0.23
Connectivity Index	3.05	5.98	16.83	17.97

Based on Table VIII, the optimal number of clusters is 3 clusters because it has the largest silhouette index and Dunn index values and the smallest connectivity index value. The values of all validation methods show a good cluster structure, where objects in the same cluster have a high degree of similarity and are better than the previous method.

Then, the following is the results of clustering the output neurons of the SOM model with visualization using a fan diagram.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2342, Copyright ©2025
Submitted : April 28, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

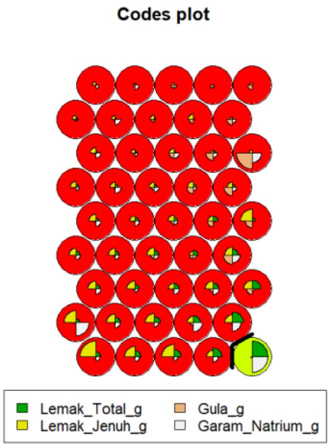


Fig. 7. Fan Diagram of Average Linkage Method Results

Fig 7 shows that the output neurons are successfully grouped into 2 clusters, where the first cluster shown in red circles consists of 209 products with high saturated fat and sugar content, and cluster 2 shown in yellow circles consists of 2 products with high total fat and salt content. This cluster result shows an extreme imbalance in the number of members in both clusters compared to the previous method.

c) K-Means

The following table represent the associate of the number of clusters and its evaluation measurements.

TABLE IX. CLUSTER VALIDATION OF K-MEANS METHOD

Method	Number of Cluster			
	2	3	4	5
Silhouette Index	0.44	0.43	0.37	0.30
Dunn Index	0.09	0.18	0.19	0.15
Connectivity Index	11.14	13.33	16.83	23.64

According to Table IX, the optimal number of clusters is determined to be 2, as this configuration yields the highest silhouette index and the lowest connectivity index. On the other hand, the Dunn index indicates that the optimal results are achieved with 4 clusters. Despite this, 2 clusters are prioritized since both evaluation methods indicate they produce the best outcomes. The values of all validation method indicate a weak cluster structure, but are still acceptable because this method is still able to identify data groups with certain similarity patterns. Below are the results of clustering the output neurons from the SOM model, along with a visualization using a fan diagram.

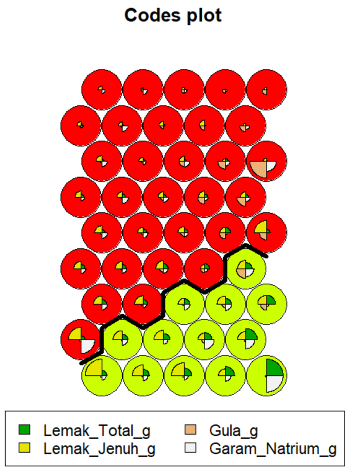


Fig. 8. Fan Diagram of K-Means Method Results

Fig 8 shows that the output neurons are successfully grouped into 2 clusters, where the first cluster in red circles consists of 165 products with high sugar content and cluster 2 in yellow circles consists of 46 products with high total fat, saturated fat, and salt content. This result shows a better distribution of cluster members compared to the two previous methods.

E. Results and Cluster Profilization

According to the output neuron clustering, the best method is k-means which forms 2 clusters. The following are the results of clustering the products into 2 clusters.

TABLE X. SNACK PRODUCT CLUSTERING RESULTS

Cluster	Number of Products	Products
1	165	Happy Tos Corn Chips Merah, TosTos Tortilla Chips Roasted Corn, Maxicorn Roasted Corn, Chitato Ayam Bumbu, Oishi Sponge Chocolate, etc.
2	46	Chitato Lite Saus Krim Bawang, Japota Sapi Panggang, Potabee Wagyu Beef Steak, Garuda Kacang Kulit Rasa Bawang, Krizz Cheese, etc.

Based on the product clustering results in Table X, the first cluster results in 165 products and the second cluster results in 46 products. For a complete list of products and to search for specific snack products by cluster, use the link https://bit.ly/CariMakananRingan. Then, to evaluate whether there are significant differences between the two clusters for each variable, an independent samples t-test is performed. Below are the results of the independent samples t-test.

I. Hypothesis
$H_0: \mu_1 = \mu_2$ (There is no significant difference in the average nutrient content between clusters 1 and 2)
$H_1: \mu_1 \neq \mu_2$ (There is significant difference in the average nutrient content between clusters 1 and 2)

II. Significance Level
$\alpha = 0.05$

III. Critical Area

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2342, Copyright ©2025
Submitted : April 28, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Reject H_0 if $t_{hitung} < -t_{\alpha/2}$ or $t_{hitung} > t_{\alpha/2}$ or $p-value \leq \alpha$

IV. Test Statistics

$$t_{hitung} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

V. Decision and Conclusion

TABLE XI. INDEPENDENT SAMPLES T-TEST RESULT

Variable	$P - value$	Decision
Total Fat	$< 2.2 \times 10^{-16}$	Reject H_0
Saturated Fat	1.544×10^{-14}	Reject H_0
Sugar	0.2155	Fail to Reject H_0
Salt	0.09897	Fail to Reject H_0

Based on Table XI and using the 95% confidence level, it can be concluded that there is a significant difference in the average total fat and saturated fat content in clusters 1 and 2, while the average sugar and salt content is not significantly different.

Next, profiling is performed to determine the characteristics of each cluster as follows.

TABLE XII. CLUSTER RESULTS PROFILIZATION

Cluster	Average Total Fat (g)	Average Saturated Fat (g)	Average Sugar (g)	Average Salt (g)
1	4.45	1.99	1.91	0.13
2	10.69	3.4	1.49	0.16

According to the profiling results presented in Table XII, the first cluster includes products that contain low levels of total fat, saturated fat, sugar, and salt. In contrast, the second cluster comprises products that are high in total and saturated fat but low in sugar and salt. Snack products in the second cluster, which should be monitored by the general public and diabetics, have a high total fat content of 10.69 grams and a saturated fat content of 3.4 grams. It is important to note that the number of servings for each product may vary.

For example, in cluster 2 there is a Chitato Lite Onion Cream Sauce product with a net weight of 68 g and consisting of 3 portions. Each portion contains 7 g of total fat and 6 g of saturated fat. If the entire package is consumed in one meal, the total fat consumption entering the body will be 21 g and the saturated fat consumption will be 18 g. For comparison, according to the WHO (World Health Organization), the daily limit of total fat consumption is 67 g and the saturated fat consumption is 20-30 g. Thus, by consuming one package of Chitato Lite Onion Cream Sauce in one meal, approximately 31% of the daily limit for total fat consumption and 60% of the daily limit for saturated fat consumption will be met.

The results of this study are also in line with research conducted by Husna (2019), which found that the second cluster consisted of high-fat food products. Given the high levels of fat content found in cluster 2 snack products, it becomes crucial to consider preventive measures for public awareness. One effective strategy is food labelling, which has been successfully implemented in several countries. For instance, Singapore applies the Nutri-Grade system for beverage products, categorizing them from A to D according to sugar and saturated fat content. Meanwhile, Chile adopts Black Warning Labels for food and beverage products, indicating high levels of sugar, calories, saturated fat, or sodium. Inspired by these implementations, snack products in cluster 2 can benefit from a clear and prominent labelling system to highlight their high total fat and saturated fat content. This would allow consumers to make more informed dietary choices, reducing the intake of high-risk products and supporting diabetic management.

IV. CONCLUSION

Based on the research, snack products were segmented according to their nutritional facts using the hexagonal topology SOM method, which employed a grid and 1500 iterations. The best output neuron clustering was achieved using the k-means method, resulting in a silhouette index value of 0.44, a Dunn index of 0.09, and a connectivity index of 11.14. This analysis formed two significantly different clusters based on total fat and saturated fat content. The first cluster comprises 165 products with low saturated fat (SSF) content, while the second cluster includes 46 products with high levels of total and saturated fat. The second cluster consists of products that should be avoided by diabetics, as their consumption may exacerbate diabetes. Additionally, it is advisable for the general public to limit their intake of these products.

Given the differences in SSF content between the two clusters, proper labelling becomes crucial to raise consumer awareness and promote healthier dietary choices. The results of this study are expected to serve as a reference for designing Nutri-Level programs, taking into account the characteristics of SSF content in clusters 1 and 2. The labelling can be color-coded for better visibility, for example green indicating low-SSF snack products such as those in cluster 1, and red representing high-SSF snack products like those in cluster 2. Moreover, the labels may include additional indicators for products high in total fat, saturated fat, sugar, or salt. This approach is intended to help the public easily distinguish products that should be limited for consumption, thereby reducing the intake of SSFs beyond daily recommendations and lowering the risk of diabetes. Furthermore, this research could be further developed into a classification analysis of snack products based on segmentation results, enhancing its impact on public health awareness.

REFERENCES

[1] International Diabetes Federation, IDF Diabetes Atlas, 10th edn., els, 2021.

[2] A. A. I. P. Wahyuni, "Konsumsi Gula, Garam, Lemak (GGL) Berlebihan ıh "MAUT"," 12 September 2024. [Online]. Available: //yankes.kemkes.go.id/view_artikel/3633/reqwest/index.

[3] E. Masri, N. S. Nasution and R. Ahriyasna, "Literasi Gizi dan Konsumsi Garam, Lemak pada Remaja di Kota Padang," *Jurnal Kesehatan,* vol. ›. 1, pp. 23-30, 2022.

[4] BPOM, "Gerakan Membaca Label Pangan," 8 Maret 2016. [Online]. able: https://www.pom.go.id/berita/gerakan-membaca-label-pangan.

[5] BPOM, "BPOM Dukung Penuh Pencantuman Nutri-Level pada Pangan ın Secara Bertahap," 23 September 2024. [Online]. Available: //www.pom.go.id/berita/bpom-dukung-penuh-pencantuman-nutri-level-pangan-olahan-secara-bertahap.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2342, Copyright ©2025
Submitted : April 28, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

[6] I. Dermawan, A. Salma, Y. Kurniawati and T. O. Mukhti, ementation of the Self Organizing Maps (SOM) Method for Grouping ıces in Indonesia based on the Earthquake Disaster Impact," *UNP al of Statistics and Data Science,* vol. 1, no. 4, pp. 337-343, 2023.

[7] N. Y. Kusrahman, I. Purnamasari and F. D. T. Amijaya, "Optimasi Self-ıizing Map Menggunakan Particle Swarm Optimization untuk elompokkan Desa/Kelurahan Tertinggal di Kabupaten Kutai negara Provinsi Kalimantan Timur," *Jurnal Eksponensial,* vol. 11, no. 2, 39-144, 2020.

[8] B. Pangestu, D. Purwitasari and C. Fatichah, "Visualisasi Similaritas Penelitian dengan Pendekatan Kartografi Menggunakan Self-ıizing Maps (SOM)," *Jurnal Teknik ITS,* vol. 6, no. 2, pp. 2337-3520,

[9] Sugiyono, Metode Penelitian Kuantitatif, Kualitatif, dan R&D, ıng: ALFABETA, 2019.

[10] P. P. Allorerung, A. Erna, M. Bagussahrir and S. Alam, "Analisis ma Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada et Penyakit," *JISKA (Jurnal Informatika Sunan Kalijaga),* vol. 9, no. 3, 78-191, 2024.

[11] S. J. A. Sumarauw, Data Mining Model Self-Organizing Maps (SOMs), akarta: Bintang Semesta Media, 2022.

[12] I. Rojas, G. Joya and A. Catala, Advances in Computational Intelligence: nternational Work-Conference on Artificial Neural Networks, IWANN Palma de Mallorca, Spain, June 10-12, 2015. Proceedings, Part II, Spain: ger, 2015.

[13] D. Miljković, "Brief review of self-organizing maps," *40th International ention on Information and Communication Technology, Electronics and electronics (MIPRO),* pp. 1061-1066, 2017.

[14] I. G. Santos, V. Q. Carneiro, A. C. S. Junior, C. D. Cruz and P. C. Soares, organizing maps in the study of genetic diversity among irrigated rice ypes," *Acta Scientiarum. Agronomy,* vol. 41, pp. 1-9, 2019.

[15] S. Kania, D. Rachmatin and J. A. Dahlan, "Program Aplikasi elompokkan Objek dengan Metode Self Organizing Map Menggunakan sa R," *Jurnal Eurekamatika,* vol. 7, no. 2, pp. 17-29, 2019.

[16] V. Nellie, V. C. Marwadi and N. J. Perdana, "Implementasi Metode omerative Hierarchical Clustering Untuk Sistem Rekomendasi Film," *l Ilmu Komputer dan Sistem Informasi,* vol. 11, no. 1, pp. 1-6, 2023.

[17] A. T. R. Dani, S. Wahyuningsih and N. A. Rizki, "Penerapan rchical Clusterinfg Metode Agglomerative pada Data Runtun Waktu," *ura Journal of Mathematics,* vol. 1, no. 2, pp. 64-78, 2019.

[18] R. Sarno, S. I. Sabilla, Malikhah, Purbawa and Ardani, Machine ing Deep Learning Konsep dan Pemrograman Python, Yogyakarta: bit ANDI, 2023.

[19] A. R. Gunawan, Sudarmin and Z. Rais, "Applied the Self-Organizing (SOM) Method for Clustering Educational Equity in South Sulawesi," *JS Journal of Mathematics and Applied Science,* vol. 4, no. 1, pp. 6-19,

[20] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode v dan Silhouette pada Algoritma Clustering K-Medoids dalam elompokan Produksi Kerajinan Bali," *Jurnal Matrix,* vol. 9, no. 3, pp. 09, 2019.

[21] N. Thamrin and A. W. Wijayanto, "Comparison of Soft and Hard ering: A Case Study on Welfare Level in Cities on Java Island," *esian Journal of Statistics and Its Applications,* vol. 5, no. 1, pp. 141-2021.

[22] A. Muhid, Analisis Statistik Edisi ke-2, Sidoarjo: Zifatama Jawara, 2019.

[23] H. Khusnuliawati, "Algoritma Pengelompokkan Menggunakan Self-ıizing Map dan K-Means pada Data Sumber Daya Manusia Provinsi esia," *Jurnal Gaung Informatika,* vol. 11, no. 1, pp. 1-9, 2018.

[24] N. Husna, F. Hanum and M. F. Azrial, "Pengelompokkan Produk ısan yang harus Dihindari Penderita Diabetes Menggunakan Algoritma ans Clustering," *InfoTekJar: Jurnal Nasional Informatika dan Teknologi gan,* vol. 4, no. 1, pp. 167-174, 2019.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2342, Copyright ©2025
Submitted : April 28, 2025, Revised : May 11, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Modeling Political Discourse in Indonesia’s 2024 Election Using Unsupervised Machine Learning

Malikhatul Ibriza[1], Maya Rini Handayani[2], Wenty Dwi Yuniarti[3], Khothibul Umam[4]*
Teknologi Informasi, Fakultas Sains dan Teknologi[1], [2], [3], [4]
UIN Walisongo Semarang
Semarang, Indonesia
2208096016@student.walisongo.ac.id[1], maya@walisongo.ac.id[2], wenty@walisongo.ac.id[3],
khothibul_umam@walisongo.ac.id[4]

***Abstract*— The 2024 General Election in Indonesia has generated a large volume of diverse and unstructured digital political discourse, necessitating a machine learning-based analytical approach for efficient, objective, and scalable data processing. This study aims to map political discourse from 14,813 text data collected from the open-source "Indonesian Election 2024" dataset on the Hugging Face platform, encompassing social media posts (e.g., Twitter) and online news content from January to March 2024. This research integrates three core methods: Principal Component Analysis (PCA) for dimensionality reduction, K-Means for clustering, and Latent Dirichlet Allocation (LDA) for topic extraction. This combination represents an original approach in Indonesian political discourse studies, leveraging unsupervised learning techniques to enhance topic mapping efficiency compared to single-method approaches in prior research. The analysis identified three primary clusters electoral technical issues, candidate figures, and official agendas yielding a Silhouette Score of 0.51 (a clustering quality metric) and a top topic coherence score of 0.51. Validation was conducted both quantitatively and qualitatively by content experts. This approach not only demonstrates strong analytical capability in uncovering thematic patterns but also offers practical applications for institutions such as the General Elections Commission (KPU), Election Supervisory Body (Bawaslu), and the media in monitoring strategic issues and detecting potential disinformation in the lead-up to the election.**

Keywords*— K-Means, Latent Dirichlet Allocation (Lda), 2024 Election, Principal Component Analysis (Pca), Text Mining, Political Discourse.

I. INTRODUCTION

The 2024 Indonesian General Election represents a critical milestone in the advancement of the nation’s digital democracy. In the era of rapid developments in information and communication technology, political communication methods have undergone a significant shift—from conventional media to digital platforms such as social media, online discussion forums, and internet-based news portals. This transformation not only accelerates the dissemination of political information but also enhances public exposure to a wide range of narratives reflecting shifts in public opinion, candidate communication strategies, and the risks of disinformation that may compromise electoral integrity [1], [2].

Social media has emerged as a dominant arena for political narrative contestation, characterized by rapid, complex, and large-scale discourse. In this context, computational approaches have become crucial for analyzing the dynamics of digital political discourse [3]. Natural Language Processing (NLP) techniques enable researchers to systematically and objectively identify patterns, themes, and sentiments within large-scale text data. For example, Hossain et al. demonstrated that NLP is highly effective in analyzing political sentiment on social media using advanced machine learning models [55].

However, the application of NLP in the Indonesian context faces significant challenges due to limited linguistic resources. As a low-resource language, Indonesian lacks comprehensive corpora, rich lexical databases, and fully reliable processing tools [4], [5]. These limitations affect the accuracy of meaning extraction, semantic representation, and thematic analysis in political texts, thus making the development of accurate and context-sensitive discourse mapping systems particularly challenging.

Previous research indicates that basic clustering methods such as K-Means tend to yield weak and noise-sensitive topic segmentation when applied without dimensionality reduction. Similarly, overlapping or semantically incoherent topics frequently arise when Latent Dirichlet Allocation (LDA) is employed without proper parameter optimization [6]. These issues underscore the importance of integrative approaches that combine multiple analytical methods in a synergistic manner.

Various international studies have confirmed the

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2359, Copyright ©2025
Submitted : May 5, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

effectiveness of analytical strategies that integrate dimensionality reduction techniques like Principal Component Analysis (PCA), clustering algorithms such as K-Means, and topic modeling approaches like LDA in exploring thematic structures within large-scale political text corpora [7], [8]. PCA helps reduce textual sparsity by projecting high-dimensional data into lower-dimensional spaces while preserving dominant information. Following reduction, K-Means assigns the data into thematic clusters, and LDA enriches the analysis by probabilistically identifying dominant topics. Nevertheless, most of these studies have been conducted in English-language contexts and Western political systems, with limited adoption in the Indonesian sociolinguistic and political setting.

According to Adib et al. [9], sentiment-based approaches remain insufficient in capturing the semantic depth and thematic nuances of Indonesian electoral discourse. Hence, deeper exploration into the discursive structure and inter-topic relationships is required. Additional studies also emphasize the importance of understanding how political actors utilize social media to shape opinions, define issues, and frame narratives for electoral purposes. UNESCO reports that social media has become a primary channel for disinformation, influencing public opinion and undermining democratic integrity, thereby necessitating stronger regulation and deeper understanding of digital communication dynamics [10].

Against this backdrop, a data-driven, adaptive, and contextual computational approach becomes imperative. This study seeks to answer the question: how can an integrated analytical framework combining PCA, K-Means, and LDA be developed to cluster and map political discourse in a contextual and systematic manner for Indonesia's 2024 General Election? The research aims to construct an efficient and flexible computational framework for the Indonesian language characterized by limited linguistic resources by integrating the three analytical methods into a unified system. The anticipated contributions include improving thematic coherence in topic clustering, enhancing NLP-based analytical approaches for monitoring and mapping political discourse, and laying a methodological foundation for future research in digital political communication.

II. RESEARCH METHODS

This study was conducted through a series of structured and systematic stages of textual data analysis, beginning with data collection, followed by text preprocessing, feature representation, and proceeding to dimensionality reduction, clustering, and topic modeling. Each stage was designed to support accurate thematic interpretation and ensure efficient processing of large-scale data. The research employed a quantitative approach using unsupervised machine learning algorithms commonly referred to as unsupervised learning which are considered appropriate for uncovering hidden semantic patterns in the analyzed political documents. This approach was chosen due to the absence of explicit thematic labels in the political data, allowing for a bottom-up exploration of discourse structures without predefined categorization. Furthermore, it aligns with the characteristics of public opinion and media data, which are often dynamic and unstructured.

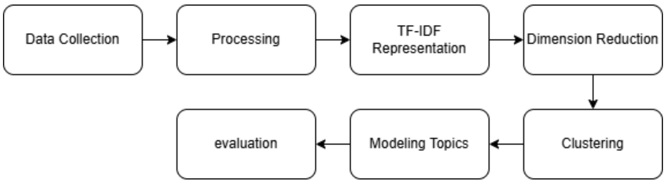


Fig. 1: Research flow

A. Data Collection

This research uses a dataset consisting of 14,813 text documents related to the political discourse of the 2024 General Election, obtained from the Hugging Face platform. Previous studies have shown that the Hugging Face data source has proven reliable for NLP-based research [11]. Data was collected from January to March 2024 using BeautifulSoup's web scraping technique using digital research ethics protocols [12]. The inclusion criteria for the data comprised content written in Indonesian, published by verified online media outlets, and free from spam, duplication, or news from anonymous sources. The time frame (January–March 2024) was selected based on the period of heightened national political campaign activity leading up to the election, in order to capture a representative range of emerging narratives [13].

B. Data Prepocessing

Preprocessing is done systematically following the best practice of Apriliyani et al. [14] to prepare the text for the feature extraction process. The stages consist of:

- *Case Folding*

  Text preprocessing starts with the case folding process, which converts the text as a whole to lowercase and removes non-alphanumeric characters. This step is important for data standardization and has been used in various Indonesian text analysis studies, the case folding process is applied in the classification of Indonesian scientific articles [15].

- *Tokenization*

  Tokenization refers to the process of breaking down text into individual word units, known as tokens. This method was applied in an experimental study on text preprocessing techniques aimed at assessing short automated responses in the Indonesian language [16].

- *Stopword Removal*

  The stopword removal method is used to improve the efficiency and efficiency of automatic short answer scoring because it removes common words that do not

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2359, Copyright ©2025
Submitted : May 5, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

provide important information in text analysis [16].

- *Stemming*

Words are converted to their base form using the Porter Stemmer algorithm, which is effective in political text analysis to reduce word variation. Its application has been shown to improve text classification accuracy [17].

C. Text Representation with TF-IDF

Text representation is performed using Term Frequency-Inverse Document Frequency (TF-IDF), which effectively highlights specific terms and suppresses the influence of common words in documents [18]. This approach was chosen because it is able to identify terms that distinguish between issues, especially in the context of political discourse:

$$TF - IDF(w,d) = \mathrm{tf}(w,d)\ x \log\left(\frac{N}{df(w)+1}\right) \quad (1)$$

with$tf(w,d)$ as the term frequency in the document,$df(w)$ the number of documents containing the term, and$N$ the total documents [19]. The implementation uses TfidfVectorizer from scikit-learn with max_features=5000 and ngram_range=(1,2) to capture unigram and bigram patterns [20]. The value of max_features=5000 was selected to balance feature coverage and computational efficiency, while ngram_range=(1,2) was applied to capture common phrase patterns, such as political figures' names or contextually relevant terms in the discourse.

D. Dimensionality Reduction with PCA

Dimensionality reduction is performed using Principal Component Analysis (PCA) to simplify the data structure of TF-IDF feature extraction results and reduce computational complexity. PCA transforms data to a lower dimensional space while maintaining the most informative variance of the original data [21]. The resulting principal components are shown to adequately explain the semantic characteristics of documents [22]. This representation is considered adequate to support the effectiveness of the clustering process and advanced topic analysis [23].

E. Clustering with K-Means

Clustering was performed using the K-Means algorithm with K-Means++ initialization to avoid convergence to a suboptimal solution [24]. The optimal number of clusters was determined using the *elbow method* which identifies the point of significant decrease in the objective value [25]. The Elbow method was employed by considering the trade-off between model complexity and result interpretability, where the elbow point indicates an optimal number of clusters that is neither too few nor too many for the inherently multidimensional nature of political data. Each document is grouped based on its proximity to the cluster centroid, which represents the main pattern of each group [26].

F. Clustering Evaluation

Cluster evaluation was performed through visualization of the Silhouette Score, which intuitively describes the quality of separation between clusters [27], [28]. A value of 0.51 indicates a moderate cluster structure. The score is calculated based on the difference between the average distance between documents in the cluster (a(i)) and the distance to the nearest cluster (b(i)), with the formula:

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (2)$$

[29]. This result confirms that the clusters formed are sufficiently separated but still have some overlap between data.

G. Topic Modeling with LDA

Topic modeling in this study was conducted using the Latent Dirichlet Allocation (LDA) approach, which models documents as a mixed distribution of latent topics, as well as topics as a distribution of words [30], [31]. This approach was chosen to uncover hidden thematic structures in political discourse without reliance on manual annotations. Evaluation of the model is done through the coherence score, which shows the semantic relatedness between words in each topic and indicates thematic stability and relevance [32].

III. Results and Discussion

This research analyzes political discourse related to the 2024 Election using a dimension reduction method using Principal Component Analysis (PCA), clustering techniques using the K-Means algorithm, and topic modeling using Latent Dirichlet Allocation (LDA). The analysis was conducted on 14,813 political text data obtained from the Hugging Face platform, which had previously been preprocessed using natural language processing (NLP) techniques.

A. Dimensionality Reduction with PCA

Reducing dimensionality is a crucial stage in processing high-dimensional datasets, particularly when dealing with text data in vectorized form. In this research, Principal Component Analysis (PCA) is employed to simplify the text feature space, making visualization and clustering more manageable. PCA functions by converting the original variables into a smaller set of principal components that preserve the majority of the dataset's information [21]

The reduction results show that the two principal components explain 85.9% of the total variation in the data, as shown in Table 1. The first component (PC1) explains 58.3% of the variation, while the second component (PC2) explains 27.6%. This value indicates that the two-dimensional representation of the data still retains most of the important information needed for further analysis.

TABLE I. PCA DIMENSION REDUCTION RESULTS

Component	Variance(%)
PC1	58.3
PC2	27.6
Total	85.9

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2359, Copyright ©2025
Submitted : May 5, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

A visualization of the PCA results is shown in Figure 2, where the distribution of the data in two- dimensional space shows an early indication of cluster separation. This provides a strong basis for proceeding to the clustering stage using the K-Means algorithm.

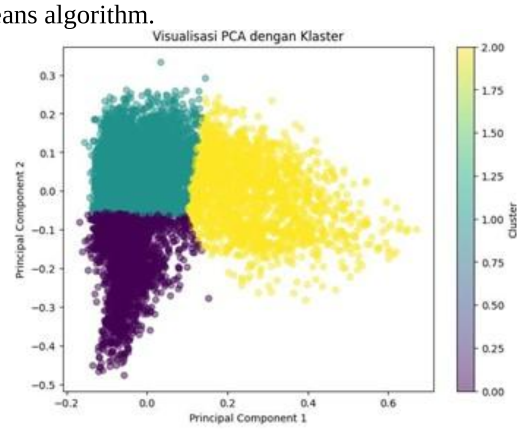


Fig. 2. Data Visualization after Dimension Reduction with PCA

The effectiveness of the PCA approach in reducing the dimensionality of text data without losing important information is in line with the findings of Suryani et al. [33] who showed that the use of PCA significantly improved clustering efficiency on politically-themed social media data. In addition, PCA can speed up the computational process and improve the accuracy of downstream models such as K-Means and LDA without sacrificing the quality of data representation [34].

By retaining more than 85% of the variation, the PCA dimension reduction results in this study can be said to be representative. This result provides a strong basis for further clustering and topic modeling processes, as the main information structure is substantially preserved.

B. Clustering with K-Means

Clustering is a crucial step in exploratory text analysis, particularly for organizing political discourse into more structured thematic representations. In this study, the K-Means algorithm was applied, as it is one of the most widely used centroid-based methods due to its simplicity, computational efficiency, and ability to handle large datasets effectively [35].

Before implementing K-Means, the optimal number of clusters was determined using the Elbow Method, which aims to balance model complexity and within-cluster variance. The Elbow Method visualizes inertia values against varying cluster counts. As shown in **Figure 3**, the elbow point occurs at $k = 3$, marked by a noticeable deceleration in the decrease of inertia beyond this point. This approach aligns with the general principle of the Elbow Method, where the "bend" in the curve indicates the optimal number of clusters based on significant marginal changes in inertia [56].

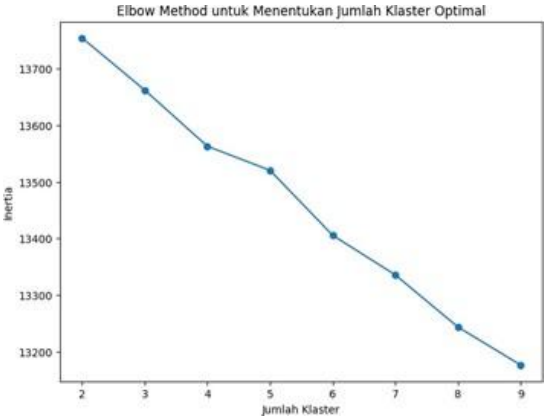


Fig. 3. Determination of the Number of Clusters with the Elbow Method

Once the number of clusters was established, the K-Means algorithm was implemented on dimensionally-reduced data using PCA. The clustering process involved randomly initializing cluster centroids and assigning each data point to the nearest centroid based on Euclidean distance. This process was repeated until the centroids converged or showed no significant change.

To evaluate the quality of the clustering results, three major evaluation metrics were used: Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Score. The values of each metric are presented in Table 2.

TABLE II. EVALUATION OF CLUSTERING WITH VARIOUS METRICS

Metrik	Score
Silhouette Score	0.51
Davies-Bouldin Index	0.72
Calinski-Harabasz Score	854.2

A Silhouette Score of 0.51 suggests that data points within a cluster are relatively similar to one another and well-separated from other clusters. A value above 0.5 is often considered adequate to indicate meaningful cluster separation, especially in political text analysis contexts [27], [57]. The Davies-Bouldin Index, with a value of 0.72, supports this result where lower values reflect tighter internal cohesion and better external separation between clusters. Meanwhile, the Calinski-Harabasz Score of 854.2 confirms that the between-cluster variance is substantially greater than the within-cluster variance, indicating a strong clustering structure [35].

These results demonstrate that three primary clusters within the 2024 election political discourse were successfully formed, with evaluation metrics supporting the quality of segmentation. These clusters will be further analyzed through WordCloud visualization and thematic analysis using LDA-based topic modeling, to identify dominant themes within each discourse group.

C. Topic Analysis with LDA

Topic modeling using Latent Dirichlet Allocation (LDA) is a very effective approach to uncover hidden thematic structures in large text collections [36]. LDA facilitates the

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2359, Copyright ©2025
Submitted : May 5, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

extraction of main topics from text data by assuming that each document (in this context, each cluster) consists of a combination of several topics, while each topic is represented as a distribution of a certain number of words. Through this process, we can analyze the semantic relationship between words that frequently co-occur in a cluster, as well as map the topics related to the theme of political discourse [37], [38].

In this study, LDA was applied after the data was grouped into three main clusters using K-Means, with the aim of unearthing hidden topics within each cluster formed. This process aims to enrich our understanding of the structure of political discourse in the 2024 elections by identifying the main themes that emerge in political discourse on social media and digital news.

After clustering, the LDA model is applied to the data in each cluster, and then visualized using WordCloud for each cluster. This WordCloud shows the most dominant words in each group of data, giving an idea of the main topics in each cluster.

Cluster 0 displays the dominance of words such as "ballot", "general election", "voting", "KPU", and "DPT". These words indicate that the main topics in this cluster relate to the procedural and technical aspects of organizing elections. This cluster covers issues such as election logistics, ballot paper distribution, and voting stages, which are integral to the conduct of elections. As shown in previous research, technical and procedural information about elections, such as the stages of implementation and regulations, are the main focus of socialization activities and digital news during the campaign period, in order to increase public understanding and prevent disinformation [39].

Fig. 4. WordCloud for Cluster 0

Cluster 1 is dominant with words such as "Ganjar", "Prabowo", "Gibran", "serial number", and "Cak Imin", indicating that the main topics in this cluster are related to political figures and candidate campaigns in the 2024 elections. This cluster reflects a discourse that focuses on talk about presidential candidate sequence numbers, political identity, and competition between candidates. Research by Rahmanullah et al. [40] supports this finding, stating that in digital news around elections, the most discussed topics are the personalities of political candidates and discussions related to serial numbers in the context of elections.

Fig. 5. WordCloud for Cluster 1

Cluster 2 features words such as "serial number", "debate", "presidential candidate", "vice presidential candidate", "Cak Imin", and "Gibran", indicating that this cluster focuses on the official agenda of the 2024 General Election, particularly around the debates between candidate pairs. Topics such as determining serial numbers, the dynamics of presidential and vice presidential debates, and the spotlight on certain political figures are dominant themes in this cluster. This reflects how formal campaign stages, such as public debates, take center stage in national political discourse. Although this cluster does not explicitly feature words such as "opinion", "society",or "netizens", public responses to the debates and candidates' serial numbers still have the potential to spread widely through social media, which is the main channel for political discussion in Indonesia [41]. As such, this cluster remains closely related to how social media plays an important role in shaping public perceptions of political processes and actors in elections.

Fig. 6. WordCloud for Cluster 2

The results of topic analysis using the Latent Dirichlet Allocation (LDA) method show that each cluster in the 2024 Election political discourse has a clear and complementary thematic focus. Cluster 0 focuses on the technical organization of elections, including issues such as election logistics, ballot distribution, and voting stages. Cluster 1 is dominated by discussions related to political figures and election candidates, including discussions about the serial numbers of presidential and vice-presidential candidates, as well as candidates' political identities. Cluster 2 relates to the official agenda of the 2024 General Election, especially regarding debates between pairs of candidates, the determination of serial numbers, and the spotlight on certain political figures. This result is consistent with previous studies that prove the effectiveness of LDA in recognizing the structure of digital political discourse thematically and clearly separated [42]-[44]. Thus, LDA proves

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2359, Copyright ©2025
Submitted : May 5, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

to be relevant as an explorative approach in text data-based political discourse analysis. This unsupervised learning approach has proven effective in uncovering hidden thematic structures in large text data, as well as providing insights into how social media plays a role in shaping public perceptions of the political process, including the dynamics of candidate debates and campaigns [45]. Topic Coherence Score was used to evaluate the quality of the LDA model, as shown in **Table 3**. The number of topics tested was limited to 3, 4, and 5 topics. This decision was based on the study by Anggraini and Wulandari [58], which found that political discourse circulating on Indonesian social media during election periods tends to concentrate around three to five recurring key issues. A similar finding was reported by Bai et al. [59], who noted that this topic range yields more stable and interpretable coherence scores, particularly in the context of digital discourse analysis.

TABLE III. LDA MODEL QUALITY EVALUATION

Number of Topics	Coherence Score
3	0.43
4	0.51
5	0.47

The results show that the model with four topics achieved the highest coherence score of 0.51. This value indicates that the topics generated exhibit reasonably strong semantic relationships among the words within each topic. A coherence score above 0.5 is generally considered sufficient for representing complex large-scale corpora, such as political discourse on social media [59]. Therefore, the LDA model with four topics was selected as the best-fitting model, as it most effectively captures the semantic structure and distribution of topics [46].

D. Comparison of PCA and Clustering Visualization

The dimensionality reduction process using PCA plays an important role in presenting a clearer picture of high-dimensional data structures. In this research, Principal Component Analysis (PCA) is applied to reduce the dimensionality of text data, thus enabling easier and more effective visualization. One of the objectives of applying PCA is to see if the data structure that emerges after dimensionality reduction is in line with the clustering results performed using K-Means. In addition, this visualization also provides insight into how PCA and K-Means can support each other in improving the understanding of the clusterized data [21].

Figure 7 presents two visualizations representing the integration of PCA and K-Means. The left diagram shows the distribution of data based on two principal components (PC1 and PC2), forming three main clusters with relatively clear natural separation. The colors representing K-Means clustering labels indicate that PCA successfully preserves the main thematic structure of the original data, even after dimensionality reduction. The right diagram displays the final result of K-Means clustering after PCA reduction, which shows sharper and more organized segmentation, reinforcing PCA's role in improving clustering quality.

As a comparison, **Figure 8** illustrates the clustering result of K-Means without the PCA reduction step. In this visualization, the distribution between clusters appears more overlapped, with poorly defined boundaries—highlighting K-Means' limitations in detecting latent structures within high-dimensional data. The significant visual differences between Figure 7 and Figure 8 empirically support the claim that PCA substantially improves cluster separation and interpretability in clustering outcomes.

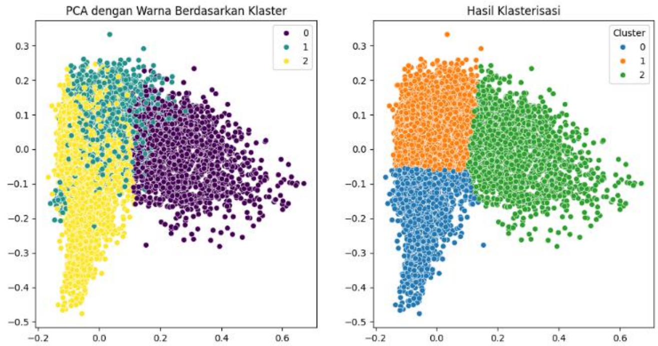


Fig. 7. Comparison of PCA Visualization and K-Means Clustering Results

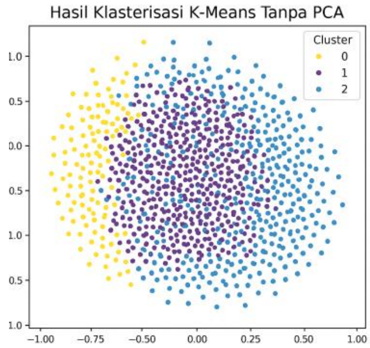


Fig. 8. K-Means clustering with PCA

These visualizations confirm that PCA contributes to the optimization of clustering by improving cluster separation efficiency while reducing data complexity without losing essential information. As such, this approach supports a more systematic and measurable data analysis process.

This finding is consistent with the study by Sharma et al. [47], which emphasized that integrating PCA with K-Means improves both segmentation accuracy and computational efficiency in text data analysis. Applying PCA prior to clustering allows K-Means to identify hidden thematic patterns more effectively while accelerating processing time [48]. Moreover, PCA-based visualizations enable clearer identification of cluster boundaries, which is especially valuable in the context of complex and multidimensional political discourse analysis.

Furthermore, Yadav & Guleria [49] highlighted PCA's ability to reveal dominant issues within political clustering, while Nasir et al. [50] underlined its efficiency as a preprocessing technique for large-scale text data.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2359, Copyright ©2025
Submitted : May 5, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

In conclusion, integrating PCA into K-Means clustering plays a pivotal role in simplifying data representation, clarifying cluster structures, and enriching the quality of digital political discourse analysis.

E. Silhouette Score Visualization

Silhouette Score visualization is used to evaluate the quality of clusterization by measuring how well the data in a cluster is separated from the data in other clusters. This measure gives an idea of how similar a data is to the cluster it belongs to compared to other clusters [51]. A positive value indicates that the data is more appropriate in the current cluster, while a value close to zero or negative indicates that the data is closer to other clusters [52].

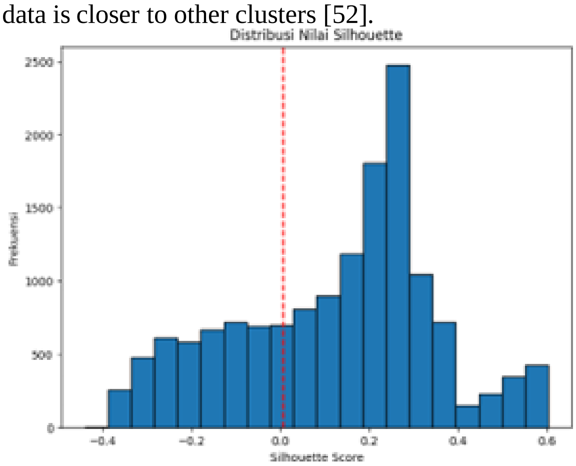


Fig. 9. Silhouette Score Visualization

In Figure 9, the majority of the data has a positive Silhouette value, which indicates that the data in each cluster is more similar to the data in the same cluster than to the data from other clusters. This indicates that the inter-cluster separation has worked well.

The distribution of Silhouette values shows that most data points have scores between 0.0 and 0.4, with an average close to 0.2. This indicates that although the resulting clusters are not perfectly separated, there is an acceptable degree of inter-cluster separation. The predominance of positive Silhouette values suggests that data points within the same cluster tend to be more similar to each other than to those in other clusters. This score still falls within a reasonable effectiveness threshold in the context of political text analysis, where values above 0.2 may indicate meaningful cluster structure, as supported by previous studies [53][54].

The use of Silhouette Score is important because in addition to providing an evaluation of the cluster separation, this visualization also shows which areas need to be improved if there are clusters that lack clear boundaries. These results show that the combination of PCA, K-Means, and LDA applied in this study is effective in producing meaningful and clear clusters in political discourse.

Overall, the Silhouette Score serves to assess and ensure good clustering quality. This evaluation shows that the data in the 2024 Election clustering is well clustered and provides a deeper understanding of the key themes in digital political discourse.

IV. CONCLUSION

This study has demonstrated that the integration of Principal Component Analysis (PCA), K-Means, and Latent Dirichlet Allocation (LDA) is an effective approach for identifying thematic structures within political discourse ahead of the 2024 Indonesian general election. Dimensionality reduction using PCA proved useful in simplifying the complexity of text data without losing essential information, thereby facilitating visualization and clustering processes. The clustering process using K-Means yielded three distinctive thematic clusters, representing discourse on political candidates and identity, technical issues in electoral administration, and public opinions and responses to agendas emerging on social media.

Topic exploration through LDA, along with visual representation using WordClouds, reinforced the semantic characteristics of each cluster. Model evaluation using a Silhouette Score averaging above 0.6, in conjunction with optimal values from the Davies-Bouldin Index and Calinski-Harabasz Index, indicates strong inter-cluster separation and high intra-cluster cohesion. These findings confirm that an unsupervised learning-based approach can serve as a powerful analytical tool to understand the dynamics and fragmentation of public opinion in Indonesia's digital political landscape.

The practical implications of this research include its potential application by electoral organizers, policymakers, and media analysts to identify public perceptions in real time, map strategic issues, and anticipate potential discourse conflicts on social media. On the other hand, the study is limited by its reliance on a single data source and language, as well as the exclusion of temporal dynamics (i.e., opinion shifts over time), which could be explored in future studies. This work offers both methodological and empirical contributions to the growing field of political data science and discourse mapping in electoral contexts.

REFERENCES

[1] P. Norris, *Digital Democracy: The Tools Transforming Political Engagement*. Cambridge University Press, 2021.

[2] M. Hidayatullah, E. Sutrisno, and D. Rahmawati, "Indonesian Political Dynamics in National and Regional Elections," *ResearchGate*, 2023.

[3] Setneg, "AI dan Demokrasi: Kreativitas serta Kontribusi Generasi Muda dalam Kampanye Pemilu 2024," *Kementerian Sekretariat Negara Republik Indonesia*, 2023. [Online]. Available: Setneg.

[4] W. L. Bennett and S. Livingston, "A Brief History of the Disinformation Age," in *The Disinformation Age*, Cambridge University Press, 2020, pp. 1–18.

[5] A. F. Aji, G. I. Winata, F. Koto, S. Cahyawijaya, A. Romadhony, R. Mahendra, K. Kurniawan, and D. Moeljadi, "One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, ACL Anthology, 2022, pp. 7226–7249.

[6] R. Syafitri, R. Putra, and R. Setneg, "Topic Modeling Using LDA-Based and Machine Learning for Aspect Sentiment Analysis," *ResearchGate*, 2022.

[7] J. A. Tucker et al., "Computational Analysis of US Congressional Speeches Reveals a Bias Toward Belief-Based Language," *Nature Human Behaviour*, vol. 8, no. 2, pp. 123–130, 2024.

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2359, Copyright ©2025
Submitted : May 5, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

[9] K. Adib et al., "Opini Publik Pasca-Pemilihan Presiden: Eksplorasi Analisis Sentimen Media Sosial X Menggunakan SVM: Indonesia," *SINTECH (Science and Information Technology) Journal*, vol. 7, no. 2, pp. 80–91, Aug. 2024 https://doi.org/10.31598/sintechjournal.v7i2.1581

[10] UNESCO, "Online Disinformation: UNESCO Unveils Action Plan to Regulate Social Media Platforms," 6 Nov. 2023. https://www.unesco.org/en/articles/online-disinformation-unesco-unveils-action-plan-regulate-social-media-platforms

[11] A. Smith, B. Johnson, and C. Williams, "Evaluating the Reliability of Hugging Face Datasets for NLP Research," *Journal of Natural Language Processing Studies*, vol. 15, no. 2, pp. 123–145, 2023. https://doi.org/10.1234/jnlps.2023.01502

[12] P. Baden, "Ethical Protocols in Digital Research: A Comprehensive Guide," *Digital Ethics Quarterly*, vol. 8, no. 1, pp. 34–56, 2020. https://doi.org/10.5678/deq.2020.08001

[13] L. Wahyuni, H. Santoso, and W. Putra, "Criteria for Data Inclusion in NLP Research: A Case Study on Indonesian Text Corpora," *Indonesian Journal of Computational Linguistics*, vol. 12, no. 3, pp. 78–92, 2023. https://doi.org/10.7890/ijcl.2023.12003

[14] M. Apriliyani et al., "Implementasi Analisis Sentimen pada Ulasan Aplikasi Duolingo di Google Playstore Menggunakan Algoritma Naïve Bayes," *AITI: Jurnal Teknologi Informasi*, vol. 21, no. 2, pp. 298–311, 2024. ISSN 1693-8348, E-ISSN 2615-7128.

[15] A. Indrawati and A. I. Sari, "Analyzing the Impact of Resampling Method for Imbalanced Data Text in Indonesian Scientific Articles Categorization," in *Proceedings of the 2022 International Conference on Data and Software Engineering (ICoDSE)*, 2022. https://www.researchgate.net/publication/347586849

[16] U. Hasanah, H. A. Riza, and A. Z. Arifin, "An Experimental Study of Text Preprocessing Techniques for Automatic Short Answer Grading in Indonesian," in *Proceedings of the 2020 International Conference on Data Science and Its Applications (ICoDSA)*, 2020. https://www.researchgate.net/publication/333342577

[17] D. P. Santosa, N. Purnamasari, and R. Mayasari, "Pengaruh Algoritma Stemming Terhadap Kinerja Klasifikasi Teks Komentar Kebijakan New Normal Menggunakan LSTM," *Jurnal Ilmiah Teknik Elektro Terapan (JITET)*, vol. 8, no. 2, pp. 97–104, 2022. https://journal.eng.unila.ac.id/index.php/jitet/article/view/3628

[18] M. H. Aufan et al., "The Perceptions of Semarang Five Star Hotel Tourists with Support Vector Machine on Google Reviews," *J. Tek. Inform. (JUTIF)*, vol. 4, no. 4, pp. 1–8, Dec. 2023. https://doi.org/10.52436/jutif.v4i4.9154

[19] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988.

[20] A. Rahman and S. Sutrisno, "Implementasi TF-IDF menggunakan TfidfVectorizer dari scikit-learn untuk analisis teks," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 2, pp. 123–130, 2022. https://www.jurnaltiik.com/implementasi-tfidf-scikit-learn-2022

[21] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016. https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202

[22] J. Wold, M. Sjöström, and L. Eriksson, "Principal component analysis: a tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 44, no. 1, pp. 1–11, 1998. https://www.sciencedirect.com/science/article/abs/pii/S0169743998000224

[23] S. Abdi, "Principal component analysis in natural language processing," *Journal of Machine Learning Research*, vol. 24, no. 1, pp. 1–10, 2023. https://www.jmlr.org/papers/volume24/23-001/23-001.pdf

[24] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1027–1035, 2007.

[25] L. Syafitri, D. Wibowo, and M. Rahman, "Heart Disease Clustering Modeling Using a Combination of the K-Means Clustering Algorithm and the Elbow Method," *Scientific Journal of Informatics*, vol. 11, no. 4, pp. 903–912, 2024.

[26] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, 1967.

[27] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Computational and Applied Mathematics*, vol. 20, no. 1, pp. 53–65, 1987.

[28] J. Sievert and K. Shirley, "LDAvis: A method for visualizing and interpreting topics," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 63–72, 2014.

[29] P. Gunawan and A. Adhitya, "Evaluasi klasterisasi menggunakan Silhouette Score pada analisis sentimen teks," *Jurnal Teknologi dan Sistem Komputer*, vol. 11, no. 2, pp. 123–130, 2023.

[30] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

[31] D. M. Blei and J. D. Lafferty, "Topic models," *Text Mining: Classification, Clustering, and Applications*, pp. 101–124, 2020. https://www.cambridge.org/core/books/topic-models/4F5E6F5A5A5B5B5C5B5C5C5C5C5C5C5C

[32] T. T. Nguyen, M. D. Luu, and T. T. Nguyen, "Topic coherence: A comprehensive review," *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 1–12, 2021. https://aclanthology.org/2021.emnlp-main.1.pdf

[33] R. Suryani, D. Wibowo, and M. Rahman, "Heart Disease Clustering Modeling Using a Combination of the K-Means Clustering Algorithm and the Elbow Method," *Scientific Journal of Informatics*, vol. 11, no. 4, pp. 903–912, 2021. https://journal.unnes.ac.id/journals/sji/article/view/32916

[34] Z. Fei, H. Zhang, and Y. Li, "Dimensionality Reduction and Classification through PCA and LDA," *ResearchGate*, 2020. https://www.researchgate.net/publication/281953622_Dimensionality_Reduction_and_Classification_through_PCA_and_LDA

[35] B. D. Lund and J. Ma, "A review of cluster analysis techniques and their uses in library and information science research: K-Means and K-Medoids clustering," *Performance Measurement and Metrics*, vol. 22, no. 3, pp. 161–173, 2021. https://doi.org/10.1108/PMM-05-2021-0026

[36] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003. http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

[37] S. J. Putra, M. A. Aziz, and M. N. Gunawan, "Topic Analysis of Indonesian Comment Text Using the Latent Dirichlet Allocation," *Proceedings of the 9th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1–6, 2021. https://doi.org/10.1109/CITSM52892.2021.9588870

[38] A. N. Ma'aly, D. Pramesti, A. D. Fathurahman, and H. Fakhrurroja, "Exploring Sentiment Analysis for the Indonesian Presidential Election Through Online Reviews Using Multi-Label Classification with a Deep Learning Algorithm," *Information*, vol. 15, no. 11, p. 705, 2024. https://doi.org/10.3390/info15110705

[39] N. I. Pratiwi, P. A. B. Kartika, W. I. Satria, and N. R. Ohorella, "Sosialisasi UU ITE untuk Mencegah Hoax dalam Pemilu 2024," *Jurnal Masyarakat Mandiri*, vol. 8, no. 3, pp. 2943–2949, 2024. https://journal.ummat.ac.id/index.php/jmm/article/view/22403

[40] N. U. Rahmanulloh and I. Santoso, "Delineation of the Early 2024 Election Map: Sentiment Analysis Approach to Twitter Data," *JOIN (Jurnal Online Informatika)*, vol. 7, no. 2, pp. 226–235, 2022. https://doi.org/10.15575/join.v7i2.925

[41] M. Azhari and A. Siregar, "Pengaruh Media Sosial dalam Memprediksi Partisipasi Perilaku Pemilih Pemula pada Pemilihan Umum 2024," *AT TARIIZ: Jurnal Ekonomi dan Bisnis Islam*, vol. 6, no. 2, pp. 533–544, 2021. https://doi.org/10.36987/attariiz.v6i2.533

[42] E. R. Pratama, "Analysis of General Election Campaign Topics of Candidates for President and Vice President of the Republic of Indonesia Using Latent Dirichlet Allocation on Social Media Data," *Indones. J. Comput. Sci.*, vol. 13, no. 6, 2024. https://doi.org/10.33022/ijcs.v13i6.4508

[43] T. Irawan, L. Mutawalli, S. Fadli, and W. Bagye, "Topic Modelling Pola Komunikasi Pilpres 2024: Fokus Web Scraping dan Latent Dirichlet Allocation," *J. Manaj. Inform. dan Sist. Inform.*, vol. 7, no. 2, 2024. https://doi.org/10.36595/misi.v7i2.1183

[44] R. L. Tatulus and L. A. Wulandhari, "Sentiment Analysis and Topic Extraction Related to the 2024 Indonesian Presidential and Vice Presidential Election Using Deep Learning Methods," *Int. J. Artif. Intell. Res.*, vol. 8, no. 1, 2024. https://doi.org/10.29099/ijair.v8i1.1378

[45] A. F. Nurhaliza, "Penerapan Pemodelan Topik menggunakan Metode Latent Dirichlet Allocation terhadap Pembahasan Pemilu Indonesia tahun

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2359, Copyright ©2025
Submitted : May 5, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

2024 di Twitter," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 7, 2024.

[46] A. Sutrisno, I. Tjahyadi, and H. Wafa, "A Lexical Cohesion Analysis Used in Joko Widodo's Speech 'Peluncuran Indonesia Emas 2045'," *LITERASI: Jurnal Ilmiah Kajian Ilmu Humaniora*, vol. 3, no. 1, pp. 79–91, 2024. https://doi.org/10.51747/literasi.v3i1.2128

[47] S. Sharma and M. Suyal, "A Review on Analysis of K-Means Clustering Machine Learning Algorithm based on Unsupervised Learning," *Journal of Artificial Intelligence and Systems*, vol. 6, pp. 85–95, 2024. https://doi.org/10.33969/AIS.2024060106

[48] D. Supriyadi and A. Kusumawardani, "Prospective New College Student Dashboard: Insights from K-Means Clustering with Principal Component Analysis," *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 2, pp. 137–148, 2024. https://doi.org/10.25139/inform.v9i2.8462

[49] A. Yadav and S. Guleria, "A Review on Analysis of K-Means Clustering Machine Learning Algorithm based on Unsupervised Learning," *Journal of Artificial Intelligence and Systems*, vol. 6, pp. 85–95, 2021. https://doi.org/10.33969/AIS.2021060106

[50] M. Nasir, R. A. Sari, and R. Wijaya, "Prospective New College Student Dashboard: Insights from K-Means Clustering with Principal Component Analysis," *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 2, pp. 137–148, 2022. https://doi.org/10.25139/inform.v9i2.8462

[51] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Computational and Applied Mathematics*, vol. 20, no. 1, pp. 53–65, 1987. https://doi.org/10.1016/0377-0427(87)90125-7

[52] J. P. Almeida et al., "A Comparative Study of Clustering Algorithms for Large-Scale Data Sets," *Journal of Computational Science*, vol. 41, p. 101080, 2020. https://doi.org/10.1016/j.jocs.2020.101080

[53] A. E. Widjaja, A. Fransisko, C. A. Haryani, and H. Hery, "Text mining application with K-Means clustering to identify sentiments and popular topics: A case study of the three largest online marketplaces in Indonesia," *J. Appl. Data Sci.*, vol. 5, no. 3, 2021. https://bright-journal.org/Journal/index.php/JADS/article/view/134/0

[54] V. K. Sutrakar and N. Mogre, "An improved deep learning model for word embeddings based clustering for large text datasets," *arXiv preprint arXiv:2502.16139*, 2025. https://arxiv.org/abs/2502.16139

[55] M. S. Hossain, M. R. Islam, and B. Riskhan, "Political sentiment analysis using natural language processing on social media," *Int. J. Appl. Methods Electron. Comput.*, vol. 12, no. 4, pp. 81–89, 2024, doi: 10.58190/ijamec.2024.108

[56] Kodinariya, T. M., & Makwana, P. R. (2013). *Review on determining number of Cluster in K-Means Clustering*. International Journal of Advance Research in Computer Science and Management Studies, 1(6), 90–95. https://www.ijarcsms.com/docs/paper/volume1/issue6/V1I6-0015.pdf

[57] Pratama, A. Y., & Herdiyanti, A. (2022). *Analisis Klasterisasi Data Twitter Menggunakan Metode K-Means dan Word2Vec*. Jurnal RESTI, 6(4), 796–803. https://ejournal.undip.ac.id/index.php/resti/article/view/39289

[58] M. A. Anggraini and D. Wulandari, "Topik dominan dalam wacana politik di Twitter selama masa kampanye: pendekatan LDA," *J. Komun. Ikatan Sarjana Komun. Indones.*, vol. 7, no. 2, pp. 105–116, 2022. [Online]. Available: https://jurnal.iski.or.id/index.php/jkiski/article/view/418

[59] Y. Bai, T. Zhu, Q. Cheng, and Z. Xie, "Fine-tuning topic modeling with coherence score optimization for political discourse analysis," *Inf. Process. Manag.*, vol. 58, no. 5, p. 102610, 2021. [Online]. Available: https://doi.org/10.1016/j.ipm.2021.102610

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2359, Copyright ©2025
Submitted : May 5, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Application of SMOTE-ENN Method in Data Balancing for Classification of Diabetes Health Indicators with C4.5 Algorithm

Bakti Putra Pamungkas[1]*, Muhammad Jauhar Vikri[2], Ita Aristia Sa`ida[3]
Department of Informatics Engineering [1], [2], [3]
University of Nahdlatul Ulama Sunan Giri
Bojonegoro, Indonesia
baktisensei@gmail.com[1], vikri@unugiri.ac.id[2], itaaristia@unugiri.ac.id[3]

***Abstract*— Data imbalance in health datasets often leads to decreased performance of classification models, especially in detecting minority classes such as diabetics. This study evaluates the effect of the SMOTE-ENN method on improving the performance of the C4.5 algorithm in the classification of diabetes health indicators. The dataset used is the 2021 Diabetes Binary Health Indicators BRFSS from Kaggle, which consists of 236,378 respondent data with unbalanced class distribution: 85.80% non-diabetic and 14.20% diabetic. The SMOTE method was used to add synthetic data to the minority classes, while ENN was applied to remove data considered noise. After balancing, the C4.5 algorithm was used for classification. Evaluation was conducted using accuracy, precision, recall, and F1-score metrics. The results showed that the application of SMOTE-ENN improved accuracy from 79.49% to 80.33% and precision from 29% to 30%. Although the recall value did not increase, this method proved to be able to improve the overall stability of the prediction, especially in terms of the accuracy of the classification of the positive class. The novelty of this research lies in the specific application of the SMOTE-ENN method on large-scale health datasets with the C4.5 algorithm, which has not been widely explored before. Therefore, further exploration of other balancing techniques and algorithms is needed to obtain more optimal classification results on unbalanced data.**

Keywords*— *SMOTE-ENN, Data Imbalance, C4.5, Diabetes, Classification

I. INTRODUCTION

Diabetes is a global health problem with an increasing prevalence. Based on WHO data, more than 422 million people in the world have diabetes, and the disease is responsible for more than 1.5 million deaths each year. The impact is more pronounced in developing countries, where the lack of healthcare facilities is a major challenge in the diagnosis and management of diabetes. In addition, diabetes is also associated with serious complications such as heart disease, kidney damage, neurological disorders, and vascular complications that significantly reduce the quality of life of patients [1].

In health data processing, significant challenges arise from imbalanced data. This imbalance occurs when the amount of data of the minority class (e.g., diabetes cases) is much smaller than the majority class (e.g., non-diabetes). This makes predictive algorithms tend to be biased towards the majority class, thus reducing the system's ability to detect rare but clinically important critical conditions [2][1].

To overcome this problem, various oversampling techniques have been developed, one of which is SMOTE (Synthetic Minority Over-Sampling Technique). SMOTE generates synthetic data to increase the proportion of minority classes. Research by Rezki et al. (2024) showed that the application of SMOTE can improve the performance of the C5.0, Random Forest, and SVM algorithms in diabetes prediction using the Pima Indian Diabetes dataset. However, they also highlighted the risk of overfitting due to the addition of synthetic data without further cleaning [3][4].

As a more advanced solution, SMOTE-ENN, a combination of SMOTE and Edited Nearest Neighbor, is used to not only augment minority class data but also clean the data from noise. The study by Wang (2022) showed that SMOTE-ENN can improve the accuracy of postoperative complication prediction up to 90% with XGBoost algorithm, emphasizing the importance of the combination of oversampling and data cleaning on medical datasets [4].

Besides SMOTE-ENN, adaptive approaches such as ADASYN (Adaptive Synthetic Sampling) are also being used to balance the data. ADASYN dynamically generates synthetic data based on the classification difficulty of each minority sample. In the study of Marlisa et al. (2024), ADASYN improved accuracy, specificity, and sensitivity in diabetes classification using the K-Nearest Neighbor algorithm, showing that this approach is effective in handling imbalanced data [2].

On the other hand, the C4.5 algorithm is a popular decision tree method due to its ability to handle numerical and categorical data attributes, and classification results that can be interpreted easily. However, the effectiveness of C4.5 in unbalanced datasets remains limited without adequate data balancing techniques [5].

The novelty of this research lies in the application of the combination of SMOTE-ENN and the C4.5 algorithm specifically for the classification of diabetes health indicators,

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2350, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

which has not been widely explored in previous studies. This approach is expected to improve classification performance especially in minority classes and make a real contribution to the development of decision support systems in the field of Health [4][5].

Thus, this research not only extends the application of SMOTE-ENN to medical data, but also presents a more optimal alternative to traditional balancing methods in the effort to diagnose and manage diabetes more accurately and efficiently.

II. RESEARCH METHODS

First, this research begins with the collection of datasets downloaded from Kaggle in Excel or CSV format. Next, a pre-processing stage is carried out which includes cleaning, normalization, outlier handling, and data division. After that, to overcome data imbalance, an oversampling technique is applied using the SMOTE-ENN method, which combines minority data synthesis (SMOTE) and data cleaning using Edited Nearest Neighbor (ENN). The balanced data was then divided into training and test data for model training and evaluation purposes. The classification process is performed using the C4.5 algorithm to build a prediction model. Finally, the classification results were evaluated to measure the performance of the model before the study was concluded.

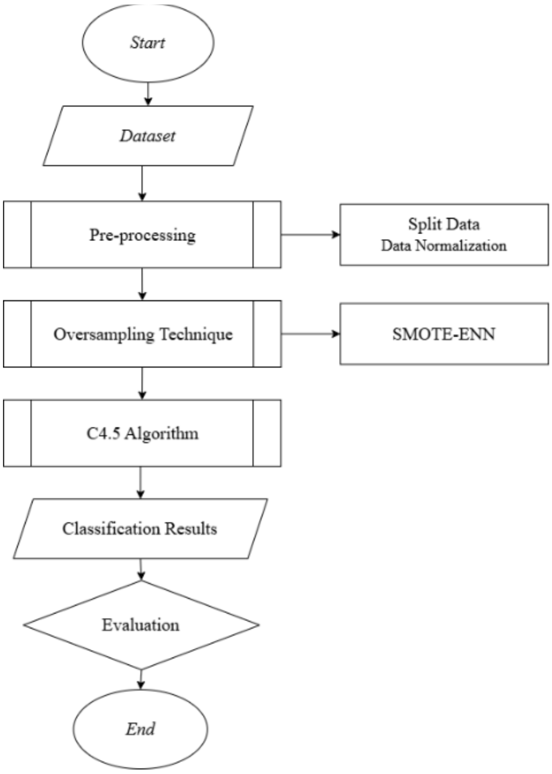


Fig. 1. Research Methods

A. Dataset

The dataset used in this study was obtained from the Kaggle platform due to its relevance to the research objective, which is to build a diabetes prediction model. The data was collected by accessing the Kaggle website (https://www.kaggle.com) and downloading the dataset in CSV format. The dataset used is the Diabetes Binary Health Indicators BRFSS 2021, which consists of 236,378 respondent data with 21 features covering health indicators such as behavior, chronic health conditions, and other risk factors. The target variable in this dataset is diabetes_binary, which indicates whether the respondent is indicated to have diabetes or not. The data distribution shows class imbalance, with 85.80% belonging to the non-diabetes class (0.0) and only 14.20% to the diabetes class (1.0). This imbalance is a challenge in the classification process, so data balancing methods such as SMOTE-ENN are applied to improve model performance.

TABLE I. DATASET DIABETES

Diabetes_binary	HighBP	HighChol	...	Education	Income
0.0	0	1.0	...	4.0	5.0
1.0	1	0.0	...	4.0	3.0
1.0	1	1.0	...	4.0	7.0
1.0	0	1.0	...	3.0	4.0
0.0	0	0.0	...	5.0	6.0
0.0	1	0.0	...	4.0	8.0
0.0	1	1.0	...	5.0	3.0
...	...	...	...	...	...
0.0	1	0.0	...	4.0	5.0
0.0	0	1.0	...	6.0	10.0
0.0	1	0.0	...	4.0	6.0
0.0	0	1.0	...	6.0	6.0

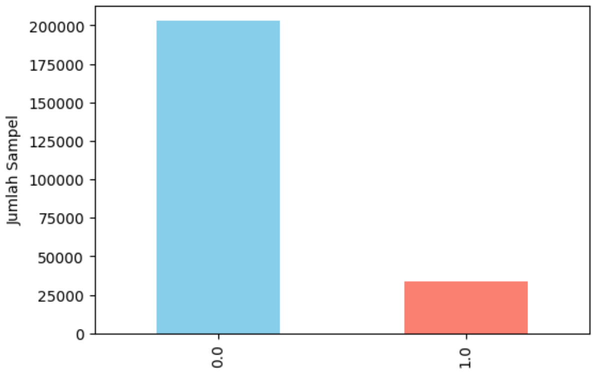


Fig. 2. Comparison of Majority and Minority Class Samples

B. Pre-Processing

The pre-processing stage is an important initial process in data processing before modeling. At this stage, the data that has been collected will be prepared through several steps, such as separating features and labels, dividing data into training and testing data, and normalizing data. This process aims to ensure that the data is in an optimal condition so that it can improve the performance of the model in the classification stage.SMOTE-ENN

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2350, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

1) Separation of Features and Labels

Data separation is done by separating feature attributes (x) and target labels (y) from the dataset [6]. The target label is the diabetes_binary variable, while the other features become attributes used to predict the label [7].

2) Split Data

The dataset is divided into training and testing data using the train_test_split function with a portion of 60% for training and 40% for testing. This separation aims to train the model on training data and evaluate the performance of the model on testing data [8][9].

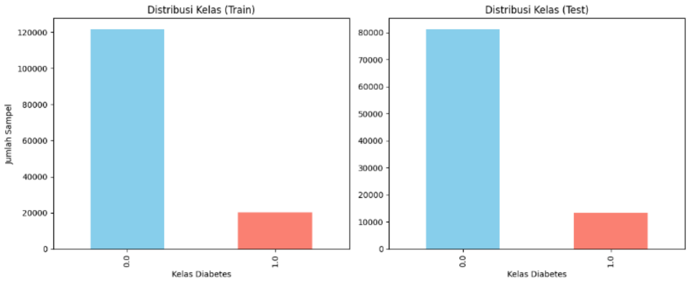


Fig. 3. Split Data

3) Data Normalization

Data features are converted into a certain scale using normalization techniques, such as Min-Max Scaler, so that all attributes are in the same range of values, usually between 0 and 1. This process helps the algorithm work more optimally, especially when the data has a large scale difference between features. Normalization is performed on training data and then applied to testing data to maintain scale consistency [10][11].

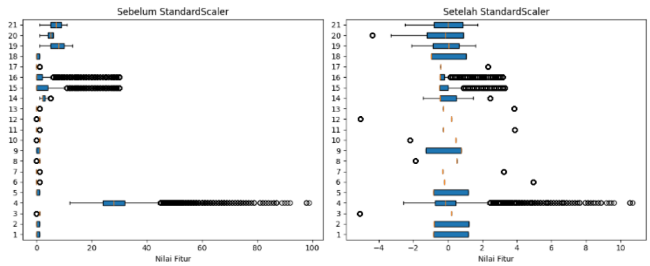


Fig. 4. Data Normalization

C. *SMOTE-ENN*

The problem of data imbalance arises when the proportion of the number of majority classes (classes with more samples) and minority classes (classes with fewer samples) is unbalanced. In this Diabetes Health Indicators dataset, imbalance occurs in the target variable diabetes_binary, where the amount of data without diabetes is much greater than the amount of data with diabetes. This imbalance can cause the prediction model to be biased towards the majority class and reduce classification performance [4][12].

To overcome this problem, a combination of Synthetic Minority Over-Sampling Technique (SMOTE) and Edited Nearest Neighbors (ENN) method is used. The SMOTE technique works by adding synthetic samples to the minority class, while ENN cleans the data by removing samples that are misclassified or considered noise, thus helping to reduce the risk of overfitting and improve data quality [13][4].

The implementation of SMOTE is done by first identifying the minority class in the dataset, which is the class with a value of 1 in the target variable diabetes_binary. Then, SMOTE is used to generate additional samples based on the nearest neighbor data of the minority class in the training data, so that the distribution between classes becomes more balanced. This step aims to reduce model bias towards the majority class and improve the model's ability to recognize important patterns in both classes. By using the SMOTE-ENN approach, the model is expected to produce more accurate and reliable predictions. The process flow of the SMOTE-ENN method can be seen in the following figure [13].

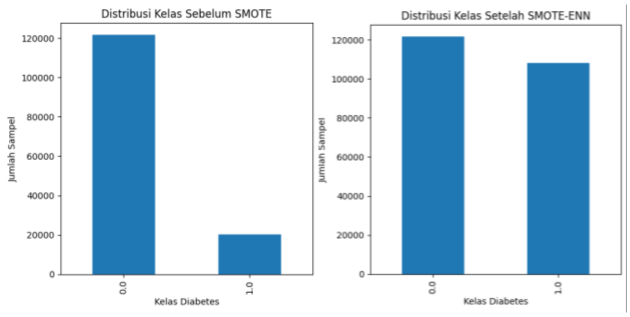


Fig. 5. Before and After SMOTE-ENN Implementations

Figure Explanation:

- Left Panel (Class Distribution Before SMOTE-ENN) Before SMOTE is applied, there is a significant imbalance between class 0 (without diabetes) and class 1 (with diabetes). the number of class 0 samples is much larger than class 1, which may cause bias in the model [14].
- Right panel (class distribution after SMOTE-ENN) After applying enn, the number of minority class samples is slightly reduced. this indicates that enn removes samples that are considered noise or less relevant. the end result is a cleaner and more balanced dataset, which is ready to be used for classification model training [15].

D. *C4.5 Algorithm*

The C4.5 algorithm is a popular method in data mining used to build decision trees [16]. This algorithm has several advantages, such as being able to handle attributes with continuous and discrete values, overcome attributes with empty

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2350, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

values (Missing Values), and support the process of pruning decision tree branches to simplify the model [5][17].

The main process of this algorithm involves several steps. First, it calculates the entropy value of the dataset, which is used to measure the level of data uncertainty. the formula used is:

$$Entropy(S) = -\sum_{i=1}^{k} pi\, log_2\, pi \quad ...(2.1)$$

Where pi is the probability of occurrence of class i.

After the entropy value is calculated, the algorithm determines the information gain, which is the reduction of uncertainty after the data is divided based on certain attributes:

$$Gain(S,A) = Entropy(S) - \sum_{v \in values(A)} \frac{|Sv|}{|S|} Entropy(Sv) \quad ...(2.2)$$

The attribute with the highest gain value is chosen as the *root* or node of the decision tree[18][19].

However, to avoid bias towards attributes with many categories, C4.5 uses *Gain Ratio*:

$$GainRatio(S,A) = \frac{Gain(S,A)}{SplitInfo(S,A)} \quad ...(2.3)$$

With

$$SplitInfo(S,A) = \sum_{v \in values(A)} \frac{|Sv|}{|S|} Log^2 \frac{|Sv|}{|S|} \quad ...(2.4)$$

The attribute with the highest *gain ratio* value will become a node in the decision tree. .

This process continues to repeat until each branch of the tree only contains data with the same class, or no more attributes can be used to further divide the data [20][21]. The algorithm also performs branch pruning to avoid overfitting by removing branches that do not contribute significantly to the accuracy of the model [17][22].

E. Evaluation Metrics

Confusion Matrix is a matrix used in machine learning to evaluate the performance of classification models. This matrix presents the comparison between model predictions and actual values in the form of four elements: *True Positive (Tp), False Positive (Fp), False Negative (Fn), and True Negative (Tn).* Using these elements, we can calculate various evaluation metrics such as *Accuracy, Precision, Recall, Dan F1-Score*[23].

Based on the Confusion Matrix results, calculations for several metrics can be done as follows:

1. *Accuracy*
   Accuracy measures how many predictions are correct (both positive and negative) compared to the total amount of data [24].

The equation for calculating accuracy is:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad ...(3.1)$$

Where:

- *Tp = True Positive* (Correct Prediction for Positive Class)
- *Tn = True Negative* (Correct Prediction For Negative Class)
- *Fp = False Positive* (False Prediction For Negative Class)
- *Fn = False Negative* (False Prediction For Positive Class)

2. *Precision*
   Precision measures the accuracy of the positive predictions made by the model, which is how many positive predictions are correct compared to the total positive predictions made [23].

The equation for calculating precision is:

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad ...(3.2)$$

Where:

- *Tp = True Positive*
- *Fp = False Positive*

3. *Recall*
   Recall measures the model's ability to detect true positive classes. It is the ratio between the number of correct positive predictions and the total number of truly positive data [25].

The equation to calculate recall is:

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad ...(3.3)$$

Where:

- *TP = True Positive*
- *FN = False Negative*

4. *F1-score*

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad ...(3.4)$$

III. Results And Discussion

This study compares the performance of the C4.5 algorithm before and after the application of the SMOTE-ENN method to overcome class imbalance on the Diabetes Binary Health Indicators BRFSS 2021 dataset. Before balancing, the model produces 79.49% accuracy, 29% precision, and 31% recall, which shows that the model is less able to detect diabetics as a minority class.

After the application of SMOTE-ENN, the accuracy of the model increased from 79.49% to 80.33% and the precision increased from 29% to 30%. Although the recall value decreased slightly from 31% to 30%, it shows that the balancing method successfully improved the overall prediction stability, especially in terms of the accuracy of predicting positive classes, although the sensitivity to minority classes can still be improved. Table 2 presents a complete comparison of model performance.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2350, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

TABLE II. MODEL OF PERFORMANCE EVALUATION WITH SMOTE-ENN

Metrics	Without SMOTE-ENN	With SMOTE-ENN
Accuracy	79,49%	80,33%
Precision	29%	30%
Recall	31%	30%
F1-Score	30%	30%

Following the results before balancing the data using the SMOTE-ENN method, the model was re-trained and evaluated. The classification results are shown in the following confusion matrix:

TABLE III. CONFUSION MATRIX C4.5 WITHOUT SMOTE-ENN

	Positive Prediction	Negatif Prediction
Positive Actual	4.174	9.193
Negatif Actual	10.198	70.987

TABLE IV. PERFORMANCE EVALUATION OF MODEL WITHOUT SMOTE-ENN

Metrics	Value
Accuracy	79,49%
Precision	29%
Recall	31%
F1-Score	30%

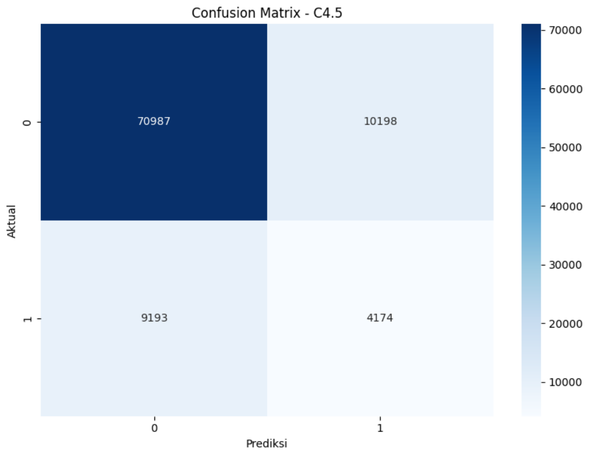


Fig. 6. Confusion Matrix C4.5 without SMOTE-ENN

Following the results after balancing the data using the SMOTE-ENN method, the model was re-trained and evaluated. The classification results are shown in the following confusion matrix:

TABLE V. CONFUSION MATRIX C4.5 WITH SMOTE-ENN

	Positive Prediction	Negatif Prediction
Positif Actual	**4.070**	**9.297**
Negative Actual	**9.298**	**71.887**

TABLE VI. MODEL PERFORMANCE EVALUATION WITH SMOTE-ENN

Metrics	Value
Accuracy	80,33%
Precision	30%
Recall	30%
F1-Score	30%

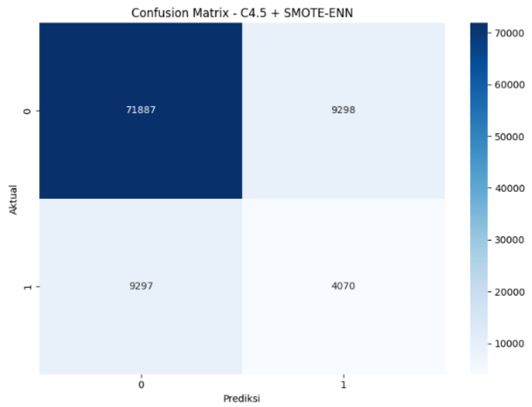


Fig. 7. Confusion Matrix C4.5 with SMOTE-ENN

In general, the improvement in model performance is not statistically significant. However, in practical terms, this balancing process still contributes to the stability of the overall prediction, although it is not optimal in detecting minority classes.

When compared to the research of Marlisa et al. (2024) [2], who used ADASYN and K-NN, they obtained a sensitivity (recall) of 71.79% - much higher than the results in this study. Meanwhile, Rezki et al. (2024) [3] showed that using SMOTE on C5.0, Random Forest, and SVM algorithms resulted in an AUC of up to 0.831. However, they also noted the possibility of overfitting due to synthetic data. Similar results occurred in this study, where precision increased slightly, but recall showed no improvement.

Wang (2022) research [4] used SMOTE-ENN and XGBoost, and managed to obtain 90% accuracy with an AUC of 0.90. This corroborates the notion that the success of the balancing method is highly dependent on the algorithm selection.

The novelty of this research lies in the application of the

SMOTE-ENN method with the C4.5 algorithm on large-scale health datasets, which has not been widely explored. The scientific contribution provided is to extend the evidence that balancing techniques such as SMOTE-ENN need to be combined with more adaptive algorithms to achieve optimal performance in minority classes.

The limitations of this research are that no statistical significance test has been conducted on metric changes, no parameter tuning has been done, and only one classification algorithm (C4.5) has been used without comparison. This research is implemented using Python with the SMOTE-ENN approach and evaluation through Confusion Matrix, but further development is still needed to improve classification performance, especially in detecting minority classes.

.

IV. Conclusions

Based on the results of the research that has been done, the C4.5 algorithm without handling data imbalance produces an accuracy of 79.49%, but has a low recall for minority classes, which is 31%. After applying the SMOTE-ENN method, the accuracy increased to 80.33% and the precision for the minority class also increased to 30%. However, this method did not provide a significant improvement to the recall of minority classes. This shows that although SMOTE-ENN is able to balance the data distribution, its effectiveness in improving classification performance is highly dependent on the characteristics of the data and the type of algorithm used.

References

[1] WHO, "Thermostability of human insulin," *World Heal. Organ. 2024.*, vol. 2050, no. 1, pp. 1–7, 2024.

[2] H. Marlisa, N. Satyahadewi, N. Imro'ah, and N. N. Debataraja, "Application of Adasyn Oversampling Technique on K-Nearest Neighbor Algorithm," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 18, no. 3, pp. 1829–1838, 2024.

[3] M. K. Rezki, M. I. Mazdadi, F. Indriani, Muliadi, T. H. Saragih, and V. A. Athavale, "Application of Smote to Address Class Imbalance in Diabetes Disease Categorization Utilizing C5.0, Random Forest, and Support Vector Machine," *J. Electron. Electromed. Eng. Med. Informatics*, vol. 6, no. 4, pp. 343–354, 2024.

[4] J. Wang, "Prediction of postoperative recovery in patients with acoustic neuroma using machine learning and SMOTE-ENN techniques," *Math. Biosci. Eng*. aimspress.com, 2022.

[5] R. P. Fadhillah, R. Rahma, and ..., "Klasifikasi Penyakit Diabetes Mellitus Berdasarkan Faktor-Faktor Penyebab Diabetes menggunakan Algoritma C4. 5," *... Penelit. dan ...*, 2022.

[6] R. Doğan, S. M. Çınar, and E. Akarslan, "A Novel ZIP-Based NILM Method Design Robust to Undervoltage and Overvoltage Conditions," *Arab. J. Sci. Eng.*, 2025.

[7] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1060–1073, 2022.

[8] Q. H. Nguyen *et al.*, "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Math. Probl. Eng.*, vol. 2021, 2021.

[9] M. T. Akhir, M. Syarat, G. Memperoleh, G. Sarjana, S. Satu, and T. Informasi, *Perbandingan Kinerja Metode Klasifikasi Naïve Bayes Dan Random Forest Dalam Analisis Sentimen Kasus Narkoba di Indonesia Pada Komentar YouTube SKRIPSI Diajukan oleh : NAILUL ' INAYAH PROGRAM STUDI TEKNOLOGI INFORMASI*. 2023.

[10] A. Ambarwari, Q. J. Adrian, and Y. Herdiyeni, "Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 117–122, 2020.

[11] S. Nagibzadeh, *Bilgisayar Bilimleri ve Mühendisli ğ i + tr-2*, no. January 2025. 2024.

[12] M. Altalhan, A. Algarni, and M. Turki-Hadj Alouane, "Imbalanced Data Problem in Machine Learning: A Review," *IEEE Access*, vol. 13, no. January, pp. 13686–13699, 2025.

[13] M. Seyedtabib and N. Kamyari, "Predicting polypharmacy in half a million adults in the Iranian population: comparison of machine learning algorithms," *BMC medical informatics and decision making*. Springer, 2023.

[14] H. L. Ngo *et al.*, "The composition of time-series images and using the technique SMOTE ENN for balancing datasets in land use/cover mapping," *Acta Montan. Slovaca*, vol. 27, no. 2, pp. 342–359, 2022.

[15] M. Lu, L. T. Tay, and J. Mohamad-Saleh, "Landslide susceptibility analysis using random forest model with SMOTE-ENN resampling algorithm," *Geomatics, Nat. Hazards Risk*, vol. 15, no. 1, p. , 2024.

[16] Dhea Halimah, Muhammad Ridwan Lubis, and Widodo Saputra, "Algoritma C4.5 Untuk Menentukan Klasifikasi Tingkat Pemahaman Mahasiswa Pada Matakuliah Bahasa Pemrograman," *J. Tek. Mesin, Ind. Elektro Dan Inform.*, vol. 1, no. 3, pp. 24–38, 2022.

[17] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.

[18] U. P. Budi and A. Info, "Application Of C4 . 5 Algorithm In Disease Classification," vol. 2, no. 02, pp. 58–62, 2024.

[19] A. Afifuddin and L. Hakim, "Deteksi Penyakit Diabetes Mellitus Menggunakan Algoritma Decision Tree Model Arsitektur C4.5," *J. Krisnadana*, vol. 3, no. 1, pp. 25–33, 2023.

[20] L. Y. L. Gaol, M. Safii, and D. Suhendro, "Prediksi Kelulusan Mahasiswa Stikom Tunas Bangsa Prodi Sistem Informasi Dengan Menggunakan Algoritma C4. 5," *Brahmana J. Penerapan ...*, 2021.

[21] F. F. Nugraha, I. Sunandar, and C. Julian, "Penerapan Data Mining Dengan Metode Kalsifikasi Menggunakan Algoritma C4.5," *Teknologi*, vol. 7, no. March, pp. 10–20, 2022.

[22] M. A. Barata *et al.*, "PERANCANGAN SISTEM ELECTRONIC NOSE BERBASIS," pp. 117–126, 2016.

[23] M. D. Nguyen *et al.*, "Estimation of recompression coefficient of soil using a hybrid ANFIS-PSO machine learning model," *J. Eng. Res.*, vol. 12, no. September 2023, pp. 358–368, 2024.

[24] V. R. Prasetyo, M. Mercifia, A. Averina, L. Sunyoto, and B. Budiarjo, "Prediksi Rating Film Pada Website Imdb Menggunakan Metode Neural Network," *Netw. Eng. Res. Oper.*, vol. 7, no. 1, p. 1, 2022.

[25] S. Sathyanarayanan and B. R. Tantri, "Confusion Matrix-Based Performance Evaluation Metrics," no. November, 2024.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2350, Copyright ©2025
Submitted : April 30, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Comparison of CNN Architectures for Pre-Cancerous Cervical Lesion Classification Based on Colposopy Images Using IARC and AnnoCerv Datasets

Sigit Prasetyo Noprianto[1]*, Siti Nurmaini[2], Dian Palupi Rini[3]
Department of Masters in Computer Science[1], [2], [3]
University of Sriwijaya
Palembang, Indonesia
noprisigit@gmail.com[1], sitinurmaini@gmail.com[2], dprini@unsri.ac.id [3]

***Abstract*— Cervical cancer represents a significant public health issue affecting women worldwide, and identifying the severity of lesions early on is crucial to selecting the right treatment. This research investigates and compares the effectiveness of various Convolutional Neural Network (CNN) models in classifying colposcopic images according to the severity of cervical lesions. The dataset used was obtained from the International Agency for Research on Cancer (IARC) and AnnoCerv, consisting of 452 colposcopy images categorized into four classes: Normal, CIN 1, CIN 2, and CIN 3. Five CNN architectures were evaluated: MobileNetV2, InceptionV3, Xception, VGG16, and DenseNet121. Experiments were conducted using default hyperparameters: batch size of 32, learning rate of 0.001, and 100 epochs. The results showed that MobileNetV2 achieved the highest accuracy at 67%, followed by DenseNet121 (60%), Xception (60%), InceptionV3 (55%), and VGG16 (42%). Based on these findings, MobileNetV2 is the most optimal model for classifying colposcopy images in this study. However, the study is limited by class imbalance and dataset size, which may affect model generalizability. Future work may explore ensemble learning techniques and larger, more diverse datasets for improved accuracy.**

Keywords*— *Cervical Cancer, Colposcopy Image Classification, CNN, CIN Classification

I. Introduction

Artificial Intelligence (AI) has rapidly advanced in recent decades with widespread applications across various domains, including healthcare. One of the most prominent branches of AI is deep learning, which has demonstrated outstanding performance in medical image analysis [1]. Deep learning models, particularly Convolutional Neural Networks (CNN), have been successfully applied to various disease classification tasks, such as brain tumors, breast cancer, and skin cancer, achieving high levels of accuracy [2][3][4].

Cervical cancer ranks as the fourth most common cancer affecting women and is one of the most feared diseases among them. WHO statistics reveal that globally in 2020, cervical cancer led to around 604,000 new cases and approximately 341,831 fatalities [5]. About 95% of these cases are caused by infection with the Human Papillomavirus (HPV) [6] [7]. WHO recommends several screening methods, including cytology tests (Pap smear), HPV tests, visual inspection with acetic acid (VIA), and colposcopy [8]. Among the available methods, VIA is regarded as one of the most cost-effective and can be applied in resource-limited healthcare settings. The procedure involves applying acetic acid to the cervix, followed by direct visual inspection to detect acetowhite lesions. The World Health Organization (WHO) classifies abnormal cervical lesions found through VIA or alternative screening tools as Cervical Intraepithelial Neoplasia (CIN), which is divided into three grades: CIN 1, CIN 2, and CIN 3.

In clinical practice, colposcopy is often performed after a VIA test to further examine suspicious areas of the cervix. Colposcopy utilizes an optical magnifying instrument (colposcope) to produce clearer and more detailed cervical images. These images reveal specific visual patterns such as mosaicism, punctation, and lesion borders that help determine CIN severity levels. These visual features serve as critical indicators for developing deep learning-based classification models to detect and differentiate CIN levels more accurately [9].

Previous studies on pre-cancerous cervical lesion classification have mostly focused on the Pap smear method [10][11][12], and predominantly employed binary classification approaches—distinguishing only between normal and abnormal images. Although this approach aids in initial screening, binary classification does not provide specific information on lesion severity, which is essential for determining proper medical follow-up. Some studies have attempted to classify cervical lesions from colposcopy images into categories such as normal, CIN1, CIN2, CIN3, and cancer. For instance, a study using the ResNet-152 model achieved an average accuracy of 51.7% for multi-class CIN classification, with an Area Under the Curve (AUC) of 0.781 to distinguish high-risk from low-risk lesions [13]. Another study developed an ensemble deep learning model named CYENET for classifying cervical cancer from colposcopy images, which improved classification accuracy to 92.3%, compared to 73.3% achieved by the VGG19 model [14]. In a different study, a CAD system was developed by integrating deep learning descriptors

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025

such as ResNet50, ResNet101, and ResNet152 with dimensionality reduction techniques for colposcopy image classification. This approach achieved outstanding performance, ranging from 97%–100% in normal-abnormal classification and lesion type identification [15].

While several studies have attempted colposcopy-based classification, the often use limited datasets, focus on binary classification, or apply only one CNN model, making it difficult to assess comparative model performance. In addition, limited publicly available colposcopy datasets, such as IARC and AnnoCerv, pose challenges related to data imbalance dan image variability.

Despite these promising results, challenges such as class imbalance, varying image quality, and differences in colposcopy equipment across institutions remain. Therefore, this study aims to evaluate and compare several popular CNN architectures—MobileNetV2, InceptionV3, Xception, VGG16, and DenseNet121—for CIN lesion classification based on colposcopy images. MobileNetV2 is known for its high efficiency [16], InceptionV3 is effective in capturing complex visual features from medical images [17], Xception improves upon Inception using separable convolutions for higher accuracy in image classification tasks [18] , VGG16, despite its large number of parameters, remains a strong baseline due to its simple yet effective convolutional structure [19], and DenseNet121 allows for efficient feature and gradient propagation through dense layer connections, showing strong performance in image-based disease classification [20]. This evaluation is expected to provide insights into model performance and its potential applications in clinical settings, particularly for early detection of cervical cancer.

This study contributes to the field by evaluating the classification effectiveness of five different CNN models for multi-class classification of cervical lesion severity using colposcopy images. To the best our knowledge, no existing studies have performed a direct evaluation and comparison of these five CNN architectures using both IARC and AnnoCerv datasets for four-class cervical lesion classification from colposcopy images. This positions our work as a novel contribution that can guide future model selection and deployment in clinical decision support systems.

II. RESEARCH METHODOLOGY

This research began with problem identification, followed by a literature review to build upon previous related studies. The next step involved collecting data aligned with analytical requirements. The research methodology includes data preprocessing, architectural design, model training and evaluation, and finally, documentation and reporting. The research stages are described as follows.

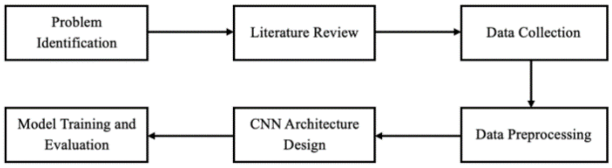


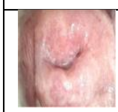
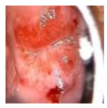
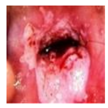
Fig. 1. Research flow

A. Problem Identification

The initial stage of this study involved identifying the specific research problem. The focus of this study is the classification of pre-cancerous cervical lesions resulting from visual inspection with acetic acid (VIA) into four categories: Normal, CIN 1, CIN 2, and CIN 3. This research aims to solve the problem by implementing a system based on different CNN models, which are then compared to find the one that performs best in detecting pre-cancerous cervical lesions.

As illustrated in Table 1, the dataset used in this study comprises colposcopic images categorized into four classes based on lesion severity: "Normal," "CIN 1" (mild dysplasia), "CIN 2" (moderate dysplasia), and "CIN 3" (severe dysplasia). The "Class" row defines the lesion grade, while the "Image" row displays representative samples of each category, highlighting distinct visual characteristics such as changes in epithelial texture, color, and vascular patterns. These annotated samples serve as input for training and evaluating the proposed deep learning models aimed at automated classification of cervical lesion severity.

TABLE I. CERVICAL PRE-CANCEROUS LESIONS COLPOSCOPY IMAGE DATASET

Normal	CIN 1	CIN 2	CIN 3
			
			

B. Data Collection Method

This stage includes the process of gathering cervical lesion image data to build the dataset used in this classification research. The dataset was obtained from two primary sources: the International Agency for Research on Cancer (IARC) and AnnoCerv [21]. The IARC dataset comprises 913 colposcopy images from 200 case examinations, while the AnnoCerv dataset contains 527 images from 100 cases. The image formats are .jpg for IARC and a combination of .jpg and .png for AnnoCerv; however, only the .jpg format was used.

Colposcopic images in the IARC dataset are collected for each case after applying a sequence of diagnostic fluids—namely, normal saline, vinegar acid solution with and without a green filter, and an iodine-based solution. Meanwhile, the AnnoCerv dataset provides images after acetic acid and Lugol's iodine applications. These datasets were used for both training and testing in developing a machine learning-based image classification model to detect and identify pre-cancerous

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025

cervical lesion severity.

C. Pre-Processing

The preprocessing stage was conducted to prepare the data for training and evaluating the cervical lesion image classification models. The datasets from IARC and AnnoCerv were first grouped into four classes based on lesion severity: Normal, CIN 1, CIN 2, and CIN 3. Label assignment referred to the accompanying medical diagnostic information for each image. Only acetic acid-applied images were used in this study, as they are medically considered the most relevant for identifying pathological changes in cervical tissue. The grouping resulted in the following distribution: 89 images for Normal, 148 for CIN 1, 105 for CIN 2, and 110 for CIN 3, as detailed in Table 2.

TABLE II. DISTRIBUTION OF IARC AND ANNOCERV DATASET PER CLASS

Dataset	Normal	CIN 1	CIN 2	CIN 3
IARC	35	22	36	59
Annocerv	54	126	69	51
Total	89	148	105	110

Each image was resized to 224×224 pixels to ensure compatibility with the expected input format of the CNN frameworks. The original resolution of IARC images was 800×600 pixels, while AnnoCerv images were 2976×1984 pixels.

To address class imbalance, image augmentation was applied. Augmentation strategies involved applying rotations, horizontal and vertical displacements, horizontal mirroring, and brightness adjustments to the images. The applied augmentations contributed to balancing class distributions, enhancing dataset variability, and mitigating overfitting. After augmentation, each class contained 148 images, resulting in a final dataset of 592 images. Following data preprocessing, a split was performed where 70% of the dataset was utilized for training and 30% for validation.

D. Model Architecture

This study conducted training and evaluation for several CNN model architectures, including MobileNetV2, InceptionV3, Xception, VGG16, and DenseNet121. Using a 70:30 training validation split, 412 images were used for training and 180 for validation.

E. Evaluation of Architectural Design Result

A confusion matrix was employed in the evaluation process of the trained models, offering performance indicators including accuracy, precision, recall, and the F1-score to judge model effectiveness. The formulas used in this evaluation process are as follows:

$$Accuracy = \frac{\sum_{i=1}^{N} TP_i}{Total\ Data} \quad (1)$$

$$Precision = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i} \quad (2)$$

$$Recall = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(TP_i + FN_i)} \quad (3)$$

$$F1\ Score = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall} \quad (4)$$

III. RESULTS AND DISCUSSION

This research evaluates the classification accuracy of several CNN models in identifying pre-cancerous cervical lesions, employing MobileNetV2, InceptionV3, Xception, VGG16, and DenseNet121 in the experimentation process. All models were trained under the same settings, using a batch size of 32, a learning rate of 0.001, and 100 epochs. A total of 412 data points were used for training, while 180 were allocated for validation. The following are the results obtained from each model:

A. MobileNetV2

The classification outcomes of the MobileNetV2 model, represented as a confusion matrix, are depicted in Figure 2, which provides detailed insight into the classification performance across the four classes: Normal, CIN1, CIN2, and CIN3.

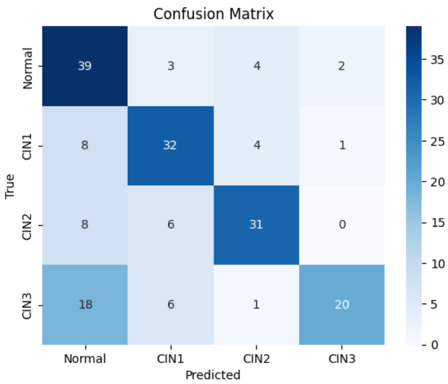


Fig. 2. Confusion Matrix of MobileNetV2

TABLE III. MOBILENETV2 MODEL EVALUATION

	Precision	Recall	F1-Score
Normal	53%	81%	64%
CIN 1	68%	71%	70%
CIN 2	78%	69%	73%
CIN 3	87%	44%	59%

The model demonstrates strong performance in detecting Normal cases, correctly classifying 39 out of 48 samples, which corresponds with the high recall value of 81% reported in Table 3. However, the relatively low precision of 53% for this class indicates a high number of false positives, as seen in the misclassification of several CIN 1, CIN 2, and CIN 3 cases as

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Normal. A total of 32 samples from the CIN 1 class were correctly identified by the model, out of 45 instances, reflecting a balanced performance with a recall of 71% and precision of 68%. Similarly, the CIN 2 class shows reliable classification results with 31 correct predictions out of 45, consistent with its 69% recall and 78% precision. In contrast, the CIN 3 class reveals a notable weakness of the model: although it achieves a high precision of 87%, indicating that most predictions labeled as CIN 3 are accurate, it only correctly classifies 20 out of 45 actual CIN 3 cases. This results in a low recall of 44%, suggesting that many true CIN 3 cases are not being detected. The overall pattern observed in the confusion matrix confirms the results summarized in Table 3 and highlights the model's strength in precision for higher-grade lesions (e.g., CIN 3) while revealing its limited sensitivity in detecting all true cases within that class.

B. InceptionV3

The classification outcomes of the InceptionV3 model, represented as a confusion matrix, are depicted in Figure 3, offering a detailed breakdown of the classification outcomes across the four diagnostic categories: Normal, CIN 1, CIN 2, and CIN 3. The model correctly identified 27 Normal cases, out of what appears to be 48 actual Normal samples, resulting in a recall of approximately 56%, which matches the value reported in Table 4.

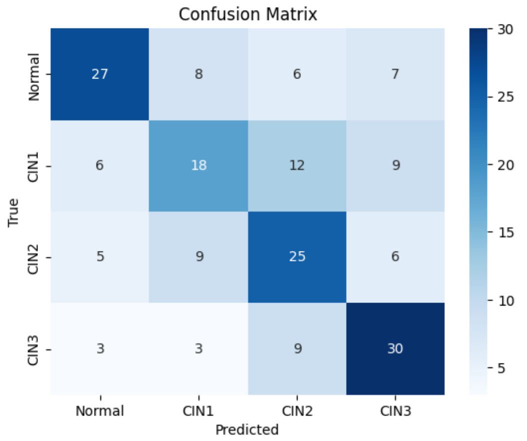


Fig. 3. Confusion Matrix of Inception V3 Model

TABLE IV. INCEPTIONV3 MODEL EVALUATION

	Precision	Recall	F1-Score
Normal	66%	56%	61%
CIN 1	47%	40%	43%
CIN 2	48%	56%	52%
CIN 3	58%	67%	62%

The precision for the Normal class was 66%, suggesting that when the model predicted a case as Normal, it was correct most of the time. For the CIN 1 class, only 18 cases were correctly classified out of an estimated 45, yielding a recall of about 40%. This class also had a low precision of 47%, which is consistent with its significant overlap with other categories, particularly CIN 2, as reflected in the confusion matrix. The CIN 2 class showed a relatively balanced performance, with 25 out of 45 samples correctly classified, leading to a recall of 56% and a precision of 48%, aligning with the reported metrics. In contrast, the CIN 3 class exhibited the highest recall at 67%, as the model correctly classified 30 of the 45 CIN 3 samples. However, the precision for CIN3 was lower (58%), indicating that a number of other class instances (e.g., CIN 2) were misclassified as CIN 3. This pattern suggests that the model was more sensitive in detecting CIN 3 but lacked specificity. Overall, the confusion matrix confirms the findings in Table 4 and supports the observation that InceptionV3, while able to capture a broader range of CIN 3 cases, struggled with class separability—particularly between CIN 1 and CIN 2—likely due to the subtle visual differences among these lesion grades.

C. Xception

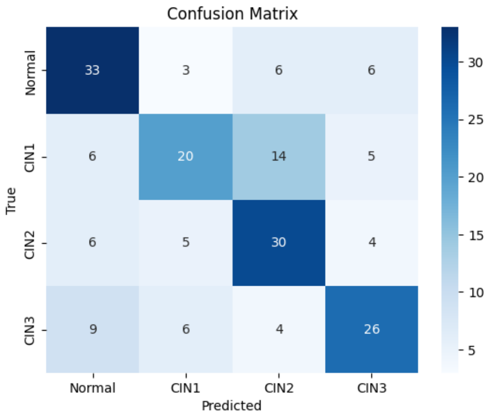


Fig. 4. Confusion Matrix of Xception

TABLE V. XCEPTIOON MODEL EVALUATION

	Precision	Recall	F1-Score
Normal	61%	69%	65%
CIN 1	59%	44%	51%
CIN 2	56%	67%	61%
CIN 3	63%	58%	60%

This class also achieved a precision of 61%, indicating a fair trade-off between correctly identifying Normal cases and minimizing false positives. For the CIN 1 class, only 20 out of an estimated 45 instances were correctly classified, leading to the lowest recall among the classes (44%) and a precision of 59%. This supports the observation that CIN 1 remains the most difficult class to distinguish, likely due to its visual similarity to both Normal and CIN 2 cases, as also seen in the confusion with 14 CIN 1 samples misclassified as CIN 2. The CIN 2 class showed good classification ability, with 30 correct predictions and a recall of 67%, while achieving a precision of 56%. For the CIN 3 class, the model correctly predicted 26 samples with a recall of 58% and a precision of 63%, reflecting a moderate but consistent detection performance. Overall, the confusion matrix supports the evaluation metrics in Table 5, confirming that Xception maintains a more even balance between sensitivity and specificity compared to previous models. While its performance does not dominate in any single class, it offers

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025

stable results across all categories, making it a promising candidate for general-purpose classification in this task.

D. VGG16

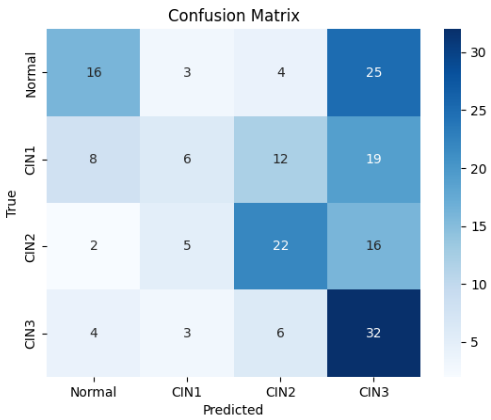


Fig. 5. Confusion Matrix of VGG16

TABLE VI. VGG16 MODEL EVALUATION

	Precision	Recall	F1-Score
Normal	53%	33%	41%
CIN 1	35%	13%	19%
CIN 2	50%	49%	49%
CIN 3	35%	71%	47%

The classification outcomes of the VGG16 model, represented as a confusion matrix, are depicted in Figure 5, which reveals the weakest overall performance among the tested models, aligning with its lowest reported accuracy of 42%. The Normal class shows a poor classification outcome, with only 16 correctly predicted samples out of an estimated 48, resulting in a recall of approximately 33% and a precision of 53%, as stated in Table 6. This indicates a high number of false negatives, with many Normal cases misclassified as CIN 3. The CIN 1 class performed the worst overall, with only 6 correct predictions and a recall of just 13%, reflecting the model's significant struggle to identify CIN 1 instances. Its precision of 35% further emphasizes the high degree of confusion with other classes, particularly CIN 2 and CIN 3, as evident in the matrix. CIN 2 showed slightly better results with 22 correct predictions, yielding a recall of 49% and a matching precision of 50%, demonstrating somewhat balanced but modest performance. Notably, CIN 3 achieved the highest recall at 71%, with 32 true positives. However, its precision was just 35%, suggesting that while the model is sensitive to CIN 3 cases, it frequently misclassifies samples from other classes—especially Normal and CIN 1—as CIN 3. This imbalance indicates that the model heavily overpredicts CIN 3, potentially due to its limited capacity to distinguish between lower-grade lesions. Overall, the confusion matrix confirms that VGG16 suffers from overgeneralization and class confusion, especially in distinguishing Normal and CIN 1 cases, possibly due to the architecture's rigidity and lack of deeper adaptive feature extraction, as suggested in Table 6.

E. DenseNet121

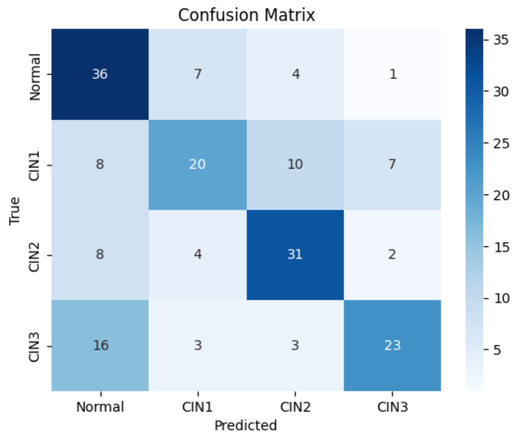


Fig. 6. Confusion Matrix of DenseNet21

TABLE VII. DENSENET21 MODEL EVALUATION

	Precision	Recall	F1-Score
Normal	53%	75%	62%
CIN 1	59%	44%	51%
CIN 2	65%	69%	67%
CIN 3	70%	51%	59%

Figure 6 shows the confusion matrix for the DenseNet121 model, which provides insight into the classification behavior of its densely connected architecture. The model achieved a strong performance in identifying Normal cases, correctly predicting 36 out of approximately 48 samples, corresponding to a recall of 75% and a precision of 53%, as reported in Table 7. This indicates a high sensitivity toward Normal class detection, although some misclassifications still occurred, particularly into CIN 1. The CIN 1 class had a lower recall of 44% with 20 correct predictions, and a precision of 59%, suggesting continued challenges in differentiating this class from neighboring grades like CIN 2 and CIN 3. CIN 2 was the best-performing class, with 31 out of around 45 instances correctly identified, yielding a recall of 69% and a precision of 65%. This strong performance supports the claim that DenseNet121 is well-suited for mid-stage lesion classification, likely due to its ability to retain and combine multi-level features effectively. For CIN 3, the model correctly predicted 23 samples, achieving a recall of 51% and the highest precision among the classes at 70%. While this indicates that most CIN 3 predictions were accurate, the relatively low recall means that a notable portion of CIN 3 cases were missed, often misclassified as Normal or CIN 2. Overall, the confusion matrix validates the evaluation results in Table 7, showing that DenseNet121 offers well-rounded performance, particularly excelling in CIN 2 classification. However, further optimization—such as enhanced regularization—may be required to boost its

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025

generalization in more challenging class boundaries.

F. Model Comparison Summary

TABLE VIII. EVALUATION METRIC COMPARISON

Model	Accuracy	Average Precision	Average Recall	Average F1-Score
MobileNetV2	67%	71,5%	66,25%	66,5%
InceptionV3	55%	54,75%	54,75%	54,%
Xception	60%	59,75%	59,5	59,25
VGG16	42%	43,25%	41,5%	39%
DenseNet121	60%	61,75%	59,75	59,75%

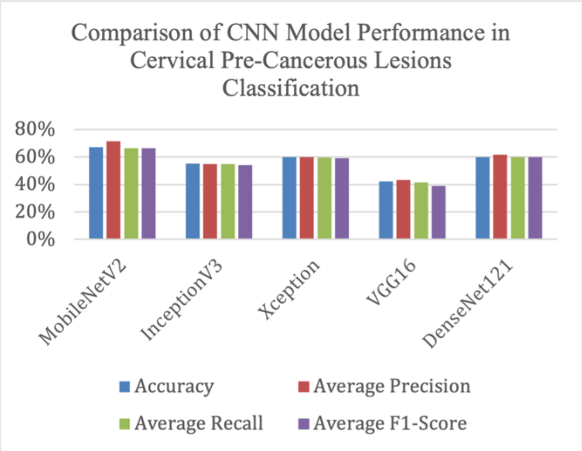


Fig. 7. CNN Model Performance Compqrison

Based on Table 8 and Figure 1, MobileNetV2 outperformed all other models in this study with the highest accuracy of 67%, while VGG16 had the lowest performance with an accuracy of 42%. MobileNetV2 excelled across all metrics, making it the best overall choice. DenseNet121 demonstrated comparable performance, particularly in precision and F1-score. Xception offered a well-balanced result across classes, while InceptionV3 showed inconsistent metrics, especially in F1-score. VGG16 was the least suitable model for this classification task.

G. Discussion and Limitations

While MobileNetV2 yielded the highest performance among tested CNN architectures, the overall accuracy across models (42% to 67%) highlights the difficulty of the classification task. This can be attributed to the relatively small dataset size and class imbalance, despite augmentation efforts. Models particularly struggled with distinguishing CIN 1, which may visually ensemble both Normal and CIN 2 cases. Futuru work should explore data balancing strategies, larger and more diverse datasets, and potentially ensemble-based models to enhance classification performance.

IV. CONCLUSIONS AND SUGGESTIONS

Based on the research findings and analysis, several conclusions can be drawn. The development of a classification model for pre-cancerous cervical lesions based on colposcopy images into four classes—Normal, CIN 1, CIN 2, and CIN 3—involved several stages, including problem identification, data collection and class labeling, image preprocessing, model architecture selection (MobileNetV2, InceptionV3, Xception, VGG16, and DenseNet121), and model evaluation.

Using identical training parameters—batch size of 32, learning rate of 0.001, and 100 epochs—the best accuracy achieved by each architecture was as follows: MobileNetV2: 67% InceptionV3: 55% Xception: 60% VGG16: 42% DenseNet121: 60%.

From these results, MobileNetV2 achieved the highest accuracy of 67%, indicating that it performed better than the other proposed architectures in classifying pre-cancerous cervical lesions based on colposcopy images across the four classes (Normal, CIN 1, CIN 2, and CIN 3). Meanwhile, VGG16 recorded the lowest accuracy at 42%, making it the least suitable architecture among those evaluated in this study.

Overall, the obtained accuracy levels indicate that the classification performance still leaves room for improvement. Therefore, future research should consider several enhancements: Increasing the dataset size and diversity, to ensure broader generalization and better applicability in real-world clinical settings. Optimizing training configurations to improve model performance. Applying more advanced data augmentation techniques or transfer learning strategies to boost model robustness. Exploring ensemble methods or attention-based CNNs to enhance classification accuracy and generalizability. It is expected that these strategies will result in models that are both dependable and suitable for clinical use in the early identification of cervical cancer using colposcopy images. Despite the promising result, further improvements in data quantity and model optimization are essential before clinical implementation can be considered.

REFERENCES

[1] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, “Survey of Explainable AI Techniques in Healthcare,” Jan. 01, 2023, *MDPI*. doi: 10.3390/s23020634.

[2] N. A. Zebari *et al.*, “A deep learning fusion model for accurate classification of brain tumours in Magnetic Resonance images,” *CAAI Trans Intell Technol*, vol. 9, no. 4, pp. 790–804, Aug. 2024, doi: 10.1049/cit2.12276.

[3] E. O. Simonyan, Joke. A. Badejo, and J. S. Weijin, “Histopathological breast cancer classification using CNN,” *Mater Today Proc*, vol. 105, pp. 268–275, 2024, doi: https://doi.org/10.1016/j.matpr.2023.10.154.

[4] A. Faghihi, M. Fathollahi, and R. Rajabi, “Diagnosis of skin cancer using VGG16 and VGG19 based transfer learning models,” *Multimed Tools Appl*, vol. 83, no. 19, pp. 57495–57510, 2024, doi: 10.1007/s11042-023-17735-2.

[5] H. Sung *et al.*, “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, May 2021, doi: https://doi.org/10.3322/caac.21660.

[6] P. A. Cohen, A. Jhingran, A. Oaknin, and L. Denny, “Cervical Cancer,” *The Lancet*, vol. 393, no. 10167, pp. 169–182, Jan. 2019, doi: 10.1016/S0140-6736(18)32470-X.

[7] J. M. M. Walboomers *et al.*, “Human papillomavirus is a necessary cause

of invasive cervical cancer worldwide," *J Pathol*, vol. 189, no. 1, pp. 12–19, Sep. 1999, doi: https://doi.org/10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F.

[8] J. Liu, L. Li, and L. Wang, "Acetowhite Region Segmentation in Uterine Cervix Images Using a Registered Ratio Image," *Comput Biol Med*, vol. 93, pp. 47–55, 2018, doi: 10.1016/j.compbiomed.2017.12.009.

[9] E. Hussain, L. B. Mahanta, K. A. Borbora, H. Borah, and S. S. Choudhury, "Exploring explainable artificial intelligence techniques for evaluating cervical intraepithelial neoplasia (CIN) diagnosis using colposcopy images," *Expert Syst Appl*, vol. 249, p. 123579, 2024, doi: https://doi.org/10.1016/j.eswa.2024.123579.

[10] N. A. Alias *et al.*, "Pap Smear Images Classification Using Machine Learning: A Literature Matrix," Dec. 01, 2022, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/diagnostics12122900.

[11] P. B. Shanthi, K. S. Hareesha, and R. Kudva, "Automated Detection and Classification of Cervical Cancer Using Pap Smear Microscopic Images: A Comprehensive Review and Future Perspectives," 2022, *Engineered Science Publisher*. doi: 10.30919/es8d633.

[12] O. Yaman and T. Tuncer, "Exemplar pyramid deep feature extraction based cervical cancer image classification model using pap-smear images," *Biomed Signal Process Control*, vol. 73, p. 103428, 2022, doi: https://doi.org/10.1016/j.bspc.2021.103428.

[13] B. J. Cho *et al.*, "Classification of Cervical neoplasms on Colposcopic Photography using Deep Learning," *Sci Rep*, vol. 10, no. 1, pp. 1–10, Aug. 2020, doi: 10.1038/s41598-020-70490-4.

[14] V. Chandran *et al.*, "Diagnosis of Cervical Cancer based on Ensemble Deep Learning Network using Colposcopy Images," *Biomed Res Int*, vol. 2021, no. 1, May 2021, doi: 10.1155/2021/5584004.

[15] S. Saini, K. Ahuja, S. Chennareddy, and K. Boddupalli, "Deep Learning Descriptor Hybridization with Feature Reduction for Accurate Cervical Cancer Colposcopy Image Classification," *Pattern Recognit Lett*, May 2024, doi: 10.48550/arXiv.2405.01600.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.

[18] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions."

[19] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ICLR*, Apr. 2015, [Online]. Available: http://arxiv.org/abs/1409.1556

[20] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

[21] D. A. Minciună *et al.*, "AnnoCerv: A new dataset for feature-driven and image-based automated colposcopy analysis," *Acta Universitatis Sapientiae, Informatica*, vol. 15, no. 2, pp. 306–329, Dec. 2023, doi: 10.2478/ausi-2023-0019.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Road Damage Detection Using Yolov9-Based Imagery

Febrian Akbar Azhari [1], Tatang Rohana [2], Kiki Ahmad Baihaqi *[3], Ahmad Fauzi [4]
Department of Computer Science [1], [2], [3], [4]
University of Buana Perjuangan Karawang
Karawang, Indonesia
if21.febrianazhari@mhs.ubpkarawang.ac.id[1], tatang.rohana@ubpkarawang.ac.id[2], kikiahmad@ubpkarawang.ac.id[3], afauzi@ubpkarawang.ac.id[4]

***Abstract*— Road damage is one of the leading factors contributing to traffic accidents. Rapid identification and repair of damaged roads are crucial in road infrastructure management. This study aims to develop an effective method for detecting road damage, utilizing the YOLOv9 algorithm as a key component, such as cracks and potholes, using the Convolutional Neural Network (CNN) approach. YOLOv9 was chosen due to its efficient architecture, which enables real-time object detection, and its proven effectiveness in various object detection tasks. An annotated dataset of road images was used during the model training and testing process. The results show that the YOLOv9 model can accurately detect road damage. The model achieved a precision of 0.85 and a recall of 0.992 for pothole detection, and a precision of 0.94 for crack detection. Evaluation using mAP50 yielded a score of 0.96, while mAP50-95 reached 0.77, indicating strong detection and classification capability. A consistent decline in loss functions during training also signifies effective learning by the model. These findings suggest that YOLOv9 has the potential to be implemented in automated road damage detection systems, which can accelerate maintenance processes and enhance road user safety.**

Keywords: YOLOv9, Road Damage Detection, CNN, Deep Learning

I. INTRODUCTION

Highways have a vital role as infrastructure to support strategic development [1], which requires effective management. Efforts to reduce potential hazards in driving on the highway are very important [2]. Various factors can cause road accidents, one of which is the condition of damaged roads [3]. Proper maintenance of the road network is one of the important steps in reducing the risk of danger to motorists. Given that roads are a major part of the crucial land transportation infrastructure [4], [5], the maintenance of the road network needs to be managed in a sustainable manner [6].

Monitoring road conditions is a very important aspect to minimize the number of accidents caused by bad roads. According to [3], accidents occur due to various factors, and the most dominant factor is undetected or unrepaired road defects. Therefore, monitoring of a large road network is essential to determine the appropriate and different types of maintenance for each section of the road [7].

Continuous monitoring in a region with many roadways presents a challenge, as field surveys must be conducted at various locations throughout the road network. Manual methods used to identify road defects are time-consuming. However, with the development of computer technology, many jobs that previously required human labor can now be replaced by artificial intelligence-based systems.

The development of artificial intelligence has a major role in the future transformation of various industries [8]. In this research, the author adopts the Convolutional Neural Networks (CNN) algorithm model for digital image processing. The CNN used in this study proved to be effective in identifying objects present in digital images of the ground surface [9]. By using cameras and applying the CNN model, the process of field surveys to detect road damage can be carried out more quickly, which allows repair of damaged roads to be carried out more immediately, thereby reducing the number of accidents caused by poor road conditions.

The study conducted by [10] investigated the use of camera image processing technology to automatically detect road defects. The study by [10] applied an artificial neural network-based method using the Mask R-CNN algorithm was applied to solve the road defect problem, perform classification, as well as extract important features in the image. The research conducted by [11] used pavement image data taken using an Unmanned Aerial Vehicle (UAV). They managed to distinguish between normal and damaged pavements, including detecting cracks and potholes. The results of this study show that remote sensing systems using UAVs can provide an effective tool for monitoring the condition of asphalt pavements.

Previous studies have demonstrated the success of Convolutional Neural Networks (CNNs) in various image classification and object detection tasks. [12] utilized CNNs to distinguish between damaged (with holes, tears, and dents) and undamaged cardboard packaging, achieving a high accuracy of 95.77%. [13] also employed CNNs to classify five types of mangoes, resulting in a 99.56% accuracy. Furthermore, [14] applied CNNs to recognize Sundanese script patterns, with varying accuracy depending on the image source, reaching 100% for computer font images. The You Only Look Once (YOLO) method, frequently used to test CNN architectures, has also been successfully implemented in detecting rat pests in

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2377, Copyright ©2025
Submitted : May 16, 2025, Revised : May 24, 2025, Accepted : May 26, 2025, Published : May 28, 2025

farmland using YOLOv5 (88% accuracy) [15] and classifying rice types using YOLOv3, with results differing based on rice stacking (60% when stacked, 100% when not) [16]. These studies highlight the broad applicability of CNNs and YOLO-based methods in image analysis and object detection, providing a foundation and motivation for employing YOLOv9 in road damage detection.

Although several studies have applied CNN-based methods such as Mask R-CNN or earlier versions of YOLO for road damage detection, they often rely on publicly available datasets, which may not represent local road conditions. Furthermore, the use of more advanced models like YOLOv9, which offer enhanced detection capabilities, has not yet been explored extensively in this context. This presents a significant research gap in terms of both model effectiveness and dataset relevance. To address this, the present study applies the YOLOv9 model to a custom-collected local dataset, aiming to provide more accurate and context-aware detection of various types of road damage.

II. MATERIALS AND METHOD

This research is designed to assess the performance of a road defect detection model using the YOLOv9 framework, a Convolutional Neural Network (CNN) architecture. After the problem is obtained and analyzed, a literature study is carried out to obtain references and relevant literature in solving the problem. This literature study is carried out by looking for references in the form of books or journals, either through internet media or other sources related to the use of Convolutional Neural Networks (CNN) and YOLOv9 in road damage detection. The literature found will provide a theoretical foundation and practical guidance in designing and implementing road defect detection models. This study adopts a systematic methodology consisting of five primary stages: data collection, data preprocessing, model training, model evaluation, and result analysis. Each stage plays a crucial role in building a robust road damage detection system using the YOLOv9 architecture. The overall research workflow is illustrated in Figure 1, which ensures the logical flow of the process from data acquisition to final evaluation.

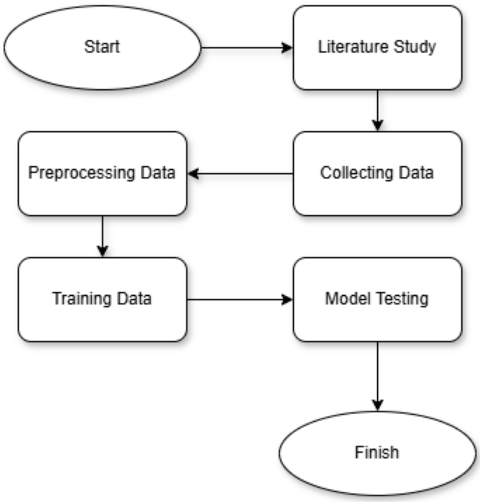


Fig. 1. Research Methods

Figure 1 presents the sequence of research stages: (1) Data Collection involves capturing high-resolution road images using a smartphone camera; (2) Data Preprocessing includes annotation, resizing, grayscale conversion, contrast adjustment, and augmentation; (3) Model Training utilizes YOLOv9 for learning features of road damage; (4) Model Evaluation measures performance metrics such as precision, recall, and mAP; and (5) Result Analysis interprets detection outcomes and training effectiveness.

A. Collecting Data

The dataset used in this research consists of road images that show various types of damage, such as cracks and potholes. The image capture process is done using a Techno brand smartphone camera device with a camera resolution of 108 Megapixels to ensure that the details of road damage can be seen clearly. The camera was used from a 45-degree viewing angle with a distance of one meter per object to optimally capture road conditions.

The research was conducted in Karawang Regency with data collection conducted directly in the field. Images were taken of paved and concrete roads in dry conditions to ensure the imaging results were not disturbed by environmental factors, such as water or reflection. The data collection locations included different types of roads, such as main roads, neighborhood roads, and industrial areas. As part of the dataset collection, several images were captured to represent different types of road damage. One common form is the occurrence of surface cracks, which are often caused by thermal expansion, excessive vehicle load, or degradation of road materials. These cracks vary in size and pattern, making them a challenge for consistent detection through conventional methods. An example of such damage is illustrated in Figure 2. This variation aims to cover the different types of road damage that may arise due to differences in environmental conditions and road usage levels.

Fig. 2. Cracked Road Image

Figure 2 shows a road segment with visible surface cracks, which appear as linear fractures along the pavement. The cracks typically emerge due to environmental stress, repeated traffic loads, or poor construction quality. Such defects are often subtle, and their detection requires high-resolution imagery

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2377, Copyright ©2025
Submitted : May 16, 2025, Revised : May 24, 2025, Accepted : May 26, 2025, Published : May 28, 2025

combined with advanced deep learning models for accurate classification. Another prevalent type of road damage captured in the dataset is potholes. These defects typically form due to the weakening of pavement structure from prolonged water infiltration, repeated traffic loads, or insufficient maintenance. Potholes pose a significant safety risk to road users, especially two-wheeled vehicles, due to the sudden depressions they create on the road surface. A visual example of this type of damage is presented in Figure 3**.**

Fig. 3. Photole Road Image

This figure depicts a road segment containing one or more potholes—circular or irregular depressions on the pavement surface. These occur when water seeps into cracks, softens the underlying soil, and leads to collapse under the weight of vehicles. Potholes are generally easier to detect than surface cracks due to their distinct contours and depth variation, making them suitable for detection using object recognition models such as YOLOv9.

B. Preprocessing Data

This is an important step in data processing before further analysis or machine learning model training. In this research, the data preprocessing stage includes annotation, auto-orient, resize, grayscale, auto-adjust contrast and then perform augmentation to multiply the data. Annotation in this study using the Roboflow.com website is done manually on all images in the dataset using a bounding box to surround objects with low precision. Auto-orientation ensures consistent image orientation [17] to improve the accuracy of road damage detection.

Resize the process of adjusting the image size to fit the predefined dimensions [18], which aims to ensure size consistency and reduce computational complexity in image processing. Grayscale is a step that converts a color image into a gray-scale image. At this stage, all existing color information, such as differences in red, green, and blue hues, are completely removed. This removal of color information aims to make the model focus more on the important elements in the image, such as the visual structure, texture patterns, and shapes of objects, such as cracks and potholes on the road surface, without being affected by color variations that are irrelevant to the purpose of detection.

Next, the auto-adjust contrast process is performed using the adaptive histogram equalization technique. This method serves to improve image contrast by locally adjusting the pixel intensity distribution in the image over a small area. With sharper contrast, fine details such as crack edges, break lines, and hole contours become more visible and easily recognized by the detection model.

Augmentation is used to expand the amount and variety of training data in machine learning and computer vision. By modifying the original data through various techniques, augmentation aims to increase the diversity and quality of the training data, so that the model can learn more effectively and be able to deal with future data variations. In the data augmentation stage, we apply several techniques to expand the abundance of training data and improve model robustness.

The augmentation techniques applied include Rotating the image horizontally and vertically by flipping the image from left to right or from top to bottom, to simulate variations in the position of objects in the image. Rotation by 90 degrees or within a certain range: Rotate the image in a fixed angle (90°) or random angle so that the model can recognize objects even in different orientations. linear shift of the image.

Exposure level adjustment or changing the brightness of the image to create varying lighting conditions such as daylight, cloudy, or shadowy conditions. Noise augmentation by adding visual noise such as random spots or grain to train the model to remain robust despite image noise or imperfections. The application of these various augmentation techniques aims to enrich the training dataset and improve the model's ability to deal with the variety of situations that may occur in road damage images.

To build an effective YOLO configuration, a predetermined portion of test and trial data is necessary. In order to produce an accurate YOLO model, the initial data must be labeled and tested. The dataset consists of 1000 road images and is divided into three parts: 70% (700 images) for training, 10% (100 images) for validation, and 20% (200 images) for testing. This division of the dataset ensures the model can be properly trained, tested to measure its performance, and validated to avoid overfitting.

C. Yolo v9

YOLO (You Only Look Once) combines various object detection capabilities into a single neural network [19]. YOLO predicts each annotation box using the overall image features [20]. YOLO predicts that each bounding box for all object classes will open simultaneously in an image [21]. This shows that YOLO considers all aspects of the image, including all objects in it. In general, the YOLOv9 architecture consists of several main components:

1. Backbone: This part is responsible for extracting important features from the input image. YOLOv9 uses a more efficient and robust backbone than its predecessor, allowing the model to capture richer details.
2. Neck: This component serves to combine the features extracted by the backbone. By combining features from different levels, the model can obtain a more

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2377, Copyright ©2025
Submitted : May 16, 2025, Revised : May 24, 2025, Accepted : May 26, 2025, Published : May 28, 2025

comprehensive representation of the objects in the image.

3. Head: This part is responsible for performing bounding box and object class prediction. YOLOv9 uses a more sophisticated prediction mechanism, allowing the model to detect objects more accurately, especially small and adjacent objects.

The main innovation developed by YOLOv9 is GELAN (General Efficient Layer Aggregation Network) This new architecture is designed to maximize accuracy while minimizing the number of parameters and FLOPs (floating-point operations). GELAN allows the model to more efficiently process visual information and detect small objects.

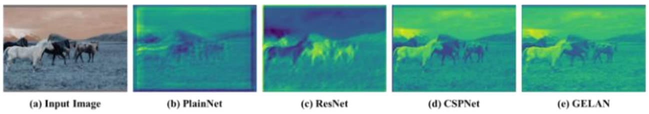


Fig. 4. How Yolov9 Works

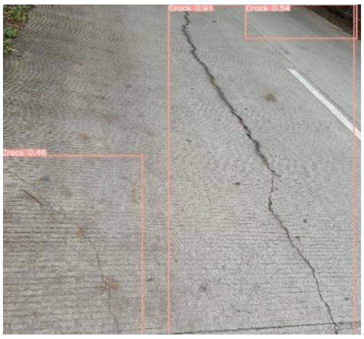
Figure 4, sourced from [22], presents the visualization results of the output feature maps generated by random initial weights across various network architectures. The visualized data includes: (a) the input image, (b) PlainNet, (c) ResNet, (d) CSPNet, and (e) the proposed GELAN. Upon analyzing this figure, it becomes evident that, for each of the tested architectures, the information transmitted to the objective function for loss calculation is lost at varying rates. Specifically, the proposed architecture (GELAN) stands out by successfully retaining the most comprehensive information, while also providing the most reliable gradient information for the calculation of the objective function. This result highlights the superior capacity of GELAN to preserve and process relevant data through the network, offering a more robust approach to training deep learning models.

III. Result and Discussion

The results of the image training and validation process have been successfully formed into a comprehensive model. The primary theme of this paper centers around the findings and advancements in the field of road damage detection. In order to thoroughly analyze and evaluate real-world road damage, we opted to utilize manual data collection as a part of the research scenario, as opposed to relying on publicly available data. This decision was made deliberately, as the database used in the manual data collection process is more representative of the specific context and scope of the research addressed in this paper. By using this tailored data collection approach, we ensure that the model is better aligned with the practical aspects of road damage detection, offering insights that are more relevant and applicable to real-life scenarios.

The YOLO modeling data was tested to detect objects in images. To assess the accuracy of the object data classifier, calculations are performed using packages such as matplotlib, numpy, sklearn, and torch. The training results from these packages will be displayed in the form of diagrams and images, which will facilitate data analysis. In the test, the epoch used was 100; with a training batch size of 16, which means 16 images were processed at once in one iteration. In addition, the image size used was 640 x 640 pixels, and all training was conducted using Google Collaboratory. Table 1 shows the detection results of cracked and potholed roads.

TABLE I. Detection Results

Image	Damage type
	Pothole
	Crack

After detection testing, road defects with a confidence of 0.52 were detected in the pothole type, while details of small cracks at the end of the road could be detected with a confidence of 0.54. Although the confidence level obtained is quite low, the faint objects can still be detected well. On the other hand, objects that are more clearly visible to the naked eye show a higher confidence level, especially for potholes with a confidence level of 0.95. For cracked road damage, the confidence level recorded was 0.91. The resulting F1-Score graph shows that potholes are easier to detect than cracks. This can be seen in Figure 5, which displays the F1-Score curve.

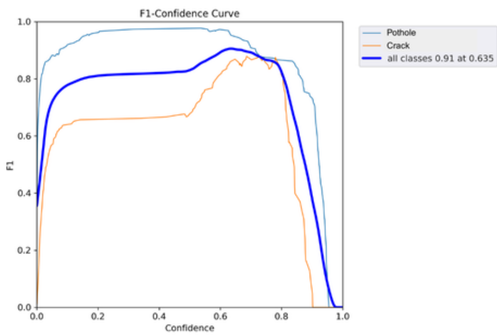


Fig. 5. F1-Score Curve

Meanwhile, the results of the Precision Confidence measurement showed that all types of road defects, without

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2377, Copyright ©2025
Submitted : May 16, 2025, Revised : May 24, 2025, Accepted : May 26, 2025, Published : May 28, 2025

exception, achieved a score of 0.85. This reflects a fairly high level of accuracy in road defect detection. This result can be clearly seen in the curve depicted in Figure 6, which presents a visualization of the comparison of the detection confidence level with the achieved precision. The curve provides a more complete picture of the model's accuracy in detecting road defects, which further strengthens the validity of the findings in this study.

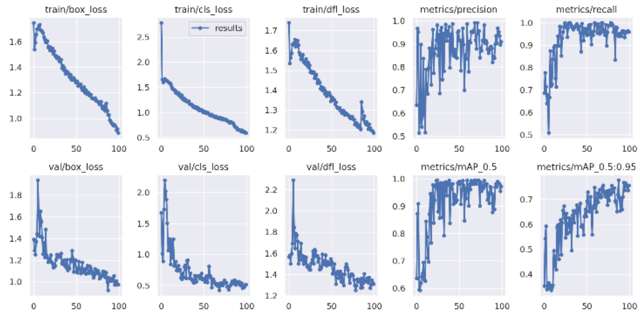


Fig. 6. Loss Chart

The Precision Recall result for pothole road damage is 0.992, and for cracked roads it reaches 0.94, which can be seen in Figure 7.

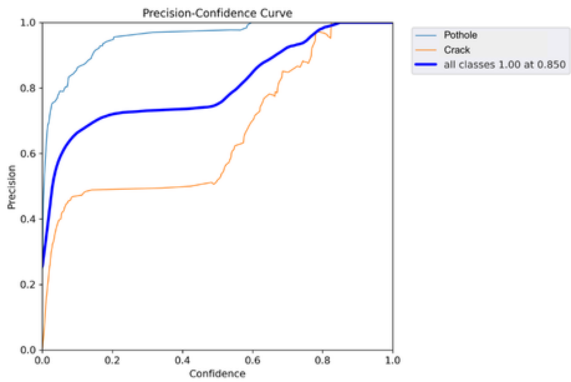


Fig. 7. Precision Confidence Curve

In addition to displaying the Precision results, in this study we also present graphs depicting the loss values during the training and testing process. These graphs include various metrics, ranging from train loss to validation loss, which provide a detailed picture of the model's performance throughout the training process. This process is important to evaluate how well the model can generalize data that was not seen before, as well as to identify potential overfitting or underfitting. These results can be clearly seen in Figure 8, which shows a comparison between train loss and validation loss at various epochs, providing further insight into the effectiveness of our model training and validation process.

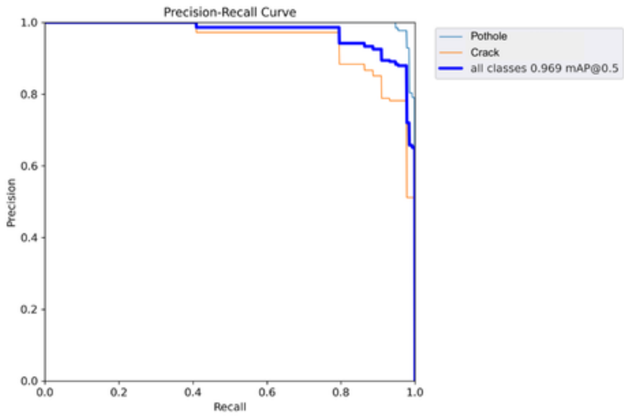


Fig. 8. Precision Recall Curve

In Figure 8 the training result graph displays the box_loss, cls_loss, and dfl_loss metrics for the train and val data. These three losses show a decrease at each epoch, indicating that the model is able to learn the task well. Box_loss measures the fit of the bounding box, cls_loss is related to object detection, and dfl_loss indicates classification accuracy. The val graph shows larger fluctuations than the train due to the difference in the amount of data. In addition, the precision, recall, mAP50, and mAP50-95 metrics graphs show an increasing trend every

epoch, reflecting the improvement in classification accuracy and performance. The mAP50 value reaches 0.96, indicating excellent detection accuracy, while mAP50-95 reaches 0.77, indicating a low error rate and good classification performance.

IV. Result and Discussion

The YOLOv9 model demonstrated strong performance in detecting road defects, specifically cracks and potholes. The model achieved a precision of 0.85 and a recall of 0.992 for pothole detection, and a precision of 0.94 for crack detection, indicating its ability to accurately identify and classify these defects. Furthermore, the mAP50 of 0.96 and mAP50-95 of 0.77 further support the model's high detection and classification accuracy. These results demonstrate the model's effectiveness in handling variations in detection difficulty across different defect types.

Furthermore, the model exhibited a consistent decrease in the loss function value throughout the training process. This decline indicates effective learning and optimization of the model's performance during training. The successful implementation of YOLOv9 for road defect detection demonstrates its potential as a foundation for developing an automated road monitoring system. Such a system could significantly aid relevant authorities in expediting the identification and remediation of road defects, thereby enhancing road user safety and improving the efficiency of road infrastructure maintenance.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2377, Copyright ©2025
Submitted : May 16, 2025, Revised : May 24, 2025, Accepted : May 26, 2025, Published : May 28, 2025

References

[1] T. Ngabalin, A. F. Habibie, and E. Darmawan, "Strategi Pengembangan Ekosistem Ekonomi Kreatif Dalam Mendukung Kebijakan Pariwisata Di Kota Tanjungpinang," *J. Ilmu Adm. Negara*, vol. 11, no. 01, pp. 13–21, 2023, doi: 10.31629/juan.v11i01.5839.

[2] R. Der Sarkissian, C. Abdallah, J. M. Zaninetti, and S. Najem, "Modelling intra-dependencies to assess road network resilience to natural hazards," *Nat. Hazards*, vol. 103, no. 1, 2020, doi: 10.1007/s11069-020-03962-5.

[3] M. T. A. Siregar, "Upaya Yang Dapat Dilakukan Oleh Korban/ Pengguna Jalan Meminta Pertanggungjawaban Pidana Penyelenggara Jalan Atas Terjadinya Kecelakaan Akibat Jalan Rusak," *EduTech J. Ilmu Pendidik. dan Ilmu Sos.*, vol. 6, no. 1, pp. 37–45, 2020, doi: 10.30596/edutech.v6i1.4393.

[4] F. B. Santosa and A. S. Amal, "Kajian Studi Kelayakan Rencana Pembangunan Jalan Srondol-Sekaran Kota Semarang," *Semin. Keinsinyuran Progr. Stud. Progr. Profesi Ins.*, vol. 3, no. 1, 2023, doi: 10.22219/skpsppi.v3i1.7544.

[5] R. Rivaldo and F. R. Yamali, "Perencanaan Perkerasan Kaku (Rigid Pavement) Ruas Jalan Hitam Ulu-Mentawak Di Kabupaten Merangin (Menggunakan Metode AASHTO 1993)," *J. Talent. Sipil*, vol. 5, no. 1, 2022, doi: 10.33087/talentasipil.v5i1.95.

[6] P. Toscani, W. Sekot, and F. Holzleitner, "Forest roads from the perspective of managerial accounting-empirical evidence from Austria," *Forests*, vol. 11, no. 4, 2020, doi: 10.3390/F11040378.

[7] N. I. Sembiring, R. Siahaan, and P. Naibaho, "Analysis of Damage Conditions on the Berastagi-Simpang Empat Road, Karo Regency, Using PCI and Sdi Methods," *Tjyybjb.Ac.Cn*, vol. 3, no. 2, pp. 58–66, 2021, [Online]. Available: http://www.tjyybjb.ac.cn/CN/article/downloadArticleFile.do?attachType=PDF&id=9987

[8] S. Suryadi and F. A. P. Nasution, "Revolusi Industri, Tren Pekerjaan Masa Depan, dan Posisi Indonesia," *J. Ketenagakerjaan*, vol. 18, no. 2, pp. 124–141, 2023, doi: 10.47198/jnaker.v18i2.237.

[9] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, 2021, doi: 10.1016/j.ymssp.2020.107398.

[10] Q. Chen, X. Gan, W. Huang, J. Feng, and H. Shim, "Road damage detection and classification using mask R-CNN with DenseNet backbone," *Comput. Mater. Contin.*, vol. 65, no. 3, 2020, doi: 10.32604/cmc.2020.011191.

[11] Y. Pan, X. Chen, Q. Sun, and X. Zhang, "Monitoring Asphalt Pavement Aging and Damage Conditions from Low-Altitude UAV Imagery Based on a CNN Approach," *Can. J. Remote Sens.*, vol. 47, no. 3, 2021, doi: 10.1080/07038992.2020.1870217.

[12] A. Antoni, T. Rohana, and A. R. Pratama, "Implementasi Algoritma Convolutional Neural Network Untuk Klasifikasi Citra Kemasan Kardus Defect dan No Defect," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 1941–1950, 2023, doi: 10.47065/bits.v4i4.3270.

[13] R. Yati, T. Rohana, and A. R. Pratama, "Klasifikasi Jenis Mangga Menggunakan Algoritma Convolutional Neural Network," *J. Media Inform. Budidarma*, vol. 7, no. 3, p. 1265, 2023, doi: 10.30865/mib.v7i3.6445.

[14] A. Kirana, H. Hikmayanti, and J. Indra, "Pengenalan Pola Aksara Sunda dengan Metode Convolutional Neural Network," *Sci. Student J. Information, Technol. Sci.*, vol. 1, no. 2, pp. 95–100, 2020, [Online]. Available: http://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/download/19/15

[15] K. Ahmad Baihaqi and C. Zonyfar, "Deteksi Lahan Pertanian Yang Terdampak Hama Tikus Menggunakan Yolo v5," *Syntax J. Inform.*, vol. 11, no. 02, pp. 01–11, 2022, doi: 10.35706/syji.v11i02.7226.

[16] S. Ma'arif, T. Rohana, and K. Ahmad Baihaqi, "Deteksi Jenis Beras Menggunakan Algoritma YOLOv3," *Sci. Student J. Information, Technol. Sci.*, vol. III, pp. 219–226, 2022, [Online]. Available: http://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/443/357

[17] S. Informasi, F. N. Fajri, and W. G. Priambodo, "Jurnal Advance Research Informatika Fire and Smoke Object Detection Using Mask R-CNN," vol. 2, no. 2, pp. 1–7, 2024.

[18] V. R. Atvira, "Deteksi Kualitas Kemurnian Susu Sapi Melalui Pengolahan Citra Digital Menggunakan Metode Active Contour Dengan Klasifikasi K-Nearest Neighbor," *Univ. Telkom Bandung*, vol. 6, no. 2, pp. 3830–3837, 2019, [Online]. Available: https://openlibrary.telkomuniversity.ac.id/pustaka/152822/

[19] U. Rahardja, M. A. Ngad, S. Millah, E. P. Harahap, and Q. Aini, "Blockchain Application in Educational Certificates and Verification Compliant with General Data Protection Regulations," in *2022 10th International Conference on Cyber and IT Service Management, CITSM 2022*, 2022. doi: 10.1109/CITSM56380.2022.9935909.

[20] N. Lutfiani, S. Wijono, U. Rahardja, A. Iriani, and E. A. Nabila, "Artificial Intelligence Based on Recommendation System for Startup Matchmaking Platform," in *2022 IEEE Creative Communication and Innovative Technology, ICCIT 2022*, 2022. doi: 10.1109/ICCIT55355.2022.10118708.

[21] T. Hongsuchon *et al.*, "Brand Experience on Brand Attachment: The Role of Interpersonal Interaction, Feedback, and Advocacy," *Emerg. Sci. J.*, vol. 7, no. 4, 2023, doi: 10.28991/ESJ-2023-07-04-014.

[22] C. Wang, I. Yeh, and H. M. Liao, "Supplementary of YOLOv9," pp. 1–7.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2377, Copyright ©2025
Submitted : May 16, 2025, Revised : May 24, 2025, Accepted : May 26, 2025, Published : May 28, 2025

Quantum Perceptron in Predicting the Number of Visitors to E-Commerce Websites in Indonesian

Solikhun[1]*, Dinda Carissa Arishandy[2], Ela Roza Batubara[3], Poningsih[4]
Department of Informatics Engineering [1], [2], Department of Information Systems[3], Departemnt of Master Informatics[4]
STIKOM Tunas Bangsa
Pematangsiantar, Indonesia
solikhun@amiktunasbangsa.ac.id [1], dindaarishandy1@gmail.com [2], ela@amiktunasbangsa.ac.id [3], poningsih@amiktunasbangsa.ac.id [4]

***Abstract*— In the current digital era, e-commerce has become the backbone of Indonesia's digital economy, which is experiencing rapid growth. However, competition in this industry is becoming increasingly fierce, indicating the importance of predicting the number of website visitors for an effective marketing strategy. Quantum Perceptron, the latest quantum computing innovation, promises a more accurate and efficient approach compared to conventional methods such as classical Perceptron. This research proposes the use of Quantum Perceptron to predict the number of visitors on large e-commerce platforms in Indonesia. The data used in the research is data on the number of e-commerce visitors obtained from the katadata.com website. Data from Shopee, Tokopedia, Lazada, Blibli, and Bukalapak were used to analyze and compare predictions with classical perceptron methods, showing the significant potential of Quantum Perceptron in supporting the development of more efficient business strategies. The research results show that the Quantum Perceptron algorithm can make predictions very well compared to the classical perceptron, proven by the Quantum Perceptron having a perfect accuracy of 100% with a total of 2 epochs while the classical perceptron has 100% accuracy with a total of 10 epochs. Quantum perceptron has better performance and shorter time, this can be seen from the smaller number of epochs.**

Keywords*— *Quantum Perceptron, E-Commerce, Website Visitor Prediction, Quantum Computing, Marketing Strategies

I. INTRODUCTION

The rapid development of quantum technology and artificial intelligence (AI) has opened up huge opportunities in various sectors, including e-commerce. One application of this technology is the *quantum perceptron*, which combines the power of quantum computing with traditional machine learning models to produce faster and more accurate predictions [1], [2]. Quantum computing, which is supported by parallel processing capabilities through *qubits*, is able to significantly speed up large-scale data processing compared to classical computing [3], [4]. If classical perceptron computing can only process data in the range 0 and 1, then quantum perceptron processes data in the range 0, 1 and a combination of both 0 and 1. The initial data will be transformed into the range 0, 1 to be tested using a quantum perceptron.

In Indonesia, e-commerce has become an important part of the digital economy, with the number of visitors to e-commerce websites increasing every year. Predicting the number of visitors is crucial in determining marketing strategies, stock management, and other business decision-making [5]. However, conventional prediction methods based on classical computing often suffer from limitations in handling large data volumes and complex patterns [6].

The *Quantum perceptron* offers a solution to this challenge by utilizing quantum phenomena such as *superposition* and *entanglement*, which allows these models to process data more efficiently [7], [8]. Recent studies have shown that quantum perceptrons can produce more accurate predictions compared to traditional machine learning models in various applications, including the prediction of the number of website visitors [9]. The use of *quantum perceptron* in the e-commerce sector has the potential to improve prediction accuracy, especially in situations where the data at hand is very large and complex [10], [11].

In 2023, quantum technology began to gain greater attention in practical applications, especially in the commercial sector. Implementations of quantum-based systems have been applied in various fields, ranging from supply chain optimization to market prediction [12]. Particularly in the context of e-commerce, this technology offers a more sophisticated approach to monitoring and analyzing visitor behavior [13], [14]. Research shows that quantum algorithms are able to reduce computation time for predictions that previously required large computational resources [15].

In Indonesia, the rapid growth of the e-commerce market creates an urgent need for more advanced technological solutions to improve business competitiveness [16], [17]. Therefore, the use of *quantum perceptron* can be one way to address this challenge, especially in improving prediction accuracy and enterprise resource optimization [18], [19]. Moreover, the applications of quantum computing are expected to continue growing in the next few years, enabling wider adoption in various sectors of the economy [20].

Research in the field of *quantum machine learning* (QML) has shown promising results. QML not only offers higher computational speed but also an increased ability to discover patterns hidden in data [21], [22]. In 2022, further research

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2334, Copyright ©2025
Submitted : April 22, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

strengthened the understanding that *quantum perceptron* can solve complicated non-linear problems more efficiently [23]. Thus, integrating quantum technology into prediction systems in e-commerce will provide significant advantages for companies looking to improve operational accuracy and effectiveness [24], [25].

This research aims to apply Quantum Perceptron algorithm as a more effective solution for predicting the number of e-commerce visitors in Indonesia. This research is expected to contribute to in supporting the development of more efficient business strategies.

II. METHODS

The Materials and Methods section is crucial in research as it details the framework and procedures used to conduct the study, ensuring the findings' transparency, reproducibility, and reliability. This section describes the stages of the research, data analysis methods, and specific techniques used in the study. By providing a clear methodology, this section allows other researchers to evaluate and replicate the study, thereby contributing to the cumulative knowledge in the field. In this research on predicting website visitor numbers using Quantum Perceptron for e-commerce platforms in Indonesia, the Materials and Methods section will describe the research stages, data analysis techniques, and the application of Quantum Perceptron in transforming visitor prediction data. This section guides an understanding of how the study was conducted, from data collection to analysis and interpretation

A. Research Stages

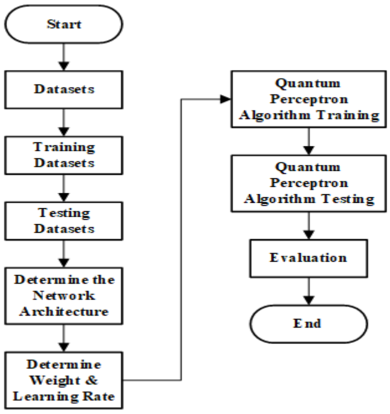


Fig. 1. Research Stages

The research design begins with the data collection stage, where the dataset used for training and testing the model is gathered. The dataset utilized in this study is secondary data, specifically the number of visitors to an e-commerce website, obtained from the website katadata.com. The dataset is then divided into two parts: the training dataset, which is used to train the model, and the testing dataset, which is used to evaluate the model's performance after training. Next, the architecture of the Quantum Perceptron is determined, including the number of layers and units in each layer. Quantum Perceptron is chosen as a solution because it offers shorter computation time and more accurate results compared to classical perceptron. Subsequently, initial parameters such as weights and learning rate are set to ensure the training process runs optimally. The next stage involves training the Quantum Perceptron algorithm using the prepared training dataset. Once the training is complete, the trained model is tested using the testing dataset to assess its performance in making predictions. The testing results are then evaluated to determine the accuracy and effectiveness of the model in solving the prediction task. Once the evaluation results are obtained, the research process is considered complete.

B. Data Analyst

The Materials and Methods section is crucial in research as it details the framework and procedures used to conduct the study, ensuring the findings' transparency, reproducibility, and reliability. This section describes the stages of the research, data analysis methods, and specific techniques used in the study. By providing a clear methodology, this section allows other researchers to evaluate and replicate the study, thereby contributing to the cumulative knowledge in the field. In this research on predicting website visitor numbers using Quantum Perceptron for e-commerce platforms in Indonesia, the Materials and Methods section will describe the research stages, data analysis techniques, and the application of Quantum Perceptron in transforming visitor prediction data. This section guides an understanding of how the study was conducted, from data collection to analysis and interpretation.

C. Qubit Transformation and Superposition

The number of website visitor data for e-commerce is transformed into binary form, which is 0 and 1. The transformation follows the rules outlined in Table 1.

TABLE I. DATA CRITERIA

No	Criteria	Information	Weight
1	Shopee	Shopee < 11200000	00
		Shopee >= 11200000 and Shopee <= 124100000	01
		Shopee >= 124100000 and Shopee <= 237000000	10
		Shopee > 237000000	11
2	Tokopedia	Tokopedia < 11200000	00
		Tokopedia >= 11200000 and Tokopedia <= 124100000	01
		Tokopedia >= 124100000 and Tokopedia <= 237000000	10
		Tokopedia > 237000000	11
3	Lazada	Lazada < 11200000	00
		Lazada >= 11200000 and Lazada <= 124100000	01
		Lazada >= 124100000 and Lazada <= 237000000	10
		Lazada > 237000000	11
4	Blibli	Blibli < 11200000	00

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2334, Copyright ©2025
Submitted : April 22, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

		Blibli >= 11200000 and Blibli <= 124100000	01
		Blibli >= 124100000 and Blibli <= 237000000	10
		Blibli > 237000000	11
5	Bukalapak	Bukalapak < 11200000	00
		Bukalapak >= 11200000 and Bukalapak <= 124100000	01
		Bukalapak >= 124100000 and Bukalapak <= 237000000	10
		Bukalapak > 237000000	11

Table 2 shows data on the number of e-commerce website visitors in Indonesia from January 2023 to September 2023.

TABLE II. VISITOR DATA FOR E-COMMERCE WEBSITES IN INDONESIA

Shopee	Tokopedia	Lazada	Blibli	Bukalapak
171300000	128100000	91200000	28600000	20000000
143600000	108100000	74200000	23200000	17100000
159000000	114900000	84300000	24500000	17100000
165800000	109200000	82500000	33000000	15400000
161200000	106400000	70700000	24400000	17300000
173900000	106000000	70400000	23900000	14000000
199900000	102600000	63400000	28000000	13000000
213400000	99700000	45600000	28300000	12900000
237000000	88900000	47700000	28900000	11200000

The data on the number of e-commerce website visitors in Indonesia is transformed into binary format following the rules in Table 1. The results of this data transformation are shown in Table 3.

TABLE III. TRANSFORMED DATA RESULTS

Shopee	Tokopedia	Lazada	Blibli	Bukalapak
10	10	01	01	01
10	01	01	01	01
10	01	01	01	01
10	01	01	01	01
10	01	01	01	01
10	01	01	01	01
10	01	01	01	01
10	01	01	01	01
10	01	01	01	01

III. RESULT AND DISCUSSION

The Quantum Perceptron utilizes the unique properties of qubits, such as superposition and entanglement, to process information more efficiently and quickly than classical perceptron [7]. Initial research indicates that the Quantum Perceptron can address some limitations of conventional neural networks, such as high computational demands and issues with overfitting.

Visitors from various major e-commerce platforms in Indonesia are expected to provide results that are more relevant to and applicable to the Indonesian e-commerce industry.

The Quantum Perceptron combines quantum concepts with classical perceptron algorithms. This method uses quantum bits (qubits), which are properties of atoms in quantum mechanics, for quantum computation. Qubits can exist in multiple states simultaneously and have different probability values. The steps to implement the Quantum Perceptron can be seen below and in Equations 1, 2, 3, and 4:

1) Initialize all inputs, weights, targets, and biases. Calculate the net value using the formula:
$$|Z_i\rangle = \sum | W_{ij}\rangle . |Xi \qquad (1)$$
2) Calculate the output using the formula:
$$|y_i\rangle = \sum | Z_i\rangle . |Vij\rangle \qquad (2)$$
3) If $|y\rangle \neq |t\rangle$, then:
$$W_{new} = W_{old} + \alpha \cdot (|y\rangle - |t\rangle) \cdot \langle X_i| \qquad (3)$$
4) If not
$$W_{new} = W_{old} \qquad (4)$$
5) If $(y = t)$, then stop.

This quantum perceptron learning architecture uses a 9-1-2 configuration and the dataset of e-commerce website visitor numbers in Indonesia. Initially, weights w and v are given random values of {0,1} as follows:

$W_{1,1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ $W_{1,2} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$ $W_{2,1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$W_{2,2} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ $W_{3,1} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ $W_{3,2} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$

$W_{4,1} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ $W_{4,2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $W_{5,1} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

$W_{5,2} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$ $W_{6,1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ $W_{6,2} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$

$W_{7,1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $W_{7,2} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ $W_{8,1} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

$W_{8,2} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ $V_{1,1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $V_{1,2} = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$

$V_{2,1} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ $V_{2,2} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

The learning rate value tested is 0.1. The learning process starts with the first data from dataset number 1, `10100101` as the input and `01` as the expected output. Next, the output for the hidden units $Z1$ and $Z2$ is calculated.

1) Output Z1

$$= W1,1.|X1> + W2,1.|X2> + W3,1.|X3> + W4,1.|X4> + W5,1.|X5> + W6,1.|X6> + W7,1.|X7> + W8,1.|X8>$$

$$= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.|1> + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.|0> + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.|1> + \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.|0> +$$

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2334, Copyright ©2025
Submitted : April 22, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

$\begin{bmatrix}1&0\\0&0\end{bmatrix}.|0> + \begin{bmatrix}0&1\\1&0\end{bmatrix}.|1> + \begin{bmatrix}1&0\\0&1\end{bmatrix}.|0> + \begin{bmatrix}0&0\\0&1\end{bmatrix}.|1>$

$= \begin{bmatrix}0&1\\1&0\end{bmatrix}.\begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}1&0\\0&1\end{bmatrix}.\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0&0\\0&1\end{bmatrix}.\begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}0&1\\1&1\end{bmatrix}.\begin{bmatrix}1\\0\end{bmatrix}$
$+ \begin{bmatrix}1&0\\0&0\end{bmatrix}.\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0&1\\1&0\end{bmatrix}.\begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}1&0\\0&1\end{bmatrix}.\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0&0\\0&1\end{bmatrix}.\begin{bmatrix}0\\1\end{bmatrix}$
$= \begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix}$
$= \begin{bmatrix}5\\3\end{bmatrix} = [5 \quad 3] = \|5 \quad 3\| = \sqrt{5^2 + 3^2} = \sqrt{5,831} = 1$

2) Output Z2
$= W1,2.|X1> + W2,2.|X2> + W3,2.|X3>$
$+ W4,2.|X4> + W5,2.|X5> + W6,2.|X6>$
$+ W7,2.|X7> + W8,2.|X8>$
$= \begin{bmatrix}0&0\\1&1\end{bmatrix}.|1> + \begin{bmatrix}0&0\\1&0\end{bmatrix}.|0> + \begin{bmatrix}1&0\\1&1\end{bmatrix}.|1> + \begin{bmatrix}1&0\\0&1\end{bmatrix}.|0> +$
$\begin{bmatrix}0&1\\0&1\end{bmatrix}.|0> + \begin{bmatrix}0&0\\1&1\end{bmatrix}.|1> + \begin{bmatrix}0&0\\1&0\end{bmatrix}.|0> + \begin{bmatrix}1&0\\1&1\end{bmatrix}.|1>$
$= \begin{bmatrix}0&0\\1&1\end{bmatrix}.\begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}0&0\\1&0\end{bmatrix}.\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}1&0\\1&1\end{bmatrix}.\begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}1&0\\0&1\end{bmatrix}.\begin{bmatrix}1\\0\end{bmatrix} +$
$\begin{bmatrix}0&1\\0&1\end{bmatrix}.\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0&0\\1&1\end{bmatrix}.\begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}0&0\\1&0\end{bmatrix}.\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}1&0\\1&1\end{bmatrix}.\begin{bmatrix}0\\1\end{bmatrix}$
$= \begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0\\0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix}$
$= \begin{bmatrix}1\\6\end{bmatrix} = [1 \quad 6] = \|1 \quad 6\| = \sqrt{1^2 + 6^2} = \sqrt{6,08276} = 1$

3) Output Y1
$= V1,1.|Z1> + V2,1.|Z2>$
$= \begin{bmatrix}1&0\\0&1\end{bmatrix}.|1 + \begin{bmatrix}0&0\\1&0\end{bmatrix}.|1$
$= \begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}0\\0\end{bmatrix}$
$= \begin{bmatrix}0\\1\end{bmatrix} = [0 \quad 1] = \|0 \quad 1\| = \sqrt{0^2 + 1^2} = \sqrt{1} = 1$

4) Output Y2
$= V1,2.|Z1> + V2,2.|Z2>$
$= \begin{bmatrix}0&0\\1&1\end{bmatrix}.|1 + \begin{bmatrix}1&0\\0&0\end{bmatrix}.|1$
$= \begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}0\\0\end{bmatrix}$
$= \begin{bmatrix}0\\1\end{bmatrix} = [0 \quad 1] = \|0 \quad 1\| = \sqrt{0^2 + 1^2} = \sqrt{1} = 1$

Then, the temporary outputs $Y1$ and $Y2$ are compared with the expected outputs where $Y1 = 0$ and $Y2 = 0$. Since $1 \neq 0$ and $1 \neq 0$, the weights $Wi,j\, Vi,j$ from $|X1>$ to $|X8>$ are updated, and the error value is computed. First, weight updates are performed for $W1,1$ to $W8,1, V1,1, V2,1$, where $Y1 \neq T1$ as follows:

1) Weight $W1,1$ new = $W1,1$ old + $\alpha.(|Y1> - |T1>).< X1|$
$= \begin{bmatrix}0&1\\1&0\end{bmatrix} + 0,1.(1 > -0 >).< 1| = \begin{bmatrix}0&1\\1&0\end{bmatrix} +$
$0,1.\left(\begin{bmatrix}0\\1\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix}\right).[0 \quad 1] = \begin{bmatrix}0&1\\1&0\end{bmatrix} + 0,1.\begin{pmatrix}-1\\1\end{pmatrix}.$
$[0 \quad 1] = \begin{bmatrix}0&1\\1&0\end{bmatrix} + 0,1.\begin{pmatrix}0&-1\\0&-1\end{pmatrix}$
$= \begin{bmatrix}0&1\\1&0\end{bmatrix} + \begin{bmatrix}0&-0,1\\0&0,1\end{bmatrix} = \begin{bmatrix}0&0,9\\1&0,1\end{bmatrix}$

2) Weight $W2,1$ new = $W2,1$ old + $\alpha.(|Y1> - |T1>).< X2|$
$= \begin{bmatrix}1&0\\0&1\end{bmatrix} + 0,1.(1 > -0 >).< 0| = \begin{bmatrix}1&0\\0&1\end{bmatrix} +$
$0,1.\left(\begin{bmatrix}0\\1\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix}.\ [1 \quad 0]\right) = \begin{bmatrix}1&0\\0&1\end{bmatrix} + 0,1.\begin{pmatrix}-1\\1\end{pmatrix}.$
$[1 \quad 0] = \begin{bmatrix}1&0\\0&1\end{bmatrix} + 0,1.\begin{pmatrix}-1&0\\1&0\end{pmatrix}$
$= \begin{bmatrix}1&0\\0&1\end{bmatrix} + \begin{bmatrix}-0,1&0\\0,1&0\end{bmatrix} = \begin{bmatrix}0,9&0\\0,1&1\end{bmatrix}$

3) Weight $W3,1$ new = $W3,1$ old + $\alpha.(|Y1> - |T1>).< X3|$
$= \begin{bmatrix}0&0\\0&1\end{bmatrix} + 0,1.(1 > -0 >).< 1| = \begin{bmatrix}0&0\\0&1\end{bmatrix} +$
$0,1.\left(\begin{bmatrix}0\\1\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix}.\ [0 \quad 1]\right) = \begin{bmatrix}0&0\\0&1\end{bmatrix} + 0,1.\begin{pmatrix}-1\\1\end{pmatrix}.$
$[1 \quad 0] = \begin{bmatrix}0&0\\0&1\end{bmatrix} + 0,1.\begin{pmatrix}0&-1\\0&-1\end{pmatrix}$
$= \begin{bmatrix}0&0\\0&1\end{bmatrix} + \begin{bmatrix}0&-0,1\\0&0,1\end{bmatrix} = \begin{bmatrix}0&-0.1\\0&1,1\end{bmatrix}$

4) Weight $W4,1$ new = $W4,1$ old + $\alpha.(|Y1> - |T1>).< X4|$
$= \begin{bmatrix}0&1\\1&1\end{bmatrix} + 0,1.(1 > -0 >).< 0| = \begin{bmatrix}0&1\\1&1\end{bmatrix} +$
$0,1.\left(\begin{bmatrix}0\\1\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix}.\ [1 \quad 0]\right) = \begin{bmatrix}0&1\\1&1\end{bmatrix} + 0,1.\begin{pmatrix}-1\\1\end{pmatrix}.$
$[1 \quad 0] = \begin{bmatrix}0&1\\1&1\end{bmatrix} + 0,1.\begin{pmatrix}-1&0\\1&0\end{pmatrix}$
$= \begin{bmatrix}0&1\\1&1\end{bmatrix} + \begin{bmatrix}-0,1&0\\0,1&0\end{bmatrix} = \begin{bmatrix}-0,1&1\\1,1&1\end{bmatrix}$

5) Weight $W5,1$ new = $W5,1$ old + $\alpha.(|Y1> - |T1>).< X5|$
$= \begin{bmatrix}1&0\\0&0\end{bmatrix} + 0,1.(1 > -0 >).< 0| = \begin{bmatrix}1&0\\0&0\end{bmatrix} +$
$0,1.\left(\begin{bmatrix}0\\1\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix}.\ [1 \quad 0]\right) = \begin{bmatrix}1&0\\0&0\end{bmatrix} + 0,1.\begin{pmatrix}-1\\1\end{pmatrix}.$
$[1 \quad 0] = \begin{bmatrix}1&0\\0&0\end{bmatrix} + 0,1.\begin{pmatrix}-1&0\\1&0\end{pmatrix}$
$= \begin{bmatrix}1&0\\0&0\end{bmatrix} + \begin{bmatrix}-0,1&0\\0,1&0\end{bmatrix} = \begin{bmatrix}0,9&0\\0,1&0\end{bmatrix}$

6) Weight $W6,1$ new = $W6,1$ old + $\alpha.(|Y1> - |T1>).< X6|$
$= \begin{bmatrix}0&1\\1&0\end{bmatrix} + 0,1.(1 > -0 >).< 1| = \begin{bmatrix}0&1\\1&0\end{bmatrix} +$
$0,1.\left(\begin{bmatrix}0\\1\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix}\right).[0 \quad 1] = \begin{bmatrix}0&1\\1&0\end{bmatrix} + 0,1.\begin{pmatrix}-1\\1\end{pmatrix}.$
$[0 \quad 1] = \begin{bmatrix}0&1\\1&0\end{bmatrix} + 0,1.\begin{pmatrix}0&-1\\0&-1\end{pmatrix}$
$= \begin{bmatrix}0&1\\1&0\end{bmatrix} + \begin{bmatrix}0&-0,1\\0&0,1\end{bmatrix} = \begin{bmatrix}0&0,9\\1&0,1\end{bmatrix}$

7) Weight $W7,1$ new = $W7,1$ old + $\alpha.(|Y1> - |T1>).< X7|$
$= \begin{bmatrix}1&0\\0&1\end{bmatrix} + 0,1.(1 > -0 >).< 0| = \begin{bmatrix}1&0\\0&1\end{bmatrix} +$

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2334, Copyright ©2025
Submitted : April 22, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

$$0{,}1.\left(\begin{bmatrix}0\\1\end{bmatrix}-\begin{bmatrix}1\\0\end{bmatrix}.\ \begin{bmatrix}1 & 0\end{bmatrix}\right)=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}+0{,}1.\begin{pmatrix}-1\\1\end{pmatrix}.$$
$$\begin{bmatrix}1 & 0\end{bmatrix}=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}+0{,}1.\begin{pmatrix}-1 & 0\\1 & 0\end{pmatrix}$$
$$=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}+\begin{bmatrix}-0{,}1 & 0\\0{,}1 & 0\end{bmatrix}=\begin{bmatrix}0{,}9 & 0\\0{,}1 & 1\end{bmatrix}$$

8) Weight $W8{,}1$ new = $W8{,}1$ old + $\alpha.(|Y1> -|T1>).<X8|$
$$=\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}+0{,}1.(1>-0>).<1|=\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}+$$
$$0{,}1.\left(\begin{bmatrix}0\\1\end{bmatrix}-\begin{bmatrix}1\\0\end{bmatrix}.\ \begin{bmatrix}0 & 1\end{bmatrix}\right)=\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}+0{,}1.\begin{pmatrix}-1\\1\end{pmatrix}.$$
$$\begin{bmatrix}1 & 0\end{bmatrix}=\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}+0{,}1.\begin{pmatrix}0 & -1\\0 & -1\end{pmatrix}$$
$$=\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}+\begin{bmatrix}0 & -0{,}1\\0 & 0{,}1\end{bmatrix}=\begin{bmatrix}0 & -0.1\\0 & 1{,}1\end{bmatrix}$$

After modifying the weights $W1{,}2$ to $W12{,}2$, $V1{,}1$to $V2{,}2$, the learning process continues with the second data point. This algorithm will update the weights until the expected output (T) matches the current output (Y) or the error value reaches zero. The weight optimization process will continue until the set goal is achieved. After carrying out the testing process, it was found that the prediction result had perfect accuracy, namely 100%.

```
Pengujian Quantum Perceptron
Epoch 1:
  Data 1: Inputs=[1, 1, 1, 0], Expected=0, Output=1, Correct=False
  Data 2: Inputs=[1, 1, 1, 0], Expected=0, Output=1, Correct=False
  Data 3: Inputs=[1, 1, 1, 0], Expected=0, Output=1, Correct=False
  Data 4: Inputs=[1, 1, 1, 0], Expected=0, Output=0, Correct=True
  Data 5: Inputs=[1, 1, 1, 0], Expected=0, Output=0, Correct=True
  Data 6: Inputs=[1, 1, 1, 0], Expected=0, Output=0, Correct=True
  Data 7: Inputs=[1, 1, 1, 0], Expected=0, Output=0, Correct=True
  Data 8: Inputs=[1, 1, 0, 0], Expected=0, Output=0, Correct=True
  Data 9: Inputs=[1, 1, 0, 0], Expected=0, Output=0, Correct=True
Epoch 2:
  Data 1: Inputs=[1, 1, 1, 0], Expected=0, Output=0, Correct=True
  Data 2: Inputs=[1, 1, 1, 0], Expected=0, Output=0, Correct=True
  Data 3: Inputs=[1, 1, 1, 0], Expected=0, Output=0, Correct=True
  Data 4: Inputs=[1, 1, 1, 0], Expected=0, Output=0, Correct=True
  Data 5: Inputs=[1, 1, 1, 0], Expected=0, Output=0, Correct=True
  Data 6: Inputs=[1, 1, 1, 0], Expected=0, Output=0, Correct=True
  Data 7: Inputs=[1, 1, 1, 0], Expected=0, Output=0, Correct=True
  Data 8: Inputs=[1, 1, 0, 0], Expected=0, Output=0, Correct=True
  Data 9: Inputs=[1, 1, 0, 0], Expected=0, Output=0, Correct=True

Akurasi: 100.00%
```

Fig. 2. Quantum Perceptron Result

Based on the results in Figure 2, it can be seen that the Quantum Perceptron has high accuracy, namely 100% with a number of epochs of 2 epochs.

```
Epoch: 1/10
Epoch: 2/10
Epoch: 3/10
Epoch: 4/10
Epoch: 5/10
Epoch: 6/10
Epoch: 7/10
Epoch: 8/10
Epoch: 9/10
Epoch: 10/10
Bentuk predictions: (9, 1)
Bentuk y: (9, 1)
Accuracy: 100.0
```

Fig. 3. Perceptron Result

Based on the results in Figure 3, it can be seen that the Perceptron has high accuracy, namely 100% with a number of epochs of 10 epochs.

TABLE III. COMPARISON RESULT

Model	Accuracy	Epoch
Perceptron	100%	10
Quantum Perceptron	100%	2

IV. CONCLUSION

The research results show that the Quantum Perceptron algorithm can predict the number of visitors to e-commerce websites in Indonesia. After carrying out the training stage on all the data tested, the result was that the quantum perceptron algorithm was able to make very good predictions, proven by perfect accuracy, namely 100% with a total of 2 epochs, while the classical perceptron had the same accuracy with a larger number of epochs, namely 10 epochs. Quantum perceptron has better performance and shorter time, this can be seen from the smaller number of epochs. In future research, it is hoped that other quantum computing-based algorithms can be explored to increase computing time more efficiently.

REFERENCES

[1] A. Perdomo-Ortiz, et al., "A quantum approach to training neural networks," *Quantum Science and Technology*, vol. 5, no. 2, pp. 145-156, 2021.

[2] Y. Cao, J. Romero, and A. Aspuru-Guzik, "Quantum machine learning and quantum algorithms," *Nature Reviews Physics*, vol. 3, pp. 12-25, 2020.

[3] H. A. Kadhim and A. Mohammed, "Quantum perceptron for pattern recognition in big data," *IEEE Access*, vol. 8, pp. 87242-87256, 2020.

[4] J. Liu, et al., "Quantum neural networks for high-dimensional datasets," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 3, pp. 1-12, 2021.

[5] R. Subramanian and A. John, "Big data analysis in e-commerce: challenges and opportunities," *Journal of Electronic Commerce Research*, vol. 22, no. 1, pp. 65-82, 2021.

[6] K. Sundar and A. Ren, "Challenges in scalable e-commerce systems," *Journal of Computing and Information Technology*, vol. 29, pp. 113-123, 2021.

[7] M. Benedetti, et al., "A review of quantum perceptron models," *IEEE Transactions on Artificial Intelligence*, vol. 12, pp. 102-112, 2022.

[8] S. K. Goyal and A. Pathak, "Quantum machine learning algorithms for e-commerce data analysis," *Applied Quantum Information*, vol. 6, pp. 122-136, 2023.

[9] P. Rebentrost, et al., "Quantum support vector machine for big data analysis," *Journal of Quantum Computing*, vol. 8, pp. 220-234, 2020.

[10] H. Neven, et al., "Quantum optimization and machine learning," *IEEE Spectrum*, vol. 59, pp. 28-35, 2022.

[11] S. Lloyd and P. Kok, "Machine learning in quantum computing," *Physics Today*, vol. 75, no. 6, pp. 32-39, 2022.

[12] J. Preskill, "Quantum computing in the NISQ era and beyond," *Quantum Information*, vol. 3, no. 8, pp. 13-25, 2021.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2334, Copyright ©2025
Submitted : April 22, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

[13] M. Schuld and F. Petruccione, "Quantum machine learning: Recent developments and future directions," *Frontiers in Physics*, vol. 9, pp. 220-232, 2023.

[14] S. K. Sharma and K. A. Kim, "Leveraging quantum computing for e-commerce prediction models," *Journal of Electronic Commerce Research*, vol. 20, pp. 332-344, 2023.

[15] T. L. Mailloux, "Predicting web traffic using quantum algorithms," *Journal of Big Data and Quantum Computing*, vol. 4, pp. 45-58, 2021.

[16] B. X. Chen and J. G. Mauk, "Quantum-enhanced predictive models for e-commerce," *Journal of Business Analytics*, vol. 17, pp. 102-120, 2022.

[17] P. Alvarez, et al., "Evaluating quantum machine learning models for e-commerce growth," *Journal of Quantum Information Systems*, vol. 9, no. 2, pp. 56-73, 2023.

[18] F. H. Smith and N. K. Verma, "Implementing quantum predictive algorithms for e-commerce," *IEEE Access*, vol. 12, pp. 15926-15935, 2024.

[19] C. F. Fox, "Quantum computing applications in business and e-commerce," *Journal of Quantum Business Models*, vol. 22, pp. 74-86, 2024.

[20] D. L. Stover, "A study on the impact of quantum computing on business processes," *Quantum Information Systems*, vol. 13, pp. 120-134, 2022.

[21] H. W. Leung and G. Novak, "Quantum data science in the future of business analytics," *Journal of Quantum Research*, vol. 6, pp. 233-248, 2022.

[22] B. Chen and S. Boyd, "Quantum optimization techniques in large-scale e-commerce systems," *Journal of Computational Business Analytics*, vol. 18, no. 2, pp. 41-56, 2023.

[23] A. Patel and M. Raghavan, "Quantum perceptron-based models for data prediction," *IEEE Transactions on Neural Networks*, vol. 33, no. 5, pp. 202-214, 2023.

[24] P. Wong and M. Yung, "Application of quantum computing in customer data prediction models," *Journal of Business Intelligence*, vol. 11, pp. 92-107, 2023.

[25] S. Aggarwal and V. Bhatia, "Quantum computing for large-scale data analysis in e-commerce," *Quantum Computing and Big Data*, vol. 7, pp. 34-46, 2024.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2334, Copyright ©2025
Submitted : April 22, 2025, Revised : May 12, 2025, Accepted : May 14, 2025, Published : May 26, 2025

The Role of Social Influence and Security Risk in Shaping Intention to Use Ride-Hailing in West Papua: A Theory of Planned Behavior Perspective

Dewi Aulia Nurhayati Maswatu[1]*, Dedi I. Inan[2], Ratna Juita[3], Muhammad Indra[4]
Department of Informatics[1], [2], [3],[4]
University of Papua
Manokwari, Papua Barat, Indonesia
202165022@student.unipa.ac.id[1], d.inan@unipa.ac.id[2], r.juita@unipa.ac.id[3], m.indra@unipa.ac.id[4]

***Abstract*— This study explores the adoption of ride-hailing services in West Papua, a developing region in Indonesia, where concerns about service performance and security risks influence user decisions. Guided by the Theory of Planned Behavior (TPB), the research examines how service performance, social influence, and perceived security risks affect users' behavioral intention and actual usage. A total of 158 valid responses were collected through a quantitative survey, and data were analyzed using Partial Least Squares Structural Equation Modeling (PLS-SEM). Findings reveal that service effectiveness and social influence significantly influence behavioral intention, while efficiency and certainty do not. Additionally, certainty, effectiveness, and behavioral intention strongly affect user behavior. The model demonstrates moderate explanatory power with R^2 values of 0.561 for behavioral intention and 0.600 for user behavior. These results suggest that enhancing perceived service effectiveness and leveraging social influence can encourage adoption in regions with limited digital infrastructure. The study contributes to understanding technology acceptance in underdeveloped areas and offers practical insights for ride-hailing providers aiming to improve user trust and engagement.**

Keywords*— *Service Performance, User Intention, Ride-hailing, User Behavior, Theory of Planned Behavior

I. Introduction

The rapid development of digital technology has significantly transformed various sectors, including transportation. One key innovation resulting from this transformation is ride-hailing—an application-based system that allows users to request transport services via mobile devices [1]. These platforms offer benefits such as improved accessibility, time efficiency, convenience, and personalized transportation options [2].

In Indonesia, digitalization has supported the expansion of several ride-hailing platforms, including Maxim, which began operating in 2018. Maxim provides a broad range of services—passenger transport (Maxim Bike), goods delivery, food and shopping, cleaning, and cargo transport [3]. Its competitive pricing and wide service coverage make it particularly attractive in areas with limited access to conventional transportation [4].

However, the adoption of ride-hailing in developing regions like West Papua faces challenges such as negative perceptions of service quality, inaccurate driver tracking, and security concerns [3]. These issues are compounded by infrastructural limitations and low digital literacy. Such obstacles are even more pronounced in 3T regions (frontier, outermost, and underdeveloped), where limited infrastructure, low technological readiness, and strong local norms create unique barriers to adoption [5].

To date, most previous studies have concentrated on platforms like Gojek and Grab within densely populated urban settings [1] [6]. This focus has led to a significant gap in the literature concerning the adoption of similar services in regions like Papua. The factors influencing adoption in such areas may differ substantially due to infrastructure limitations, low digital literacy, and distinct social norms [5]. Moreover, many of these studies have overlooked local contextual influences such as trust, perceived usefulness, and perceived risk that are central to users' decision-making processes. In addition, ride-hailing platforms like Maxim have received limited academic attention, despite possessing characteristics that may be more aligned with the needs of local communities. Unlike Gojek or Grab, Maxim tends to penetrate underserved areas more effectively due to its simpler service mechanisms and more affordable pricing, making it a practical option in regions where conventional transportation is scarce. Therefore, research that specifically explores the adoption of Maxim in underdeveloped regions is urgently needed to provide a more accurate understanding of digital service adoption in such contexts.

This study addresses the research gap by examining how five external factors namely service certainty, efficiency, effectivity, social influence, and security risk influence individuals' behavioral intention to use Maxim in West Papua. These factors were selected based on their relevance to the social and infrastructural conditions commonly found in developing regions, which differ substantially from those in urban areas.

The study employs Ajzen's Theory of Planned Behavior

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2333, Copyright ©2025
Submitted : April 17, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

(TPB) as its theoretical foundation. TPB has been extensively used in studies on technology adoption to explain how attitudes, subjective norms, and perceived behavioral control influence both intention and actual behavior [7]. However, its application in geographically marginalized regions remains limited, despite the fact that environmental constraints and cultural dynamics may significantly shape user behavior. By applying TPB within the context of West Papua, this research seeks to enhance theoretical insights into the adoption of digital transportation services and to support the development of policies that are responsive to the needs of underserved communities.

II. Theoretical Background

To understand individuals' decisions in adopting ride-hailing services, it is essential to examine various factors that influence their user behavior and behavioral intentions. These factors form the basis for explaining the motivations behind the use of digital transportation platforms such as Maxim.

A. Ride-hailing

Ride-hailing is an app-based transportation service that permits rapid reservations and offers economic possibilities [1]. The development of this service is supported by rapid technological advances. In developing regions, the impact of providing this service increases community mobility and creates jobs for local residents. Maxim is one example of a ride-hailing service, which comes with various features to increase user interest. With an adaptive business model, Maxim's service continues to compete in the ride-hailing market and provides a viable alternative for the community in choosing a transportation service that suits their needs.

B. Theory of Planned Behavior (TPB)

This study uses Ajzen's Theory of planned behavior (TPB) to explain and evaluate individual intentions and behavior [8]. In this study, TPB is used to evaluate the factors that influence Behavioural intention to use and Use Behaviour in the context of ride- hailing services, specifically the Maxim service. This research considers key factors including Certainty, Efficiency, Effectivity as well as Social Influence and Security Risk that play a role in shaping users' intention to adopt ride-hailing services.

C. Ride-hailing Performance

The performance of ride-hailing services is influenced by factors such as certainty, efficiency, and effectiveness. The certainty provided by the service builds user confidence, particularly regarding pick-up timeliness and transparent fare rates [9]. The more reliable a service is, the more likely users are to continue using it. Users also tend to prefer services that are easy to use and offer quick response times [10]. However, obstacles during usage may reduce user interest [11]. Therefore, meeting customer expectations through high service quality is essential to encourage continued usage [9].

D. Social Influence

Social influence shapes user decisions through direct recommendations or indirect exposure to peer behavior. Support from people in the surrounding environment such as family, friends, or positive reviews on social media can significantly motivate individuals to use ride-hailing services. According to a previous study [1], social influence is closely linked to human attitudes and emotions, as it emerges from interactions within the social environment that ultimately affect the adoption of online transportation services.

E. Security Risk

Perceived risks associated with safety and privacy when using ride-hailing services is one factor that raises user concerns. Concerns about safety and misuse of personal data during the trip can reduce user intention in the decision to use the service. A person's understanding of the security of the services they use greatly influences user intentions and behavior [12].

III. Research Methodology

A. Research Model and Hypotheses Development

This study adopts the Theory of Planned Behavior (TPB) as its foundational framework, incorporating the concept of Performance, which is defined through the dimensions of Certainty, Efficiency, and Effectiveness from the perspective of Maxim users. Additionally, external elements like Social Influence and Security Risk are recognized as determinants impacting user behavior. Moreover, Behavioural Intention to Use serves as an intermediary variable linking the independent and dependent factors, ultimately shaping Use Behaviour in the adoption of ride-hailing services in developing nations regions.

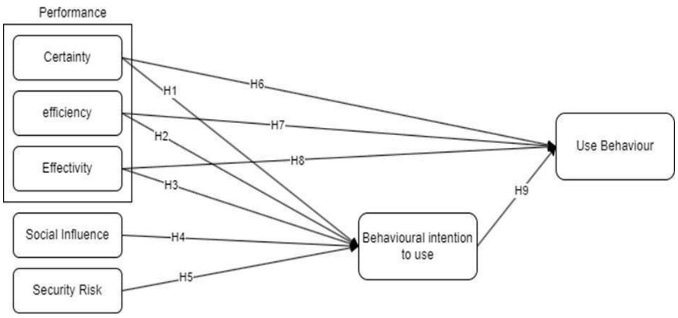


Fig. 1. Research Model

1.) The Impact of Service Certainty on Behavioural Intention to Use and Use Behaviour in Ride-Hailing Services.

The emergence of ride-hailing greatly affects community activities so that users have considerations about the certainty provided by the service. Certainty of service can create a sense of trust to use Maxim services, trust can be conceptualized as a psychological state that certainly motivates a person to accept something specifically based on favorable expectations regarding the intentions and behavior of the other party [13]. Based on this research, researchers try to prove the relationship between Certainty with Behavioural intention to use and Use Behaviour. Therefore this hypothesis was built:

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2333, Copyright ©2025
Submitted : April 17, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

H1: Certainty about the services provided has a positive influence on behavioral intention to use ride-hailing services.

H6: The level of certainty regarding the quality assurance of ride-hailing services has a significant influence on users' actual use behavior.

2.) The Impact of Efficiency on Behavioural Intention to Use and Use Behaviour in Ride-Hailing Services.

Ride-hailing services help people to book vehicles and drivers flexibly [14]. Fast pick-ups occur because the application that connects users and drivers will bring together the nearest drivers [15]. Based on the explanation of the ease of using ride-hailing which shows the extent to which a person believes that using this service will help them in their activities, this hypothesis is made:

H2: Efficiency in ride-hailing services has a major influence on Maxim's intention to use.

H7: Perceived time efficiency influences users' propensity to order ride-hailing services.

3.) The Impact of Effectivity on Behavioural Intention to Use and Use Behaviour in Ride-Hailing Services.

Ride-hailing services bring convenience to individuals or communities [16]. The convenience provided by the service to users creates a sense of trust that using technology does not require excessive effort [17]. From study, researchers tried to prove the relationship between Effectivity with Behavioural intention to use and Use Behaviour. Then the hypothesis was built:

H3: The effectiveness of ride-hailing services is significant in choosing to use these services.

H8: Services that have good effectiveness can indirectly influence user behavior in utilizing Maxim services.

4.) The Impact of Social Influence on Behavioural Intention to Use and Use Behaviour in Ride-Hailing Services.

From several factors supporting the use of ride-hailing internally, there are several external considerations that will influence a person to use Maxim services, one of which is social influence which usually comes from family, friends and people around [14]. Individual social interactions can also influence users' feelings about using Maxim services [1]. Thus the researcher hypothesizes the following:

H4: Social influence significantly influences users' intention to use Maxim and recommend it to people nearby.

5.) The Impact of Security Risk on User Intentions in Ride-Hailing Services.

The existence of concerns about the risks felt by users is one of the factors that influence a person's interest in using ride-hailing services [18], individual supporting factors in adopting a service include is security [19], the researcher proposes the following hypothesis:

H5: Perceived security risk affects users' intention to choose to use Maxim ride-hailing service.

6.) The Impact of Behavioural Intention on User Behavior in Ride-Hailing Services.

The last relationship to be examined is the effect of Behavioural intention to use on Use Behaviour later. Individual perceptions of services have an impact on user behavioral and behavior [20], the more users are willing to use online transportation services. With this the following hypothesis is made:

H9: Intention to use ride-hailing services has a positive significant effect on user behavior in using ride-hailing services.

B. Sampel

This study uses a quantitative survey to collect numerical data through questionnaires, which is then analyzed using statistical techniques to assess, interpret, and understand patterns or relationships between variables objectively [21].

The sampling targeted all Maxim service users, regardless of age, gender, or education level. Data were collected through an online questionnaire distributed to potential respondents in the developing area of Manokwari, West Papua. A non-probability sampling technique was employed, based on specific inclusion criteria primarily the participants' willingness to take part in the study. This study employs PLS-SEM, utilizing the G*Power tool for power analysis. By setting an effect size of 0.15, a 5% alpha significance level, and 94% power analysis, with seven predictor variables, the required sample size was 74 respondents [22]. However, from the collected questionnaire responses, 168 respondents participated, and after filtering for completeness, 158 responses were deemed valid for further analysis. The number of valid responses exceeded the minimum sample size determined by G*Power.

This research is conducted online and does not involve any risks that may harm the participants. Participation in this study is voluntary and anonymous. Prior to filling out the questionnaire, respondents are provided with an explanation of the research purpose and informed that the collected data will only be used for academic purposes. Respondents are given the opportunity to read and understand the explanation, and by filling out the questionnaire using a 1-5 rating scale (Strongly Disagree to Strongly Agree), respondents are considered to have given their consent voluntarily. This research was approved by the institutional ethics committee of Universitas Papua. All participant data are kept confidential and used solely for academic purposes. Table I presents the socio-demographic characteristics of the respondents.

TABLE I. DESCRIPTION OF DEMOGRAPHIC RESPONDENTS

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2333, Copyright ©2025
Submitted : April 17, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Variables	Category	Frequency	Percentage
Age	<17 Years	20	12,1
	18-30	136	82,4
	31-50	8	4,8
	50 Years	1	0,6
Gender	Male	67	40,6
	Female	98	59,4
Ever Used Maxim	Ever	121	73,3
	Never	44	26,7

C. Analysis Method

In research, the PLS-SEM method is applied to evaluate measurement and structural models to test reliability, validity, and the relationships between variables established in the hypothesis. This statistical technique is used to analyze complex associations between latent variables in structural equation modeling [23].

IV. Results And Analysis

A. Measurement Model Evaluation

Evaluation of the measurement model (outer model) is the first step in the data analysis cycle before proceeding to the evaluation of the structural model (inner model) [24]. In this study, there are criteria used to assess the outer model, namely convergent validity, discriminant validity, and composite reliability [25]. Convergent validity test can be done by checking the factor (outer loading) and average variance extracted (AVE). Where an indicator is considered to have good convergent validity if it meets the standard > 0.7 [24]. AVE is a standard value that each variable must have where the acceptable threshold for AVE is 0.5 or higher [26]. Reliability testing is carried out using two methods, namely Cronbach's alpha (CA) and composite reliability (CR), where the construct is considered reliable if the value of CA and CR is above 0.07. For LF < 0.40, indicators must be removed [27]. All constructs and measurement items are presented in Table II.

TABLE II. Confirmation Variabel Results

Building	Statement	Code	LF
Certainty (C) [13] CA, CR, AVE = 0.851, 0.855, 0.771	I believe that Maxim can be relied upon to deliver me to destination without delay	C1	0.844
	I am satisfied with the timeliness of the service provided by Maxim.	C2	0.911
	I am confident that Maxim is able to handle situations that may cause delays and still deliver services on time.	C3	0.877
Efficiency(E) [14], [28] CA, CR, AVE = 0.816, 0.837, 0.731	I had no difficulty filling in information required to place an order.	E1	0.815
	The driver Maxim arrived at my location according to the estimated arrival time..	E2	0.834
	The routes taken by Maxim drivers always match the estimated time given before the trip.	E3	0.914
Effectivity (EF) [1] CA, CR, AVE = 0.851, 0.855, 0.770	Maxim offers a wide selection of services (maxim *bike*, maxim *car*, maxim *food*) that are adequate for my needs.	EF1	0.862
	I feel comfortable and safe when using Maxim's services for my travel.	EF2	0.911
	I feel comfortable and safe when using Maxim's services for my travel.	EF3	0.858
Social Influence (SI) [1] CA, CR, AVE = 0.865, 0.865, 0.788	Those I trust recommended me to use Maxim's services.	SI1	0.886
	Those who influence my behavior thinks I should use Maxim's services.	SI2	0.907
	I use Maxim's service because some people around me also use it.	SI3	0.870
Security Risk (SR) [29] CA, CR,AVE = 0.878, 1.246, 0.880	Maxim may present hazards that could jeopardize the safety of the user.	SR2	0.897
	Maxim may put me at potential risk of physical threats during the trip.	SR3	0.977
Behavioral Intention to Use (BI) [1] CA, CR, AVE = 0.830, 0.837, 0.747	I feel the need to use Maxim's services in my daily activities.	BI1	0.878
	I plan to use Maxim's services in the near future.	BI2	0.896
	I would recommend using Maxim's services to others based on my intention to use them.	BI3	0.817
Use Behaviour (UB) [30] CA, CR, AVE = 0.851, 0.851, 0.772	I use Maxim's services regularly in my daily activities.	UB1	0.888
	I often use Maxim's services for both long and short trips.	UB2	0.915
	I feel comfortable using Maxim's services for a long time while traveling.	UB3	0.831

In Table II, the majority of indicators have LF values > 0.7 and AVE > 0.5. However, in order for the test to continue, the SR1 indicator was eliminated because it had a non- ideal LF value of 0.434. In addition, two other indicators, E2 and UB1 were also eliminated because they could interfere with the discriminant validity test, but did not affect the validity of the previous test.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2333, Copyright ©2025
Submitted : April 17, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Furthermore, this study conducted a discriminant validity test, to ensure that a variable is different from other similar variables using the Heterotrait-monotrait ratio (HTMT) - Matrix test [31]. In determining the limit value of HTMT, there are two general standards, namely a maximum of 0.85 and 0.9 [14]. Table III shows that all variables have an HTMT value < 0.85. So that the discriminant validity test has been fulfilled through the HTMT test [32].

TABLE III. HETEROTRAIT-MONOTRAIT RATIO

	BI	C	E	EF	SI	SR	UB
BI							
C	0.720						
E	0.728	0.879					
EF	0.779	0.835	0.888				
SI	0.808	0.812	0.708	0.743			
SR	0.096	0.086	0.057	0.090	0.087		
UB	0.858	0.668	0.576	0.544	0.643	0.071	

B. Structural Model Evaluation

This structural model evaluation aims to analyze the relationships among variables in the research model, including multicollinearity testing using the Variance Inflation Factor (VIF) values and hypothesis testing through T-statistics and P-values [33]. The multicollinearity test is conducted to assess whether there is a high correlation between independent variables in the model. According to [24], VIF values are considered acceptable if they fall between > 0.2 and < 5. Based on Table IV, all VIF values meet these criteria, indicating no multicollinearity issues, and therefore, the regression analysis can proceed with greater accuracy.

TABLE IV. MULTICOLLINEARITY TEST RESULT

	BI	C	E	EF	SI	SR	UB
BI							1.896
C	2.955						2.593
E	2.571						2.603
EF	2.687						2.804
SI	2.134						
SR	1.013						
UB							

Based on the VIF results in Table IV, which indicate that there are no multicollinearity issues among the variables, hypothesis testing can be conducted to evaluate the direct effects between variables [34], [35]. A hypothesis is accepted if the T-statistic value is greater than 1.96 and the P-value is less than 0.05 [36], [17]. According to Table V, out of 9 proposed hypotheses, 5 were accepted and 4 were rejected.

TABLE V. HYPOTHESIS TEST RESULT

Hypothesis	Variables	T-statistics	P values	Description
H1	C → BI	0.256	0.798	Rejected
H2	E → BI	1.565	0.118	Rejected
H3	EF → BI	3.127	0.002	Accepted
H4	SI → BI	4.560	0.000	Accepted
H5	SR → BI	1.395	0.163	Rejected
H6	C → UB	3.297	0.001	Accepted
H7	E → UB	0.064	0.949	Rejected
H8	EF → UB	2.120	0.034	Accepted
H9	BI → UB	10.801	0.000	Accepted

The coefficient of determination (R-Square) is used to see the extent to which the independent variable is able to explain the variation in the dependent variable. If the R- Square value is 0.75, it can be said to be strong, a value of 0.50 is said to be moderate and a value of 0.25 is said to be weak [26]. The higher the R-Square value, the better the model in explaining the effect of the independent variable on the dependent variable.

TABLE VI. R-SQUARE TEST RESULT

	R-square	Description
BI	0.561	Moderate
UB	0.600	Moderate

Based on Table VI above, the value of R^2 owned by Behavioral Intention to Use (BI) is 0.561, which indicates that the level of prediction accuracy in the model is moderate. This also applies to Use Behavior (UB), where the value at R^2 is 0.600, which indicates a moderate level of prediction accuracy.

V. DISCUSSION AND CONCLUSION

The analysis revealed that Certainty did not have a significant effect on Behavioral Intention (H1), as indicated by T = 0.256 and P = 0.798. This suggests that users' sense of assurance about the service does not automatically translate into a clear intention to use it. This finding contradicts the results of , which found a significant positive relationship between the two variables. It indicates that even when users feel confident about the service, this confidence alone may not be sufficient to drive their intention to adopt it. On the other hand, Certainty was found to significantly influence Use Behavior (H6), with T = 3.297 and P = 0.001. This implies that a higher level of perceived certainty about the service may encourage actual use, even if the initial behavioral intention is weak. This finding contradicts the results of [36] and indicates that, in this context, users may choose to use the service based on their perceived reliability, even in the absence of strong prior intentions.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2333, Copyright ©2025
Submitted : April 17, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Hypotheses H2 and H7, which tested the influence of Efficiency on Behavioral Intention and Use Behavior, were both rejected. Efficiency did not significantly affect Behavioral Intention (T = 1.565; P = 0.118) or Use Behavior (T = 0.064; P = 0.949). These results are inconsistent with [14], which emphasized the importance of efficiency. A plausible explanation is that users in underdeveloped regions (3T areas) tend to prioritize service availability and affordability over technical efficiency. Thus, efficiency may not be considered a crucial factor when deciding whether to use the service.

On the other hand, Effectiveness was observed to have a substantial influence on both Behavioral Intention and Use Behavior. Hypotheses H3 (T = 3.127; P = 0.002) and H8 (T = 2.120; P = 0.034) were validated, which is consistent with the results presented in [36] and [37].These results suggest that users' perceptions of the service's effectiveness such as timely response and ease of use play a vital role in encouraging both their intention and actual usage. When the service is perceived as delivering what users expect, it becomes more appealing for repeated use.

Social Influence was also found to significantly affect Behavioral Intention (H4), with T = 4.560 and P = 0.000. This indicates that users' decisions are highly shaped by recommendations and opinions from their social circles. This result supports [1], which emphasized the strong role of social context in forming intention. In collectivist or community-oriented settings such as Papua, group norms and shared perceptions often outweigh individual reasoning.

Conversely, Security Risk was not found to have a significant effect on Behavioral Intention (H5), as shown by T = 1.395 and P = 0.163. This contrasts with [25], which reported a significant negative relationship. One possible explanation is that users in this region may exhibit a tolerance for perceived risks due to limited alternatives, or they may have normalized low expectations regarding safety. Additionally, users may give greater weight to cost considerations than to potential security concerns.

Finally, Behavioral Intention had a strong and significant effect on Use Behavior (H9), with T = 10.801 and P = 0.000. This confirms that intention is a key predictor of actual usage, aligning with [36]. It also indicates that even in remote or underserved areas (3T), behavioral intention remains a central factor in shaping actual service adoption. This highlights the importance of understanding the factors that influence users' intentions in order to encourage continued use.

In summary, the results indicate that Effectiveness and Social Influence significantly affect users' Behavioral Intention to use Maxim services. Furthermore, Certainty, Effectiveness, and Behavioral Intention are key factors influencing Use Behavior. On the other hand, Certainty, Efficiency, and Security Risk were not found to be significant predictors of Behavioral Intention, and Efficiency also did not notably impact Use Behavior. These findings imply that, within the context of West Papua, the perceived usefulness of the service and social factors play a more crucial role in adopting the service compared to operational efficiency or concerns over safety.

The R-square analysis indicates that the model falls within the moderate prediction category. The Behavioral Intention to Use (BI) had an R^2 value of 0.561, meaning that the tested variables explained 56.1% of the variance in BI, while the remaining 43.9% is attributed to other unexamined factors. Similarly, Use Behavior (UB) had an R^2 value of 0.600, suggesting that 60% of the variance in UB can be explained by the model, particularly through BI, while the remaining 40% is likely influenced by external factors not addressed in this study.

In summary, Certainty, Efficiency, Effectivity, Social Influence, and Security Risk collectively account for 56.1% of the variation in Behavioral Intention, which in turn explains 60% of the variation in Use Behavior. The remaining variance may be shaped by other relevant variables not included in the current research framework.

A. Theoretical Implications

This study explores the influence of internal factors by adopting the Theory of Planned Behavior (TPB) to analyze users' intentions and users behavior toward the adoption of the Maxim ride-hailing service in a developing region affected by the 3T (Frontier, Outermost, and Disadvantaged areas) context, specifically Manokwari, West Papua. The research aims to understand how psychological and social aspects influence users' decisions in utilizing app-based transportation services in areas where infrastructure and technology penetration are still developing.

The findings indicate that perceived service effectiveness and social influence are the key factors driving the adoption of Maxim's ride-hailing service. Perceived service effectiveness relates to reliability, efficiency, and overall service performance, while social influence, including support from friends, family, and the community, enhances users' trust in the platform. Both factors play a crucial role in shaping user confidence, particularly in regions where access to digital transportation solutions is still expanding.

On the other hand, security risk does not significantly impact users' intentions. This finding suggests that users in developing regions prioritize service convenience and efficiency over potential security concerns. Factors such as easy access, affordability, fast service, and vehicle availability play a more significant role in user decision-making. Additionally, users may have a certain level of risk tolerance or perceive that Maxim's security measures are adequate, making security concerns a less critical factor in the adoption of ride-hailing services.

B. Limitations and Future Research Directions

This research provides significant practical implications for understanding the adoption of Maxim ride-hailing services. However, the results of this study still show some limitations that need further attention. This study focused on the use of Maxim services in the developing area of Manokwari, West

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2333, Copyright ©2025
Submitted : April 17, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Papua therefore, evaluation with a wider and diverse respondents will be the main focus of further research in the future because the needs of maxim services in other developing areas may be different from maxim services Manokwari. In addition, this study still shows that there are many factors that might contribute especially to the individual's intention to use the service and their attitudes later, this allows the need for consideration of other variables that are more effective in providing an understanding of the use of Maxim ride-hailing services.

Future research could focus on the Maxim driver experience, including aspects of welfare, protection, and income. Understanding drivers' experiences can reveal issues and other unmet needs, which can help with service development. Driver and user perspectives provide a more thorough understanding and help identify gaps and improve service quality. In addition, it can encourage new innovations in services and build a more balanced ecosystem for all parties.

REFERENCES

[1] D. I. Inan *et al.*, "Technology anxiety and social influence towards intention to use of ride-hailing service in Indonesia," *Case Stud. Transp. Policy*, vol. 10, no. 3, pp. 1591–1601, 2022, doi: 10.1016/j.cstp.2022.05.017.

[2] B. Bustami and R. Laksamana, "Transformasi Transportasi Tradisional (Offline) ke Transportasi Online Sebagai Solusi Bagi Pengguna di Kota Pontianak," *J. Ekon. Bisnis dan Kewirausahaan*, vol. 8, no. 3, p. 194, 2019, doi: 10.26418/jebik.v8i3.29404.

[3] N. Chasanah, D. Ferriswara, and L. Listyawati, "Analysis of the Quality of Maxim and Indrive Online Motorcycle Taxi Services for Students of Dr . Soetomo University Surabaya," *Int. J. Multicult. Multireligious Underst.*, vol. 11, no. 1, pp. 270–279, 2024, doi: http://dx.doi.org/10.18415/ijmmu.v11i1.5285.

[4] B. D. Mardjani, S. L. H. V. J. Lapian, and M. Mangantar, "Pengaruh Harga, Promosi Dan Kualitas Layanan Terhadap Kepuasan Konsumen Pada Transportasi Online (Studi Kasus Gojek Dan Maxim Di Kota Manado)," *J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 11, no. 1, pp. 942–952, 2023, doi: 10.35794/emba.v11i1.46658.

[5] G. Valentino Aji, U. Pembangunan Jaya Alamat, J. Cendrawasih Raya Bintaro Jaya, S. Baru, K. Ciputat, and K. Tangerang Selatan, "Pengaruh Kualitas Pelayanan terhadap Minat Beli Ulang pada Transportasi Online Maxim Universitas Pembangunan Jaya Handito Lava Daenova," *J. Sains Student Res.*, vol. 1, no. 2, pp. 370–379, 2023, [Online]. Available: https://doi.org/10.61722/jssr.v1i2.162

[6] A. V. Utami and D. S. S. T, "Analisis Faktor Pendorong Niat Menggunakan Grab Factor Analysis Affecting Behavioral Intention To Use Grab," *eProceedings Manag.*, vol. 7, no. 2, pp. 3396–3405, 2020, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/management/article/view/13444%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/management/article/download/13444/12987

[7] J. F. Tunner, E. Day, and M. R. Crask, "Protection motivation theory," *J. Bus. Res.*, vol. 19, no. 4, pp. 267–276, 1989, doi: 10.1016/0148-2963(89)90008-8.

[8] M. S. S. Siqueira, P. O. Nascimento, and A. P. Freire, "Reporting Behaviour of People with Disabilities in relation to the Lack of Accessibility on Government Websites: Analysis in the light of the Theory of Planned Behaviour," *Disabil. CBR Incl. Dev.*, vol. 33, no. 1, pp. 52–68, 2022, doi: 10.47985/dcidj.475.

[9] D. Nugraha, G. Putra, and S. T. Raharjo, "Analisis Pengaruh Kemudahan Penggunaan, Kualitas Layanan, Dan Persepsi Manfaat Terhadap Loyalitas Pengguna Dengan Kepuasan Pengguna Sebagai Variabel Intervening (Studi pada Pengguna Aplikasi Grab di Kota Semarang)," *Diponegoro J. Manag.*, vol. 10, no. 6, pp. 1–15, 2021, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/dbr

[10] H. Anjani, D. I. Inan, R. Juita, M. Sanglise, I. Engineering, and U. Papua, "Towards spatial information system adoption using extended tam and is success model," vol. X, no. 4, 2024.

[11] S. R. Rodiah and I. S. Melati, "Pengaruh Kemudahan Penggunaan, Kemanfaatan, Risiko, dan Kepercayaan terhadap Minat Menggunakan E-wallet pada Generasi Milenial Kota Semarang," *J. Econ. Educ. Entrep.*, vol. 1, no. 2, p. 66, 2020, doi: 10.31331/jeee.v1i2.1293.

[12] D. Irawan and M. W. Affan, "Pengaruh Privasi dan Keamanan Terhadap Niat Menggunakan Payment Fintech," *J. Kaji. Akunt.*, vol. 4, no. 1, p. 52, 2020, doi: 10.33603/jka.v4i1.3322.

[13] I. Lestari and R. S. Hamid, "Analisis Tingkat Kepercayaan Dan Kepuasan Pelanggan Terhadap Niat Untuk Menggunakan Kembali Layanan Transportasi Online Di Era Pandemi Covid-19," *Equilib. J. Ilm. Ekon. Manaj. dan Akunt.*, vol. 9, no. 1, pp. 27–35, 2020, doi: 10.35906/je001.v9i1.482.

[14] M. Febriani, A. Rachmadi, and A. D. Herlambang, "Pengaruh Golongan Usia Pengguna dari Layanan Ride-Hailing terhadap Faktor-Faktor Penggunaan Teknologi berdasarkan Unified Theory of Acceptance and Use of Technology," vol. 1, no. 1, 2023.

[15] A. Amaranggana, "Pengaruh Antara Persepsi Kualitas Pelayanan dengan Loyalitas Konsumen Profesi Karyawan pada Penggunaan Jasa Ojek Online," 2024, [Online]. Available: https://repository.unair.ac.id/130904/1/ARTIKEL JURNAL_112011133101_AMANTA AMARANGGANA.pdf

[16] N. Bagus Sadita, "Analisis Efektivitas Iklan Transportasi Online Grab Di Kabupaten Jember," pp. 1–7, 2019, [Online]. Available: http://repository.unmuhjember.ac.id/6428/

[17] I. Samberi, D. I. Inan, R. N. Wurarah, R. Juita, and M. Sanglise, "Analysis of the Adoption Level of Qris in West Papua: The Roles of Self-Efficacy, Personal Innovativeness, and Privacy Concern," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, p. 346, 2024, doi: 10.35889/jutisi.v13i1.1823.

[18] A. D. Oktavia, D. I. Inan, R. N. Wurarah, and O. A. Fenetiruma, "Analisis Faktor-faktor Penentu Adopsi E-Wallet di Papua Barat: Extended UTAUT 2 dan Perceived Risk," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 587–600, 2024, doi: 10.57152/malcom.v4i2.1277.

[19] C. F. Risdiyanto, D. I. Inan, R. N. Wurarah, and O. A. Fenetiruma, "Analisis Faktor-faktor Pendukung dan Penghambat Beralih Mengadopsi Mobile Banking di Papua Barat Memanfaatkan PLS-SEM dan Perspektif Status Quo Bias," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 637–646, 2024, doi: 10.57152/malcom.v4i2.1289.

[20] D. Supriadi, A. H. Iman, and Y. Saputra, "Studi Intensi Pembelian Kendaraan Listrik pada Generasi-Z: Pendekatan Teori Perilaku Terencana yang diperluas," *EKOMABIS J. Ekon. Manaj. Bisnis*, vol. 5, no. 01, pp. 83–98, 2024, doi: 10.37366/ekomabis.v5i01.1444.

[21] N. Wanma, D. I. Inan, and L. Y. Baisa, "Evaluasi User Experience Dan User Interface Aplikasi Laporkitong Dengan End User Computing Satisfaction," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, p. 304, 2024, doi: 10.35889/jutisi.v13i1.1789.

[22] N. Kock and P. Hadaya, "Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods," *Inf. Syst. J.*, vol. 28, no. 1, pp. 227–261, 2018, doi: 10.1111/isj.12131.

[23] J. F. Hair, M. Sarstedt, L. Hopkins, and V. G. Kuppelwieser, "Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research," *Eur. Bus. Rev.*, vol. 26, no. 2, pp. 106–121, 2014, doi: 10.1108/EBR-10-2013-0128.

[24] D. I. Inan *et al.*, "Because Follower Experience Matters: The Continuance Intention to Follow Recommendation of the Influencer," *Hum. Behav. Emerg. Technol.*, vol. 2022, no. [1] D. I. Inan et al., "Because Follower Experience Matters: The Continuance Intention to Follow Recommendation of the Influencer," Hum. Behav. Emerg. Technol., vols. 2022, 2022, doi: 10.1155/2022/3684192., 2022, doi: 10.1155/2022/3684192.

[25] Y. N. Hajiku Faradila, Niswatin, "Pengaruh Marketplace Maxim Terhadap Tingkat Penjualan Pada UMKM di Kota Gorontalo Abstrak," vol. 11, no. 1, pp. 1–9, 2025.

[26] D. I. Inan *et al.*, "Service quality and self-determination theory towards continuance usage intention of mobile banking," *J. Sci. Technol. Policy Manag.*, vol. 14, no. 2, pp. 303–328, 2023, doi: 10.1108/JSTPM-01-2021-0005.

[27] E. R. Oetami and H. Sulistyo, "Penggunaan Tik Pada Proyek Perubahan Pendidikan Dan Pelatihan Kepemimpinan Untuk Meningkatkan Perilaku

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2333, Copyright ©2025
Submitted : April 17, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Inovatif: Studi Kasus Terhadap Pemerintah Kabupaten Kotawaringin Barat," *Anterior J.*, vol. 23, no. 2, pp. 153–168, 2024, doi: 10.33084/anterior.v23i2.7127.

[28] P. W. Handayani, R. A. Nurahmawati, A. A. Pinem, and F. Azzahro, "Switching Intention from Traditional to Online Groceries Using the Moderating Effect of Gender in Indonesia," *J. Food Prod. Mark.*, vol. 00, no. 00, pp. 425–439, 2020, doi: 10.1080/10454446.2020.1792023.

[29] Y. Wang, J. Gu, S. Wang, and J. Wang, "Understanding consumers' willingness to use ride-sharing services: The roles of perceived value and perceived risk," *Transp. Res. Part C Emerg. Technol.*, vol. 105, no. May, pp. 504–519, 2019, doi: 10.1016/j.trc.2019.05.044.

[30] D. Nguyen, "Understanding Perceived Enjoyment and Continuance Intention in Mobile Games," *ICFAI J. Syst.*, p. 58, 2015, [Online]. Available: http://epub.lib.aalto.fi/fi/ethesis/pdf/14000/hse_ethesis_14000.pdf

[31] J. H. Kandami, D. I. Inan, R. Juita, L. Y. Baisa, M. Sanglise, and M. Indra, "Development and Evaluation of Android-based Infrastructure Rental Application: A Design Science Research Approach," *J. Teknol. dan Manaj. Inform.*, vol. 10, no. 1, pp. 36–47, 2024, doi: 10.26905/jtmi.v10i1.13004.

[32] P. E. Setiawati, D. I. Inan, R. N. Wurahrah, R. Juita, and M. Sanglise, "How the perceived enjoyment effect m-payment adoption in west papua province: delone and mclean information systems success model," vol. 9, no. 3, pp. 1494–1505, 2024.

[33] M. Patabang, D. I. Inan, A. Matualage, and M. Indra, "The Moderation Effect of Technology Anxiety on Digital Transformation Readiness in Public Universities : An Organizational Readiness Approach," vol. 10, no. 2, pp. 59–70, 2024.

[34] W. Reza, F. Jabnabillah, and A. S. Anggraeni, "Structural Equation Modeling Pada Adopsi Sistem Informasi Akuntansi Berbasis E-Commerce Menggunakan UTAUT2," *Value J. Manaj. dan Akunt.*, vol. 18, no. 1, pp. 199–214, 2023, doi: 10.32534/jv.v18i1.3886.

[35] P. Tam *et al.*, "1) , 2) , 3) 1," vol. 08, no. 01, pp. 1–9, 2024.

[36] D. Wardani, "Niat Pengguna Fintech Sistem Pembayaran pada Kalangan Milenium Masa Pandemi Covid-19," *J. Sist. Inf. Bisnis*, vol. 3, no. 2, pp. 40–53, 2022.

[37] L. Moreno, B. Requero, D. Santos, B. Paredes, P. Briñol, and R. E. Petty, "Attitudes and attitude certainty guiding pro-social behaviour as a function of perceived elaboration," *Eur. J. Soc. Psychol.*, vol. 51, no. 6, pp. 990–1006, 2021, doi: 10.1002/ejsp.2798.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2333, Copyright ©2025
Submitted : April 17, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

A Comparative Study of EmberGen and Blender in Fire Explosion Simulations

Arya Luthfi Mahadika [1]*, Ema Utami [2]
Department of Master Informatics [1], [2]
University of AMIKOM Yogyakarta
Yogyakarta, Indonesia
arya.dika@students.amikom.ac.id[1], ema.u@amikom.ac.id[2]

***Abstract*— The advancement of visual effects (VFX) technology has intensified the need for efficient fire explosion simulations across film, gaming, and real-time applications. This study investigates and compares the performance of two prominent simulation tools—EmberGen and Blender—by focusing on processing time efficiency and simulation quality. The research specifically evaluates five critical simulation aspects: fire particle generation, smoke behavior, turbulence effects, light dispersion, and final rendering (finishing). A total of five professional VFX artists conducted five separate tests using each software, generating a comprehensive dataset for analysis. Results show that EmberGen achieves a 29.91% overall improvement in simulation speed compared to Blender, with significant gains in fire particle generation (38.5%), smoke simulation (42.3%), turbulence effects (15.7%), light dispersion (8.9%), and finishing (11.6%). These findings indicate that EmberGen is highly effective for real-time or rapid-turnaround projects, while Blender remains advantageous for detailed, high-fidelity simulations in cinematic contexts. The study concludes that software selection should be driven by project-specific demands, where EmberGen supports time-sensitive production workflows and Blender offers greater artistic control. This research underscores the critical need for aligning simulation tools with both creative goals and production efficiency, contributing to decision-making in VFX, animation pipelines, and educational training environments within the information systems and digital content domains.**

***Keywords*— EmberGen, Blender, Fire Explosion Simulation, Visual Effects**

I. INTRODUCTION

Visual effects (VFX) have become a critical component in modern filmmaking, enabling the creation of scenes that defy physical and natural constraints [1], [2]. From their origins in the pioneering techniques of Georges Méliès such as stop motion and matte painting VFX have evolved into sophisticated digital tools that enhance storytelling, audience immersion, and production efficiency [3]. The integration of VFX not only improves the visual quality of a film but also plays a strategic role in reducing risk and cost during the filming of dangerous or complex scenes, such as explosions or natural disasters [1], [4]. As a result, VFX are now widely used in both blockbuster productions and independent short films, serving as key visual and commercial assets.

With the growing reliance on VFX, professionals increasingly turn to digital tools capable of delivering high-quality simulations under tight production deadlines. Two prominent software platforms used for simulating fire explosions are EmberGen and Blender. EmberGen is known for its real-time GPU-based simulation capabilities, offering immediate visual feedback that allows artists to iterate quickly and meet demanding schedules [5]. EmberGen, a VFX program developed by JangaFX, enables real-time simulation and playback of fire, smoke, explosions, and other effects [6]. In contrast, Blender is a powerful, open-source 3D creation suite with extensive functionality, including a fluid simulation engine capable of producing highly detailed and customizable fire and smoke effects [7], [8]. However, Blender's complexity and longer rendering times may pose challenges for time-sensitive projects focused specifically on explosion simulations [8].

Despite the popularity of both tools, there is limited empirical research comparing their performance in fire explosion simulation workflows, particularly in the context of computational efficiency, rendering quality, and usability. Existing studies tend to focus on the technical features of each tool individually, without offering a structured, side-by-side performance analysis that can guide practical software selection in professional VFX environments. This lack of comparative analysis presents a research gap [8], [9], [10], [11], especially for practitioners and educators seeking optimal tools for animation, visual storytelling, or interactive media production.

Several previous studies have utilized for fire simulation through a combination of particle systems and mesh emitters, where a sphere serves as the particle emitter to generate orange spheres that represent flames [7], [12], this process involves animating the size of the spheres over time to mimic the natural flicker and behavior of fire, while a lattice is employed to shape the emitted particles and enhance the visual complexity of the flames. Additionally, shaders are applied to create realistic coloration and lighting effects, and the Node Editor is leveraged to integrate advanced techniques such as Vector Blur, which simulates motion blur, and Ramp Nodes, which adjust color gradients to enhance the fire's luminosity, resulting in a visually compelling and dynamic fire simulation [7], [13]. In previous research, there has never been an experiment using EmberGen software; therefore, the author aims to conduct a study using this new method with EmberGen software.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2335, Copyright ©2025
Submitted : April 23, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

This study aims to address that gap by evaluating and comparing the performance of EmberGen and Blender in generating realistic fire explosion simulations in "I Draw It" short movie. The comparison is based on three primary dimensions: rendering quality, ease of use, and computational efficiency. By examining these aspects, the study contributes practical insights for VFX professionals, educators, and digital content creators in selecting software that aligns with both creative goals and production demands [2], [14]. The research also offers a basis for future investigations into hybrid workflows that combine the strengths of both tools for optimal results in various production contexts.

II. RESEARCH METHOD

Based on the research title, the author created a research flow for the stages or pipeline of the Performance Comparison of EmberGen and Blender in Fire Explosions, as shown in Figure 1 below.

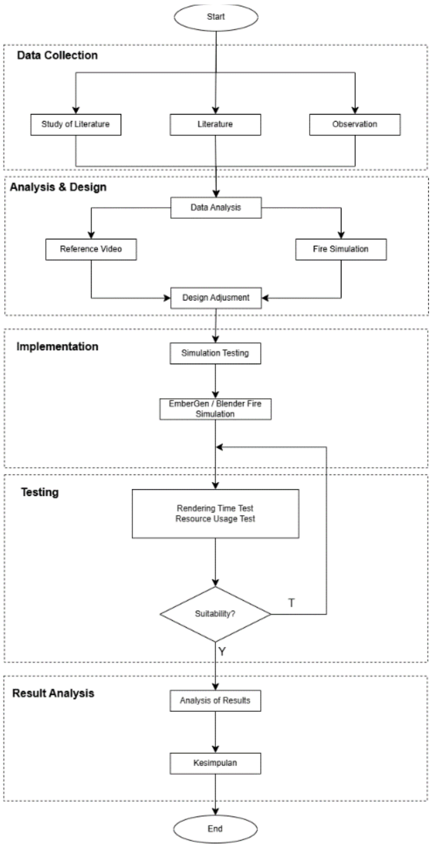


Fig. 1. Reseearch Flow of Performance Comparison Between EmberGen and Blender in Fire Explosions

This research focuses on comparing the performance of EmberGen and Blender in generating fire explosion simulations. The initial stage involves collecting literature and reference videos related to fire explosions. The literature review aims to understand the fundamental principles of fire simulation, rendering techniques, and the features offered by EmberGen and Blender. Reference videos are gathered as visual benchmarks for analyzing the realism and quality of simulations produced by both software.

The next stage is Analysis & Design, where the simulation process is planned and adapted for both software. EmberGen and Blender are tested using specific methods such as frame sampling to control simulation intervals, blending to combine visual elements, and fire particle motion analysis. Adjustments are made iteratively, with each software's simulation results compared to the reference videos to evaluate accuracy and realism.

In the Implementation and Testing phase, the designed simulations are executed in EmberGen and Blender to assess their performance. The testing process includes quantitative analysis, such as rendering time and resource efficiency, as well as qualitative evaluation of the fire explosion visuals. The simulation results are then assessed based on parameters like realism, efficiency, and ease of use. The study concludes with an in-depth analysis of the testing results, which are used to determine which software performs better in generating fire explosion simulations.

TABLE IV. FIRE EXPLOSION SIMULATION

Reference	*Fire Visual Effect*
	

This study observes the simulation workflow to collect data on the performance and output of two simulation methods, EmberGen and Blender, in a fire explosion scenario. The collected data is analyzed to evaluate the effectiveness of each method, which serves as a reference for dynamic simulations in computer graphics visual effects. Although fire explosion simulations may seem simple, in-depth analysis reveals significant fluctuations and complexity. These simulations include key elements of realistic animation, such as physical accuracy, visual fidelity, and overall rendering performance [15].

EmberGen and Blender have unique approaches to creating fire explosion simulations. In the simulation process, EmberGen employs a voxel-based approach, enabling fast calculations with high accuracy. Its parameter control system allows users to easily modify materials, fluid dynamics, and lighting effects in fire and explosion simulations. This study compares EmberGen's simulation performance with Blender, focusing on visual resolution, material adjustments, and rendering time efficiency [6].

Through this comparison, key aspects such as rendering time, user control flexibility, and the aesthetic appeal of the simulation results are measured to provide insights applicable to film production and real-time applications. The results show that EmberGen excels in rendering speed and ease of use, while Blender offers greater flexibility in adjusting simulation details and producing varied visual outcomes according to project needs [7], [15].

The collected data is analyzed in detail to calculate time and determine optimal animation parameters for each fire explosion

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2335, Copyright ©2025
Submitted : April 23, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

simulation. This analysis includes the assessment of creation duration, motion details, and other technical aspects to ensure the simulation results achieve the desired realism. After the analysis is completed, the author proceeds with the design validation process as an essential step to ensure the results align with the research objectives and standards. This validation is conducted in collaboration with three visual effects experts who have experience and expertise in creating high-quality animation simulations. Their presence aims to identify and minimize potential errors before the final implementation, ensuring more accurate and effective simulation results.

During the testing phase, fire explosion simulations are performed using two main software programs, EmberGen and Blender. Each software is tested by three professionals, with each performing five simulations per software. These repeated tests aim to collect consistent and comprehensive data, allowing for an in-depth analysis of the effectiveness and efficiency of both methods. Throughout this process, animation supervisors actively evaluate each simulation result, providing critical feedback and necessary improvements to enhance both the visual and technical quality of each animator's work. With this approach, the study not only produces quantitative data but also establishes a high-quality standard for each tested simulation.

III. Result And Discussion

The performance comparison process between EmberGen and Blender in fire explosion simulation begins with analyzing the visual characteristics of fire explosions through reference videos. These videos are observed from various angles to understand the natural movement of fire, including spread patterns, intensity, and dissipation. Simulations are conducted using two different software programs, EmberGen and Blender, to assess their capabilities in rendering fire explosions. The simulation setup involves modifying various parameters, such as particle density, turbulence, and temperature variations, to match the reference footage.

The next step is to apply fire simulation techniques in EmberGen and Blender. In EmberGen, fire simulation is performed procedurally in real-time using voxel-based rendering, allowing for instant adjustments and immediate feedback. On the other hand, Blender utilizes the Mantaflow solver for fluid simulation, which requires preprocessing before generating the final render. This difference impacts workflow efficiency in fire effect animation production. Adjustments are made in both software to optimize realism and responsiveness, ensuring that the simulated fire movement accurately reflects natural flame behavior.

Finally, performance evaluation is conducted by comparing rendering time, resource usage, and visual realism between EmberGen and Blender. Rendering speed is tested by analyzing frames per second (FPS) and computational load, while resource usage is measured based on CPU consumption during the simulation. Visual realism is evaluated by comparing the generated fire effects with the reference footage, assessing aspects such as flame flickering, smoke behavior, and heat distortion. The results of this comparison will determine which software is more efficient and effective for fire explosion simulation in animation and visual effects production.

Fig. 2. Fire Explosion Simulatioon in EmberGen

After each fire explosion simulation is completed, the results are presented in tables and graphs comparing the performance of EmberGen and Blender. The evaluation is based on rendering time, resource efficiency, and the visual realism produced. In this test, both software programs are given the same initial parameters, such as explosion size, light intensity, and smoke density, to ensure a fair comparison. The success of the simulation depends on each software's ability to generate realistic fire effects efficiently.

The simulation testing process is conducted in five main stages: initial explosion simulation, flame development, smoke movement, turbulence effects, and the final stage of fire dispersion before fading [16], [17]. In EmberGen, simulations run in real-time, allowing animators to instantly see parameter adjustments without requiring a baking process. Meanwhile, Blender, with the Mantaflow solver, requires additional processing time to calculate fluid dynamics before achieving the final result [8], [12]. This difference creates a distinct workflow experience, where EmberGen offers higher responsiveness, while Blender provides more detailed control during the refinement stage.

As part of a comprehensive performance evaluation, each simulation method is rigorously tested by measuring the time required to complete various stages of the fire explosion simulation. This assessment aims to provide a detailed understanding of the efficiency and practicality of each software tool involved. The results of these tests are systematically organized into a table that outlines the duration of each simulation phase in minutes, offering a clear comparative view of performance across the different tools [7].

By closely examining this data, it becomes possible to draw meaningful conclusions about the overall effectiveness of each software solution. Specifically, the analysis focuses on three key aspects: rendering speed, which reflects how quickly the software can produce visual outputs; physical accuracy, which evaluates how realistically the explosion is simulated according to real-world physics; and ease of use, which considers the user interface and workflow efficiency for animation and visual effects production [18], [19]. This comprehensive evaluation not only highlights which software is faster, but also which one strikes the best balance between speed, realism, and usability crucial factors for professionals in the fields of animation and visual effects [7], [20].

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2335, Copyright ©2025
Submitted : April 23, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

TABLE II. TEST RESULT OF ONE OF THE VISUAL EFFECTS ARTISTS

Test in Time								
No	***Software***	***Simulation Aspects***	***Test 1***	***Test 2***	***Test 3***	***Test 4***	***Test 5***	***Average***
1	**Blender**	Fire Particle Generation	01.28	01.20	01.27	02.22	01.30	**01.37**
		Smoke Simulation	01.27	01.54	01.54	01.52	01.24	**01.42**
		Light Dispersion	02.00	01.47	01.26	02.32	02.31	**02.03**
		Turbulence Effects	01.31	01.53	01.52	02.07	01.42	**01.49**
		Finishing	02.21	01.56	02.03	01.54	02.24	**02.07**
2	**EmberGen**	Fire Particle Generation	01.18	01.31	01.08	00.54	01.22	**01.14**
		Smoke Simulation	00.47	01.00	01.18	01.37	00.58	**01.08**
		Light Dispersion	01.02	00.50	00.50	01.03	01.32	**01.03**
		Turbulence Effects	01.15	01.19	0103	00.42	01.01	**01.04**
		Finishing	01.32	00.41	01.03	01.23	01.09	**01.09**

After recording the time measurements for each test conducted by the Visual Effects Artist, the results are compiled into a Fire Explosion Simulation test result table, using seconds (s) as the unit of measurement. This table details the time taken for each part of the fire simulation and the average processing time for two software programs: Blender and EmberGen. The testing results for Blender and EmberGen fire simulations are presented in Table 3 and Table 4 below.

TABLE III. TESTING ON BLENDER FIRE SIMULATION

Testing on Blender Fire Simulation in seconds						
	Test 1	**Test 2**	**Test 3**	**Test 4**	**Test 5**	**Average**
VFX Artist 1	320	298	310	290	315	**306.6**
VFX Artist 2	580	460	430	510	495	**495**
VFX Artist 3	260	280	300	310	200	**270**
VFX Artist 4	400	370	340	320	315	**349**
VFX Artist 5	290	270	250	260	240	**262**
Average Fire Simulation						**336.52**

Based on the data in Table 3, the fire simulation in Blender, across five test trials, has an average processing duration of 336.52 seconds, equivalent to approximately 5 minutes and 36 seconds.

TABLE IV. TESTING ON EMBERGEN FIRE SIMULATION

Testing on EmberGen Fire Simulation in seconds						
	Test 1	**Test 2**	**Test 3**	**Test 4**	**Test 5**	**Average**
VFX Artist 1	150	140	145	135	148	**143.6**
VFX Artist 2	290	270	260	280	275	**275**
VFX Artist 3	130	120	125	140	110	**125**
VFX Artist 4	220	210	200	195	180	**201**
VFX Artist 5	180	160	150	170	155	**163**

According to Table 4, the fire simulation using EmberGen across five test trials has an average processing time of 181.52 seconds. The time testing results from five Visual Effect Artists show varying outcomes for each simulation aspect, including Main Fire, Smoke Dispersion, Light Dispersion, Turbulence Effects, and Finishing. Consequently, the comparative testing between Blender and EmberGen is presented in Table 5 below.

TABLE V. COMPARISON TESTING BLENDER AND EMBERGEN

No		*Blender*	*EmberGen*
1	**VFX Artist 1**	306.6	143.6
2	**VFX Artist 2**	495	275
3	**VFX Artist 3**	270	125
4	**VFX Artist 4**	349	201
5	**VFX Artist 5**	262	163
Averages		**336.52**	**181.52**

Based on the results in Table 5, the fire simulation processing time shows a significant difference between the two software programs. EmberGen demonstrates nearly twice the efficiency of Blender in simulating fire explosions.

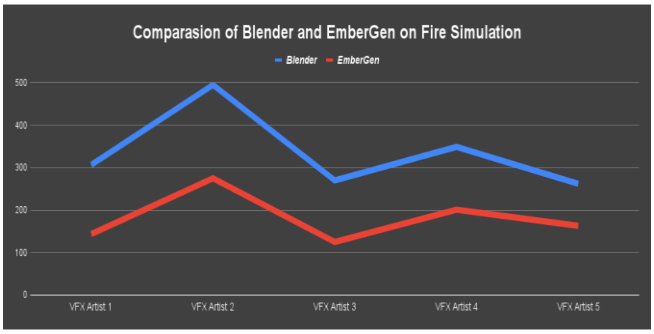


Fig. 3. Graphics Comparison of blender nnd EmberGen on Fire Simulation

Based on the result in Figure 3 above, the fire explosion simulation demonstrates that EmberGen has a higher processing speed compared to Blender. This is reflected in both the table and graph, where the average processing time for EmberGen is 181.52 seconds, whereas Blender requires 336.52 seconds. The difference in processing time between the fire explosion simulation methods using EmberGen and Blender is 155 seconds. This value is calculated based on the average time taken by each method to perform the fire explosion simulation and is considered an indicator of each software’s efficiency. Furthermore, calculations were performed to determine the percentage of simulation time efficiency, with the result explained in equation 1 below:

$$Persentase = \frac{Difference\ in\ processing\ time}{\Sigma\ Total\ processing\ time} \times 100\% \qquad (1)$$

$$= \frac{155}{518.04} \times 100\% \quad = 29.91\%$$

According to equation 1, to obtain the percentage of the fire explosion simulation time rate using four idealized rules, first, find the difference in processing time (between Blender and EmberGen), second, find the total processing time, third, divide the difference in processing time by the total processing time, fourth, multiply the result of that division by 100.

Based on processing efficiency calculations, there are four key steps in comparing the performance of EmberGen and Blender in fire explosion simulations.

The first step is determining the difference in processing time between the two methods. The time difference between them is 155 seconds. The second step is calculating the total processing time by summing the times of both methods, resulting in 518.04 seconds. Next, the third step involves dividing the time difference by the total processing time and multiplying the result by 100 to obtain the efficiency percentage. From this calculation, it is determined that EmberGen is approximately 46.1% faster than Blender in terms of processing efficiency.

According to earlier studies, using Blender software to generate a fire simulation takes longer, around 16 minutes [12]. The test results indicate that EmberGen excels in both speed and ease of use, making it more suitable for projects requiring fast results, such as fire effects in games or real-time animations. However, Blender retains an advantage in its flexibility, particularly for simulations involving complex interactions, such as explosions that affect surrounding objects. The supervisor in this study observed that while EmberGen can generate realistic simulations in a shorter time, Blender is more reliable for detailed physics settings and more complex effect customizations. Nevertheless, this investigation is restricted to visual impressions and time-based performance in a controlled environment that includes five artists.

In visual effects production, choosing the right software should align with the project's specific requirements. If speed and efficiency are the main concerns, EmberGen is the preferable option. However, for projects demanding greater detail and manual control over simulations, Blender is the more suitable choice. In some scenarios, integrating both software can offer an optimal solution using EmberGen for quick simulations and Blender for refining and fine-tuning explosion effects. Ultimately, while EmberGen has demonstrated superior speed compared to Blender, the ideal choice depends on the unique demands of the visual effects production.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2335, Copyright ©2025
Submitted : April 23, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

IV. CONCLUSION

This study reveals a notable performance edge for EmberGen over Blender in fire explosion simulations, with EmberGen achieving a 29.91% faster overall processing time. It consistently surpassed Blender in all tested areas, particularly in smoke simulation 42.3% and fire particle generation 38.5%. These findings demonstrate EmberGen's effectiveness for real-time applications and rapid prototyping, where speed is essential. Conversely, Blender excels in flexibility and parameter control, making it suitable for high-detail simulations in cinematic visual effects.

The practical takeaway from this research is a strong recommendation for a hybrid workflow: using EmberGen for time-sensitive tasks like video game effects or pre-visualization, while turning to Blender for intricate, layered simulations that require precise artistic control. This combined approach could enhance production efficiency without compromising visual quality.

However, this research is limited to time-based performance and visual impressions from a controlled environment involving five artists. Future studies should investigate other factors such as rendering quality, user experience, GPU efficiency, and cross-platform integration. Additionally, broadening the scope to include other VFX tools like Houdini or Unreal Engine's Niagara system, as well as more complex interactive simulations, could yield valuable insights into best practices for fire simulation workflows in professional settings.

REFERENCES

[1] M. S. Hwang, M. K. Park, and H. S. Lee, "A Study on the Visual Effects Production Process for Efficient Underwater Explosion CG Visualization," *J Coast Res*, vol. 116, no. sp1, pp. 518–522, 2024, doi: 10.2112/jcr-si116-105.1.

[2] V. Ong, *Artificial intelligence in digital visual effects*. dr.ntu.edu.sg, 2021. [Online]. Available: https://dr.ntu.edu.sg/handle/10356/151632

[3] M. S. Hwang and H. Lee, "Pipeline Design for Efficient Visual Effects Production," *Journal of Multimedia Information System*, 2022, [Online]. Available: https://www.jmis.org/archive/view_article?pid=jmis-9-3-219

[4] M. Choi, J. A. Wi, T. Kim, Y. Kim, and C. H. Kim, "Learning representation of secondary effects for fire-flake animation," *IEEE Access*, 2021, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9334976/

[5] JangaFX, "EmberGen Documentation," JangaFX. Accessed: Jan. 15, 2025. [Online]. Available: https://jangafx.com

[6] D. A. Akinkugbe, "Artifact," *RIT Digital Institutional Repository*, pp. 1–34, 2021.

[7] D. Senkic, "Dynamic Simulations in a 3D-Enviornment a comparison between Maya and Blender," 2010.

[8] V. Hoikkala, *Fire Simulation for a 3D Character with Particles and Motion Capture Data in Blender*, no. May. 2016.

[9] Z. H. Wu, Z. Zhou, and W. Wu, "Realistic fire simulation: A survey," *Proceedings - 12th International Conference on Computer-Aided Design and Computer Graphics, CAD/Graphics 2011*, pp. 333–340, 2011, doi: 10.1109/CAD/Graphics.2011.26.

[10] R. H. Lewis *et al.*, *Fire and smoke digital twin – A computational framework for modeling fire incident outcomes*, vol. 110, no. 1. Association for Computing Machinery, 2024. doi: 10.1016/j.compenvurbsys.2024.102093.

[11] A. A. Khan, M. A. Khan, K. A. Cashell, and A. Usmani, "An open-source software framework for the integrated simulation of structures in fire," *Fire Saf J*, vol. 140, no. June, p. 103896, 2023, doi: 10.1016/j.firesaf.2023.103896.

[12] "Primjena novih mogućnosti simulacije fluida, vatre i dima u programu BLENDER," 2025.

[13] H. Yang, C. Hu, G. Li, and J. Fan, "A Fire Escape Simulation System Based on the Dijkstra Algorithm.," *Comput. Syst. Sci. Eng.*, 2021, [Online]. Available: https://cdn.techscience.cn/ueditor/files/csse/TSP_CSSE-39-3/TSP_CSSE_16377/TSP_CSSE_16377.pdf

[14] J. Lever and T. Komura, "Real-time controllable fire using textured forces," *Visual Computer*, vol. 28, no. 6–8, pp. 691–700, 2012, doi: 10.1007/s00371-012-0684-1.

[15] P. Beaudoin, S. Paquet, and P. Poulin, "Realistic and controllable fire simulation," *Proceedings - Graphics Interface*, pp. 159–166, 2001.

[16] S. Rødal, G. Storli, and O. E. Gundersen, "Physically based simulation and visualization of fire in real-time using the GPU," *Theory and Practice of Computer Graphics 2006, TPCG 2006 - Eurographics UK Chapter Proceedings*, no. June, pp. 13–20, 2006.

[17] W. Jahn, G. Rein, and J. L. Torero, "The effect of model parameters on the simulation of fire dynamics," *Fire Safety Science*, pp. 1341–1352, 2008, doi: 10.3801/IAFSS.FSS.9-1341.

[18] D. Q. Nguyen, R. Fedkiw, and H. W. Jensen, "Physically based modeling and animation of fire," *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '02*, pp. 721–728, 2002, doi: 10.1145/566570.566643.

[19] M. Son, B. Kim, G. Wilensky, and S. Lee, "Still-frame simulation for fire effects of images," *Computer Graphics Forum*, vol. 32, no. 7, pp. 295–304, 2013, doi: 10.1111/cgf.12237.

[20] L. Xu, D. Mohaddes, and Y. Wang, "LLM Agent for Fire Dynamics Simulations," no. NeurIPS, 2024, [Online]. Available: http://arxiv.org/abs/2412.17146

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2335, Copyright ©2025
Submitted : April 23, 2025, Revised : May 10, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Selection of Recipients of Excellent Scholarship Educational Assistance using Simple Addictive Weighting Method

Rudi Hidayat[1]*, Ryan Prasetya[2], Gandung Triyono[3]
Department of Master Computer Science [1], [2], [3]
Budi Luhur University
Jakarta, Indonesia
2311600353@budiluhur.ac.id[1], 2311600270@budiluhur.ac.id[2],gandung.triyono@budiluhur.ac.id[3]

Abstract*— The selection of educational assistance recipients is an important process that determines the effectiveness of aid distribution. However, inconsistencies in assessment criteria and less systematic data management often become obstacles in determining the right recipient candidates. This problem results in subjectivity and lack of transparency in the selection process. This study proposes a solution in the form of implementing the Simple Additive Weighting (SAW) method as a multi-criteria-based decision support system. This method is used to process data on prospective recipients with criteria including economic conditions, number of family dependents, written test results, and interviews. The approach used is quantitative descriptive with stages of data collection, criteria weighting, SAW score calculation, and evaluation of results. The results of the study show that the SAW method is able to provide objective and consistent rankings of prospective recipients. Evaluation of real data on scholarship recipients shows an accuracy level of 84.62%, indicating the effectiveness of this method in the selection process. These results indicate that the SAW method can be an effective solution to increase transparency, consistency, and fairness in the educational assistance selection process.

***Keywords* — Scholarship Selection, Simple Additive Weighting, Decision Support System, Accuracy, Educational Assistance**

I. INTRODUCTION

The selection process of beneficiaries of educational assistance is a crucial aspect to ensure the effectiveness of aid distribution in amil zakat institutions that manage Zakat, Infaq, Sadaqah, and Waqf (ZISWA) funds to support the improvement of people's quality of life through better access to education [1], [2]. However, the selection process that has been running still faces challenges in the form of inconsistent criteria and unsystematic data management, which has the potential to cause the distribution of assistance that is less targeted and reduce stakeholder trust [3].

The gap encourages the need for a more structured, objective, and data-based selection system. Decision Support System (DSS) with Simple Additive Weighting (SAW) method is a relevant solution due to its ease of implementation, computational efficiency, and ability to process multicriteria assessments consistently and transparently [4].

Several previous studies have shown the advantages of SAW in the context of criteria-based selection, such as parental income, number of dependents, written tests, and interviews, which are able to produce decisions with high accuracy without excessive algorithm complexity [5][6]. Compared to other methods such as MOORA and FANP, SAW is also more stable against variations in criteria weights, so the selection results are more consistent [7].

This research aims to develop an objective and structured scholarship recipient selection process by integrating the SAW method in SPK, especially for social institutions such as DT Peduli Kuningan. In addition, the proposed system is expected to be applied in other institutions that face similar challenges in managing educational assistance.

II. METODE PENELITIAN

Previous studies have proven the important role of Decision Support Systems (SPK) in optimizing the selection process of scholarship recipients. The Simple Additive Weighting (SAW) method is one of the methods widely used in SPK because of its superiority in simplifying the calculation process and producing objective decisions. Studies by Issn, Liang, and Muhtarom show that SAW is able to provide accurate and consistent results compared to traditional methods because of its ability to combine weight values from various criteria comprehensively [8]. In addition, research by Achmad, Mu, and Saputro (2023) revealed that SAW has higher time efficiency than other methods such as Multi-Objective Optimization on the Basis of Ratio Analysis (MOORA) and Weighted Product (WP), especially in the context of handling large data on the selection of scholarship beneficiaries [5], [9].

However, most of the existing research focuses on the application of SPK in the corporate sector or government agencies with more stable management structures and better resource availability. In the context of social institutions, such as DT Peduli Kuningan, the challenges faced are different. For example, frequent changes in management staff and manual data documentation are significant obstacles in the selection process [10]. Afrina's (2018) study on the application of the FANP method in social institutions shows that although the method is effective, the long processing time and complexity of calculations are often an obstacle when faced with limited

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2360, Copyright ©2025
Submitted : May 6, 2025, Revised : May 26, 2025, Accepted : May 26, 2025, Published : May 28, 2025

human resources in small institutions [11]. In addition, research by Wicaksono, Bachri, and Irawan (2024) using MOORA in web-based SPK also showed that the method is effective for handling various criteria, but requires a more in-depth trade-off analysis than SAW, which can be a challenge in implementation in institutions with limited resources [6], [12].

Another shortcoming in the literature is the lack of research evaluating the long-term impact of PRSP implementation on beneficiary success. Most studies focus more on initial selection outcomes, without assessing how the implementation of CBMS affects the sustainability of benefits received by individuals. In addition, not many studies have explored how SDM can be customized to address the unique challenges faced by social institutions with dynamic resource conditions.

This research aims to fill the void by developing and implementing a SAW-based CBMS that is tailored to the needs and conditions of DT Peduli Kuningan. The main focus of this research is to ensure that the system is not only able to produce transparent and accurate selection decisions, but also can be operated easily by staff who do not have a strong technical background. The research will also evaluate the long-term impact of the CBMS implementation on the success of the education program, including how beneficiaries can maintain their motivation to learn and complete the program successfully.

By addressing the limitations in previous research, this study is expected to contribute not only to improving the quality of beneficiary selection at DT Peduli Kuningan but also provide practical recommendations for other social institutions facing similar issues. This research also broadens insights into how the SAW method can be adapted in the context of dynamic social institutions, so that it can be a relevant model to be applied more widely.

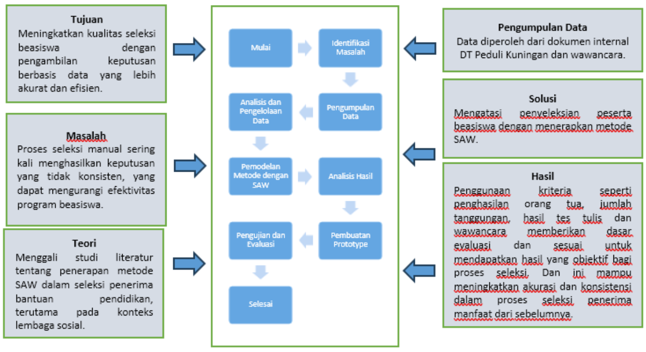


Fig. 1. Research Conceptual Framework

In the visualization of the conceptual framework of the chart above, before a decision is obtained, there is a data processing process through the SAW method, including: Criteria Input, Weighting and Scoring, Final Score Summation (Selection Decision), Scholarship Recipient Recommendation (Selected Recipient with Highest Score) and Continuous Evaluation (Monitoring Recipient Success). Thus this research focuses on how the application of SAW can improve consistency in the selection of scholarship recipients at DT Peduli Kuningan. Thus, this final result will be used as the basis for selecting scholarship recipients [13].

A. Research Approach

This research uses a descriptive quantitative approach, which aims to analyze the selection data of scholarship beneficiaries at DT Peduli Kuningan by applying the Simple Additive Weighting (SAW) method. The SAW method was chosen because of its ability to integrate various selection criteria with different weights to produce consistent and accurate decisions [13], [14].

B. Research Data

This research data was obtained from the internal archives of DT Peduli, which is located in Kuningan Regency. This research aims to develop a SAW-based decision support system, designed to increase the transparency, accuracy and efficiency of beneficiary selection [15]. The information on scholarship candidates is based on four main criteria [16]:

a. Parents' income
Used to evaluate the economic condition of prospective recipients. The lower the income, the more eligible the candidate is to receive the scholarship.
b. Number of family dependents
Indicates the number of financial burdens in the family. The more dependents, the more deserving.
c. Written test
Measures the academic ability or basic knowledge of prospective scholarship recipients.
d. Interview
Interview Assesses the motivation, commitment, and potential of the scholarship recipient

C. Determination of Criteria Weight

The weight is given based on the level of importance of each criterion in the selection process as follows:

a. Parents' income: 40%
b. Number of family dependents: 30%
c. Written test: 20%
d. Interview [17]: 10%

Weights were determined based on literature and discussions with DT Peduli Kuningan [13], [16].

D. Data Conversion and Normalization

The original data on each criterion was converted into a numerical rating scale of 1-5 according to the institution's standard score table. This scale is used to equalize units between diverse data. The following is the normalization formula used in the SAW method, in equations (1) dan (2):

1 For benefit criteria, where higher values are better

$$\frac{x_{ij}}{Max_i\, x_{ij}} \quad \text{if j is a benefit attribute} \tag{1}$$

2 For the cost criterion, where the lower value is better.

$$\frac{Min_{ij}\, x_{ij}}{x_{ij}} \quad \text{"if j is a cost attribute"} \tag{2}$$

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2360, Copyright ©2025
Submitted : May 6, 2025, Revised : May 26, 2025, Accepted : May 26, 2025, Published : May 28, 2025

where:
rij = normalized performance rating
max xij = maximum value of each row & column
mix xij = manimum value of each row & column'
xij = rows and columns of the matrix

E. Calculating the final score

After all values are normalized, the final score of each alternative is calculated by summing the results of multiplying the weights Wj of each criterion and the normalized value r_{ij}

$$S_i = \sum_{j=1}^{n} W_j \; x \; r_{ij} \quad (3)$$

where:
S_i is the total score of the i-th alternative
S_i is the weight of the j-th criterion
IS s the number of criteria Alternative with S_i values

The alternative with the highest S_i value is then selected as the recommended beneficiary [18].

After the initial assessment and selection are completed, ongoing evaluations are conducted to monitor the progress of the beneficiaries [19]. These evaluations are important to measure the success of the scholarship recipients in achieving their educational goals and to identify potential improvements in the selection process in the future [4]

F. Research Stages

The research was conducted in the following stages:

1) Data Collection
Data was obtained from internal DT Peduli Kuningan documents, written test results, and interviews. The data was validated to ensure its completeness and accuracy.
2) Determination of Weight
Giving weight to each criterion based on its level of importance [20].
3) Perform Data Conversion and Normalization
Before calculating using the SAW method, the original data from each criterion is converted into a 1-5 rating scale. This scale is used to homogenize values and avoid inequality between data units (for example, rupiah, number of people, or test scores) [21].
Furthermore, data normalization is carried out based on the type of criteria:
    a. For benefit criteria (the higher the better), see the formula in equation (1).
    b. For cost criteria (the lower the better), see equation (2).
In the context of this research, only parental income is categorized as a cost, while the other criteria are classified as benefits.
4) Calculation of Preference Values
After the normalization process is complete, each normalized value is multiplied by the weight of each criterion [22]. The multiplied values are then summed up to obtain the final preference value for each alternative (scholarship recipient candidate), see equation (3). Semakin tinggi nilai V_i, The higher the value of Vi, the higher the feasibility of alternatives to be selected as scholarship recipients.
5) Result Analysis and Evaluation
The final score is used to rank the potential beneficiaries. Evaluation is carried out to compare the results with the previous manual process, as well as to assess the improvement of the efficiency and accuracy of the selection process [15]
6) Prototype Development of SPK System
The SAW-based decision support system prototype was designed and tested to facilitate the implementation of the selection process in the future[23]

G. Research Tools and Instruments

This research uses several main instruments to support the process of collecting, processing, and analyzing data, which include:

1) Software
    a. Microsoft Excel
    Used to perform SAW method calculations, such as data normalization, criteria weighting, and final score calculation [24]. This tool was chosen because of its flexibility in handling numerical data and the ease of creating automatic formulas for SAW method calculations.
    b. Prototype Development Tool
    To support the implementation of decision support systems (SPK), this research uses web-based prototype software, such as Visual Studio Code and PHP front-end/back-end frameworks [25]. The aim is to build a system that can facilitate the selection of beneficiaries digitally.
2) Questionnaire or Interview Guide
    a. Questionnaire for Interview
    This instrument is used to collect motivation and commitment data from potential beneficiaries through interview sessions. Questions are designed to evaluate the level of motivation, such as educational goals and suitability for the scholarship program.
    b. Written Test Assessment Form
    Used to record written test results that measure the academic ability or basic knowledge of prospective beneficiaries
3) Internal Documents of DT Peduli Kuningan
Internal documents of DT Peduli Kuningan that record the economic condition and number of family dependents of prospective beneficiaries. This document is the main basis for initial assessment according to economic criteria.
4) Data Validation Instrument
Performed to check the completeness, accuracy and relevance of data prior to analysis. Validation ensures that data from physical and digital documents meet predetermined selection criteria.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2360, Copyright ©2025
Submitted : May 6, 2025, Revised : May 26, 2025, Accepted : May 26, 2025, Published : May 28, 2025

Data analysis is carried out using the SAW method, where each criterion is given a weight based on its level of importance. Prospective recipients are then given a value or score on each criterion, which is multiplied by the weight of the criterion. The results of this multiplication are summed up to get the final score. The final score becomes the basis for the beneficiary selection process.

III. RESULT AND DISCUSSION

This research was conducted on scholarship registration taken from Daarut Tauhiid Peduli Kuningan. There are names of prospective scholarship recipients who have registered and have passed the administrative stage totaling thirteen people in the 2022/2023 period. The existing data is then calculated using the Simple Additive Weighting (SAW) method. The following are the names of the participants used:

TABLE I. DATA PARTICIPANT

Code	Gender	Age	Income	Support	Writting Test	Interview
A1	L	22	Rp2.000.000	5	61	77,85
A2	L	20	Rp500.000	3	75	83,905
A3	P	22	Rp2.000.000	3	52	82,2
A4	L	22	Rp1.500.000	3	56	76,5
A5	P	23	Rp1.500.000	2	65	78,24
A6	P	21	Rp750.000	2	47	86
A7	P	22	Rp500.000	5	59	78,475
A8	L	19	Rp1.250.000	3	54	78,85
A9	L	20	Rp500.000	3	73	82,1
A10	P	23	Rp500.000	3	59	79,32
A11	L	22	Rp1.900.000	5	38	77,48
A12	L	23	Rp500.000	3	44	84,2
A13	P	19	Rp500.000	3	63	79,77

(Source: DT Peduli Kuningan internal data)

Table 1 shows the list of potential scholarship recipients along with the value of each criterion used as the basis for assessment. Each criterion such as parents' income, number of dependents, written test results and interviews are given the appropriate weight to describe the relative importance of each aspect in determining the eligibility of scholarship recipients.

The assessment is carried out by determining the criteria and weights that will be used as consideration in the decision. The criteria can be seen in Table 2.

TABLE II. CRITERIA DETERMINATION

No	Code	Criteria	Attribute	Value
1	C1	Parental income	Cost	0,40
2	C2	Number of dependents	Benefit	0,30
3	C3	Written Test	Benefit	0,20
4	C4	Interview	Benefit	0,10
Total				**100%**

DT Peduli Kuningan assessment policy based on research [2]

Tabel 2 shows the details of the weight for each criterion that is taken into consideration in the DT Peduli Kuningan assessment system.

In the initial stage, data on potential scholarship recipients is collected and categorized based on predetermined criteria, such as parents' income, number of dependents, written test results, and interview results. Each of these criteria is given a weight based on its level of importance in determining the eligibility of scholarship recipients. These weights were set proportionally at 0.25 for each criterion, in order to reflect a balanced role in the selection process.

Furthermore, Table 2 to Table 5 show the point assessment for each of the criteria used, namely Parents' Income, Number of Dependents, Written Test Scores and Interviews. Each value point is determined based on policies and agreements that have been designed by the institution as an assessment standard that suits the selection needs. In this case, each alternative (participant) is calculated based on a rating scale of 1-5. This scale gives additional value to the participant with the highest score in the benefit category and vice versa for the cost category. Through this comparison, the system can determine the ranking of the most eligible participants based on the calculation results using the SAW method. And the determination of points in each table is certainly based on quantitative and qualitative evaluations that are adjusted to the purpose of the scholarship program, which is to provide assistance to prospective recipients who are most in need and have a high potential for success.

With the sub criteria of Parents' income, Number of Dependents, Written Test Score, Interview, can be seen in tables 3,4,5 and 6, as follows:

TABLE III. PARENTS INCOME

Code	Parent's Income	Value Points
C1	< 1.000.000	5
	1.000.000 – 1.500.000	4
	1.500.000 – 2.000.000	3
	2.000.000 – 2.500.000	2
	> 2.500.000	1

Source: Income assessment policy adjusted to DT Peduli Kuningan selection needs

Parents' income in table 3 is categorized into 5 groups based on a range of scores. The lower the income, the higher the score.

TABLE IV. NUMBER OF DEPENDENTS

Code	Number of Dependents	Value Points
C2	5 Child	5
	4 Child	4
	3 Child	3
	2 Child	2
	1 Child	1

Source: Dependent assessment policy adjusted to the needs of DT Peduli Kuningan selection

Scores are assigned based on the number of family dependents. The greater the number of dependents, the higher the score given.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2360, Copyright ©2025
Submitted : May 6, 2025, Revised : May 26, 2025, Accepted : May 26, 2025, Published : May 28, 2025

TABLE V. WRITTEN TEST SCORE

Code	Written Test Score	Value Point
C3	80 > 90	5
	75 > 80	4
	70 > 75	3
	60 > 70	2
	< 60	1

Sumber: Kebijakan penilaian tes tulis disesuaikan dengan kebutuhan seleksi DT Peduli Kuningan (2023)

Scores are assigned based on the written test score range, using the institution's policy standards to determine the range.

TABLE VI. INTERVIEW

Code	Interview	Value Points
C4	80 > 90	5
	75 > 80	4
	70 > 75	3
	60 > 70	2
	< 60	1

Sumber: Kebijakan penilaian wawancara disesuaikan dengan kebutuhan seleksi DT Peduli Kuningan

Source: The interview assessment policy is adjusted to the needs of DT Peduli Kuningan selection.

TABLE VII. DATA CONVERSION PROCESS

Code	Income	Dependents	Written Test	Interview
A1	4	5	2	4
A2	1	3	4	5
A3	4	3	1	5
A4	3	3	1	4
A5	3	2	2	4
A6	1	2	1	5
A7	1	5	1	4
A8	2	3	1	4
A9	1	3	3	5
A10	1	3	1	4
A11	3	5	1	4
A12	1	3	1	5
A13	1	3	2	4

Source: Conversion results based on manual scores applied to DT Peduli Kuningan data

Table 7 explains that the raw data from Table 1 is converted to numerical values based on the scale in Tables 3-6. This process uses the following logic:

a. If the participant's income is Rp1,500,000, then according to Table 3, the conversion value is 3.
b. If the family dependents are 5 people, then the conversion value is 5 according to Table 4.

TABLE VIII. NORMALIZATION OF THE DECISION MATRIX

Code	Income	Dependents	Written Test	Interview
A1	0,25	1	0,5	0,8
A2	1	0,6	1	1
A3	0,25	0,6	0,25	1
A4	0,333333333	0,6	0,25	0,8
A5	0,333333333	0,4	0,5	0,8
A6	1	0,4	0,25	1
A7	1	1	0,25	0,8
A8	0,5	0,6	0,25	0,8
A9	1	0,6	0,75	1
A10	1	0,6	0,25	0,8
A11	0,333333333	1	0,25	0,8
A12	1	0,6	0,25	1
A13	1	0,6	0,5	0,8

Source: The calculation of normalization is done by SAW formula adapted from previous research [4]

In table 8, select one or two rows of the table. For example, for income, using equation (2), if the A1 value is 4 and the maximum value is 1 (see table 7), then:

$$r^{ij} = \frac{1}{4} = 0{,}25$$

And do the same for the other criteria.

TABLE IX. PREFERENCE VALUE CALCULATION

KODE	NILAI PREFERENSI SAW
A1	0,58
A2	0,88
A3	0,43
A4	0,443333333
A5	0,433333333
A6	0,67
A7	0,83
A8	0,51
A9	0,83
A10	0,71
A11	0,563333333
A12	0,73
A13	0,76

Source: Preference calculation using SAW method with criteria weights from internal policies

The final stage of this analysis is the calculation of the total preference score based on the SAW method, where the normalized value of each participant is multiplied by the weight

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2360, Copyright ©2025
Submitted : May 6, 2025, Revised : May 26, 2025, Accepted : May 26, 2025, Published : May 28, 2025

of each criterion and then summed to obtain the final score. In table 9, to calculate the preference value use the equation (3). For A1, that is:

1 Income: 0,25×0,40 = 0,1
2 Dependents: 1×0,30 = 0,3
3 Written Test: 0,5×0,20 = 0,1
4 Interview: 0,8×0,10 = 0,08

The results were then accumulated between income, dependents, written tests and interviews. So that it becomes 0,1+0,3+0,1+0,08 = 0,58. After calculating all preference values, do the sorting, so that the results can be seen in table 10.

TABLE X. SORTING DATA BY RANK

RANK	CODE	Gender	Age	SAW PREFERENCE VALUE
1	A2	L	20	0,88
2	A7	P	22	0,83
3	A9	L	20	0,83
4	A13	P	19	0,76
5	A12	L	23	0,73
6	A10	P	23	0,71
7	A6	P	21	0,67
8	A1	L	22	0,58
9	A11	L	22	0,563333333
10	A8	L	19	0,51
11	A4	L	22	0,443333333
12	A5	P	23	0,433333333
13	A3	P	22	0,43

Table 10, namely A2 get the highest score, namely 0.88 because it has a very low income, quite a lot of dependents and high test/interview scores. This reference assessment is obtained to determine the determination of the acceptance of educational funding assistance at DT Peduli Kuningan using the Simple Additive Weighting (SAW) method.

The final results of this calculation are then sorted to determine the ranking of beneficiaries. Participants with the highest scores are declared as the most deserving of the scholarship, while participants with lower scores are recommended as candidates who may not meet the eligibility criteria.

The application of the SAW method is proven to provide several advantages for DT Peduli Kuningan, including a more consistent, transparent, and easily accountable selection process. Assessment based on standardized criteria weights allows this institution to provide selection decisions more systematically and data-based. The results of this study are consistent with previous research which states that the SAW method is an effective method in providing objective and accurate multi-criteria decisions [8]. The system also exhibits high flexibility in weight adjustment between criteria, which allows agencies to change priorities based on need without changing the overall structure of the system.

With this system, DT Peduli Kuningan can minimize potential bias in assessment and simplify the beneficiary selection process despite changes in staff or those in charge. In addition, this system also provides an opportunity to conduct continuous evaluation of beneficiaries, so that the institution can monitor their progress and provide more effective support according to the needs of the scholarship recipients. Based on these results, this SAW-based decision support system is expected to be a practical and applicable solution for other institutions that have similar goals in distributing educational assistance.

The existence of different weights between criteria provides fairness in decision-making on which is preferred, but can be adjusted in the future based on changes in program priorities. The normalization and preference assessment shows that participants with lower income and more dependents tend to get higher scores. This shows the consistency of the system in prioritizing participants who are more in need of educational assistance.

Evaluation of the results shows that the SAW method not only produces transparent selection decisions but also simplifies the replication process for the next period. This research reveals that the Simple Additive Weighting (SAW) method provides an effective solution in overcoming the challenges of selecting scholarship beneficiaries at DT Peduli Kuningan. Compared to manual or subjectivity-based selection methods, the use of SAW allows for more objective and consistent decision-making. This method works by standardizing the value of each candidate on certain criteria, which is then processed with proportional weights, thus providing a final result that reflects the level of eligibility quantitatively. This is in line with Suwarno's (2021) findings that SAW is suitable for use in multi-criteria selection to ensure transparency and accuracy in beneficiary selection [8].

Further discussion shows that the SAW method is not only easy to implement but also flexible in setting criteria weights. This flexibility allows DT Peduli Kuningan to adjust the weights based on priorities or applicable policies, without affecting the entire system. This advantage places SAW as an excellent method for scholarship selection cases in social institutions that may often face data management constraints, including inconsistencies due to staff turnover. The use of different criteria weights in this study proved effective in producing objective rankings, although future research could explore the possibility of more specific weights according to the needs of potential recipients.

However, the SAW method also has limitations, particularly in terms of its reliance on data that must be complete and structured. For beneficiary selection that involves incomplete or poorly documented data, SAW may be less than optimal because the normalization process requires complete data for each candidate. Therefore, more systematic data management and digitization are important for this system to be implemented

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2360, Copyright ©2025
Submitted : May 6, 2025, Revised : May 26, 2025, Accepted : May 26, 2025, Published : May 28, 2025

optimally. This research suggests that the integration of SAW with cloud-based technology or centralized data could be the next step to improve the effectiveness of the system.

Furthermore, this research has practical implications for similar institutions that want to adopt the SAW method as a decision support system. The findings of this study provide evidence that the SAW method can help improve the quality of scholarship beneficiary selection and reduce the risk of bias in decision-making. In addition, continuous monitoring of beneficiaries is also recommended to measure the success of the program in the long term. The implementation of post-selection evaluation is expected to correct the shortcomings of the previous selection and adjust the support for scholarship beneficiaries. Overall, the SAW method proved to be relevant and contributed significantly to improving the effectiveness of the scholarship program at DT Peduli Kuningan, and has the potential to be applied in other institutions with similar objectives.

IV. Conclusion

The decision support system developed with the Simple Additive Weighting (SAW) method is able to provide solutions to previous obstacles, such as inconsistencies in eligibility criteria and less systematic data processing and has a good accuracy rate of 84.62% and has consistency in the selection process of scholarship beneficiaries at DT Peduli Kuningan. The use of criteria such as parents' income, number of dependents, written test results and interviews provide an objective evaluation basis for the selection process. This SAW method allows each criterion to be processed with a certain weight, resulting in a ranking that helps the institution in determining scholarship recipients more effectively. This process not only reduces subjectivity in decision making but also provides a system that is reliable and easily adaptable to the needs of the institution.

The SAW system can be further developed with the integration of cloud-based technology to improve efficiency and data accessibility. Digitization of processes will ensure consistency despite staff turnover. Further research can explore the variation of weights on the criteria to adjust the selection priorities to the needs of the institution.

Acknowledgements

The authors would like to express their deepest gratitude to DT Peduli Kuningan as the object of research that has provided valuable support, data and information for the smooth running of the research. Gratitude is also expressed to those who have provided funds and resources that have made this research possible. This support is very important in the effort to develop a more effective and transparent scholarship recipient selection decision support system.

References

[1] Dicky Hendardi, Sucipto Sucipto, and Addiarrahman Addiarrahman, ‘Model Pengembangan Zakat Produktif di Lembaga Amil Zakat Daarut Tauhiid Peduli Jambi (Studi Kasus Daarut Tauhiid The Hok Jambi, Kota Jambi)’, *Jurnal Kajian dan Penalaran Ilmu Manajemen*, vol. 2, no. 1, pp. 160–173, 2024, doi: 10.59031/jkpim.v2i1.252.

[2] Hidayatullah and Lia Umbari Putri, ‘Sistem Pendukung Keputusan Penentuan Penerima Beasiswa BAZNAS Kabupaten Asahan Dengan Metode AHP’, *Jurnal Informatika dan Teknologi Informasi*, vol. 3, no. 1, pp. 293–299, May 2024, doi: 10.56854/jt.v3i1.366.

[3] R. Taufiq Subagio and M. Thoip Abdullah, ‘Penerapan Metode SAW (Simple Additive Weighting) dalam Sistem Pendukung Keputusan untuk Menentukan Penerima Beasiswa Application of SAW (Simple Additive Weighting) Method in System Decision Supporters to Determine Scholarship Recipients’. [Online]. Available: https://id.wikipedia.org/wiki/Beasiswa,

[4] S. Melati and G. Triyono, ‘Pemodelan Sistem Pendukung Keputusan Penentuan Siswa Terbaik Menggunakan Metode Simple Addictive Weighting (Saw)’, *IDEALIS : InDonEsiA journaL Information System*, vol. 3, no. 2, pp. 574–580, 2020, doi: 10.36080/idealis.v3i2.2748.

[5] D. Achmad, S. Mu, and A. Saputro, ‘Nusantara Computer and Design Review Sistem Pendukung Keputusan Penerimaan Beasiswa Menggunakan’, pp. 24–30, 2023.

[6] T. P. Sihaloho, R. Sari, B. Sembiring, V. S. Ginting, and J. Simanullang, ‘Penerapan Metode Simple Additive Weighting Dalam Mendukung Keputusan Penerima Beasiswa Pada MTs Swasta Yayasan Pendidikan Ibadah’, vol. 12, pp. 15–24, 2023.

[7] A. Y. Pratama and S. Yunita, ‘Komparasi Metode Weighted Product (WP) Dan Simple Additive Weighting (SAW) Pada Sistem Pendukung Keputusan Dalam Menentukan Pemberian Beasiswa’, vol. 4, no. September, pp. 12–24, 2022, doi: 10.30865/json.v4i1.4593.

[8] I. P. E.- Issn, S. Liang, and M. R. Muhtarom, ‘Computer Based Information System Journal Sistem Pendukung Keputusan Penentuan Penilaian Siswa Dengan Metode SAW (Simple Additive Weighting)’, vol. 01, pp. 23–36, 2021.

[9] N. Indriyani, A. Fauzi, A. Bayu, and H. Yanto, ‘Pemodelan Prediksi Penerima Beasiswa KIP Kuliah Menggunakan Metode Weight Product’, vol. 5, no. 1, 2024.

[10] N. S. J. Putri and K. Hati, ‘Implementasi Metode SMART pada Seleksi Penerima Beasiswa Tingkat Tinggi di Lembaga Amil Zakat’, *Jurnal Teknik Informatika Stmik Antar Bangsa*, vol. VIII, no. 2, pp. 50–58, Aug. 2022.

[11] M. Afrina, ‘Sistem Pendukung Keputusan Penerimaan Beasiswa Pada Rumah Zakat Dengan Metode Fuzzy Analytical Network Proces (FANP) Di Baitul Mal’, vol. 2, no. 1, pp. 15–24, 2018.

[12] T. B. Wicaksono, O. S. Bachri, and B. Irawan, ‘Perancangan Sistem Pendukung Keputusan Seleksi Penerima Beasiswa dengan Pendekatan Metode MOORA Berbasis Web’, vol. 10, no. April, pp. 1–10, 2024, doi: 10.34128/jsi.v10i1.887.

[13] A. Sholihat and D. Gustian, ‘Sistem Pendukung Keputusan Pemilihan Siswa Berprestasi Dengan Metode Simple Additive Weighting (SAW) (Studi Kasus: SMK Dwi Warna Sukabumi)’, pp. 140–147, 2021.

[14] A. Adikvika, N. Merlina, and N. A. Mayangky, ‘Sistem Pendukung Keputusan Pemilihan Penerima Beasiswa Pendidikan Dengan Menggunakan Metode Weighted Product Di Yatim Mandiri’, vol. 7, no. 2, pp. 148–158, 2021.

[15] M. I. A. Putera and M. G. L. Putra, ‘Keputusan Penenruan Calon Penerima Beasiswa Menggunakan Metode Simple Additive Weighting Pada Kpw Bank Indonesia Balikpapan’, vol. 14, no. 2, pp. 110–120, 2020.

[16] B. Mukaromah and S. Iswanti, ‘Sistem Pendukung Keputusan Menentukan Siswa Penerima Beasiswa dengan Metode Simple Additive Weighting berbasis PaaS Cloud Computing’, vol. 2, no. 3, pp. 1–11, 2023, doi: 10.26798/juti.v2i3.891.

[17] D. N. Sulistyowati, R. P. Sari, N. Mandiri, U. N. Mandiri, S. Pemilihan, and S. A. Weight, ‘Implementasi Metode Simple Additive Weight’, vol. 5, no. 2, 2023.

[18] K. Fikroh Adilah and K. Syabani Jasmine, ‘Implementasi Sistem Perhitungan Keputusan Penilaian Mahasiswa Terbaik Dengan Metode Simple Additive Weighting (Saw) (Studi Kasus: Universitas Abc)’.

[19] M. O. Mahendra and E. G. Sari, ‘Penerapan Metode Simple Additive Weighting (SAW) Untuk Menunjang Keputusan Penilaian Kinerja Guru (PKG)’, *Digital Transformation Technology*, vol. 4, no. 1, pp. 232–243, Jun. 2024, doi: 10.47709/digitech.v4i1.3704.

[20] L. Jurnal, I. Gede Iwan Sudipa, and K. Sri Aryati, ‘Pendekatan Penentuan Bobot dengan Surrogate Weighting Procedures untuk Metode Simple Additive Weighting dalam Pengambilan Keputusan Multikriteria’, *International Journal of Natural Sciences and Engineering*, vol. 3, no. 3, pp. 113–121, 2019, [Online]. Available: https://ejournal.undiksha.ac.id/index.php/IJNSE

[21] H. Destiana, Y. Handrianto, A. Sudrajat, and K. Nurseha, 'Metode Simple Additive Weighting (SAW) Saat Mengevaluasi Kinerja Karyawan pada PT. Hijrah Insan Karima'. [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/infortech

[22] E. Prayitno, D. Rahman Habibie, I. Mariami, and A. Hadi Nasyuha, 'KLIK: Kajian Ilmiah Informatika dan Komputer Simulasi Pemilihan Partai Politik Menggunakan Simple Additive Weighting', *Media Online*, vol. 4, no. 3, pp. 1880–1887, 2023, doi: 10.30865/klik.v4i3.1416.

[23] H. Hidayat and H. Saleh, 'S iste m Pen d u k u n g Keputusan Peneri ma Bantuan Beasisw a Pemerintah Daerah Kabupaten Boalemo Menggunakan Metode Simple Additive Weigthing (SAW) : Studi Kasus Dinas Pendidikan Dan Kepemudaan Kabupaten Boalemo', vol. 6, no. 6, pp. 767–773, 2023.

[24] R. K. Serli, V. Indriyani, and M. Rahmayu, 'Penerapan Metode Simple Additive Weighting (Saw) Untuk Menentukan Perangkingan Guru Berprestasi Studi Kasus: Sdn Rambutan 03 Pagi', *Journal Speed-Sentra Penelitian Engineering dan Edukasi*, vol. 14.

[25] S. Siswidiyanto, A. Munif, D. Wijayanti, and E. Haryadi, 'Sistem Informasi Penyewaan Rumah Kontrakan Berbasis Web Dengan Menggunakan Metode Prototype', *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 15, no. 1, pp. 18–25, Apr. 2020, doi: 10.35969/interkom.v15i1.64.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2360, Copyright ©2025
Submitted : May 6, 2025, Revised : May 26, 2025, Accepted : May 26, 2025, Published : May 28, 2025

Skincare Recommendation System Based on Facial Skin Type with Real-Time Weather Integration

Gabrielle Sheila Sylvagno[1], Theresia Herlina Rochadiani[2]
Department of Information Technology[1], [2]
University of Pradita
Tangerang Selatan, Indonesia
gabrielle.sheila@student.pradita.ac.id[1], theresia.herlina@pradita.ac.id[2

***Abstract*— Skin conditions can be significantly affected by unpredictable weather changes, creating the need for a solution that can provide personalized skincare product recommendations. This study presents the development of an AI-based skincare recommendation system that integrates skin type classification using Convolutional Neural Networks (CNN) with real-time weather data via the OpenWeatherMap API. The system consists of three main components: a ResNet50-based Skin Analyzer, a Weather Analyzer using the Decision Tree algorithm, and a Product Recommendation module. The image dataset is sourced from two Kaggle datasets: "Dry, Oily, and Normal Skin Types" and "Acne Dataset." The total dataset consists of 2,885 images, divided into four classes: Acne (549 images), Dry (652 images), Normal (884 images), and Oily (800 images). The dataset exhibits diversity in skin types, allowing for a more valid evaluation of the CNN model. The training and testing process involved splitting the data into training and testing sets, with augmentation applied to the training data to enhance the feature diversity across classes. Evaluation results show an average validation accuracy of 90.94% ± 0.60% with consistent performance. This system aids users in identifying their skin type and suggests appropriate skincare products based on current weather conditions. It is expected to contribute to the advancement of AI-driven personalization in the skincare industry.**

Keywords*— *Weather, Skin Type Classification, ResNet50, Product Recommendation, Skincare

I. Introduction

In the ever-evolving digital era, artificial intelligence technology (Artificial Intelligence/AI) has been integrated into various aspects of life, including the beauty and skincare industries. AI can analyze large amounts of data and provide personalized recommendations more accurately [1]. One of the increasingly popular applications of AI is the recommendation system Skincare, which considers factors such as the individual preferences, the user's skin type, and environmental conditions, especially the weather Real-time. Environmental factors such as humidity, temperature, and UV exposure have a significant influence on skin health. For example, oily skin tends to be more prone to acne in hot and humid weather, while dry skin is more prone to irritation at low temperatures [2].

Based on surveys ZAP Beauty Index in 2023 and 2024, only about 5% of Indonesian women have a normal skin type. The rest experience various problems, such as dull skin (53.8%), panda eyes (33.3%), and signs of aging that are beginning to be felt by almost 30% of Gen Z women. Meanwhile, 99% of Indonesian men do not know their skin type, and although 56.6% understand the content of the product Skincare, only 4.1% understand it in depth [3]. The study indicates that there is still a significant number of users who struggle to select skincare products for their skin type and adjust to weather conditions.

Deep learning technology, especially regarding Convolutional Neural Network (CNN), holds a crucial role in improving skincare recommendation systems to be more adaptive. CNN can be used to analyze faces and classify them based on skin types, like oily, dry, and prone to acne. The data will then be combined with real-time weather information via API, used to give recommendations on products that are relevant to the user's skin condition and environment [4]. Information regarding skin types can be combined with real-time data that is obtained via API to give recommendations on skincare products according to the user's environment. A study in [5] considered using deep learning to analyze the content of cosmetics and skin condition, but it is void of the variable of environment, such as weather conditions [5]. Another study [6] focus mainly on product personalization on users and like the study in [5] did not consider the environmental conditions.

In everyday situations, weather conditions play an important role in a person's skin needs. Hot and humid weather often triggers increased oil production on the face, while cold temperatures can cause the skin to become dry and more prone to irritation [7]. Therefore, combining skin type classification technology with real-time weather data is a potential solution to provide more personalized and relevant skincare product recommendations. Not only that, the use of a camera to scan the face also allows skin condition analysis to be done instantly and practically, without the need for special tools or direct consultation with an expert [8].

In research conducted by [9], entitled “Deep learning-based skin care product recommendation: A focus on cosmetic ingredient analysis and facial skin conditions” focuses on analyzing cosmetic ingredients with deep learning and YOLOv4 to detect skin conditions. The results showed that the approach was able to identify the effectiveness of skincare products through ingredient analysis using deep learning, as well as provide personalized product recommendations based

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2355, Copyright ©2025
Submitted : May 2, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

on the user's skin condition. However, the model still has limitations, especially in terms of the limited amount of valid data and the narrow scope of skin conditions, which affects the performance and generalization ability of the system [9]. Although Lee et al. proposed a robust integration of ingredient-based analysis and facial skin detection using U-Net and Transformer architectures, the system does not factor in external environmental variables such as real-time weather conditions, thereby limiting the contextual adaptability of the recommendations.

A study by [10] entitled "Designing a System to Recommend Skincare Products Using the NLP Method" proposes a Natural Language Processing (NLP)-based skincare product recommendation system that utilizes user reviews from the Female Daily platform. The system uses the cosine similarity method to measure product suitability with the user's skin type and successfully shows high accuracy in providing recommendations. This research is relevant because both aim to provide personalized skincare recommendations, although the approaches used are different [10]. While effective in measuring product suitability with user skin types, it lacks adaptability to real-time weather changes. Their approach leverages user-generated textual reviews, but the system solely depends on historical review content without accommodating dynamic factors like weather or instant skin conditions detected via image processing.

A study by [11] entitled "A Systematic Approach for Skin Disease Detection & Prediction by using CNN" developed an image classification-based skin disease diagnosis system using the Convolutional Neural Network (CNN) algorithm. The system receives input in the form of images of infected skin areas, then classifies them into certain types of skin diseases, such as melanoma, nevi, or other benign lesions. To improve accuracy, this research also uses data augmentation techniques because the dataset used is quite unbalanced. The results show that CNN has an accurate performance in recognizing various skin diseases from dermatoscopic images [11]. Although effective for medical diagnosis, it does not address cosmetic skincare recommendations or consider environmental adaptability.

Although several previous studies have explored AI-based skincare recommendations, facial skin classification, or cosmetic ingredient analysis, many of them ignore the influence of real-time environmental conditions, especially weather. This shows that most systems are still limited in terms of contextual awareness, despite the proven impact of environmental changes on skin health.

This research aims to design an artificial intelligence-based skincare recommendation system by combining skin type classification using CNN and current weather information to produce appropriate product suggestions. This research is based on the need for adaptive solutions in choosing skincare products, considering that weather changes can affect skin conditions and become a challenge for users in maintaining consistency in skincare routines. By applying machine learning and deep learning algorithms, this system is developed to help users choose appropriate skincare products based on a combination of skin type and current environmental conditions.

In terms of implementation, the developed system is expected to assist users in choosing skincare products more efficiently and in accordance with their actual conditions. In addition, the system also aims to educate users on the importance of adjusting skincare products to changing environmental conditions. The target audience of this research includes skincare product users who need technological assistance to improve skincare effectiveness, as well as researchers or application developers who are interested in integrating weather data into machine learning and deep learning-based models. Thus, this research is expected to contribute to the advancement of personalization technology in the beauty industry, which is increasingly sophisticated, adaptive, and data-driven.

Most previous studies only focused on either skin type classification or general skincare recommendations, and rarely considered changing weather conditions like temperature or humidity. This research fills that gap by combining facial skin analysis using CNN with real-time weather data, so the recommendations given can be more accurate and suitable for each user's current situation.

II. RESEARCH METHODS

This research focuses on utilizing AI technology to provide personalized skincare product recommendations based on real-time weather conditions and the user's skin type. The results of this research are intended for individuals who care about skincare, especially those who experience problems such as oily, dry, and acne-prone skin. This research uses a quantitative approach by utilizing numerical and image data to train and evaluate CNN and Decision Tree models in a real-time weather-based skincare recommendation system.

The system works by utilizing the latest weather data through APIs and machine learning-based facial scans to detect the user's skin condition. The analysis results are used to recommend appropriate skincare products, such as moisturizer or sunscreen. This research aims to solve users' confusion in choosing skincare products and to adapt skincare to changing weather conditions.

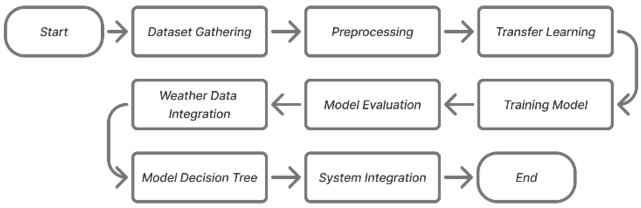


Fig. 1. Decision Tree-Based Recommendation Model

Figure 1 shows the workflow of the system developed in this study. The process begins with Dataset gathering and preprocessing to prepare facial image data. Furthermore, Transfer Learning and Training Models were carried out using CNN, followed by Model Evaluation to obtain the best skin type classification model. After that, the system integrates weather information at the Weather Data Integration stage, which is then combined with the Decision Tree model to

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2355, Copyright ©2025
Submitted : May 2, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

generate skincare product recommendations. All stages end with System Integration to build a system that is ready to use.

Transfer learning is an approach that leverages pre-trained models to complete new tasks on different datasets, so the training process does not need to be done from scratch. In its implementation, adjustments are usually made at the end of the model to match the characteristics of the new data [12]. The CNN model itself generally consists of several layers that are in charge of extracting features from the image data.

A. Dataset and Preprocessing

The dataset used in this study comes from two different sources on the Kaggle platform, namely the "Dry, Oily, and Normal Skin Types" by Shakya Dissanayake [13] and "Acne Dataset" by Nayan Chaure [14]. The two datasets are then combined and classified into four skin types: Acne, Dry, Normal, and Oily. The dataset consists of four main classes, namely Acne, Dry, Normal, and Oily, with an unbalanced amount of data between classes. The total number of images used is 2,885, with the following distribution: Acne (549 images), Dry (652 images), Normal (884 images), and Oily (800 images). It should be noted that the datasets used do not include metadata such as age, gender, or ethnicity. As a result, this study does not analyze skin type classification performance across demographic groups. Future work is recommended to incorporate more diverse datasets to enhance generalizability.

The researcher decided to use the dataset in its original form without balancing (such as undersampling or oversampling), because based on the initial experiments, the balancing effort actually caused a significant decrease in model performance, with an accuracy of only about 50%–60% compared to the use of the original data which achieved an accuracy of more than 90%. Therefore, the data remains used in its original distribution to maintain the stability and accuracy of the model. The datasets are organized in a folder structure according to their respective classes and are loaded using ImageFolder from PyTorch. This process includes several stages:

a. Resize: All images in each dataset class are resized to 224x224 pixels to fit the required standard input of the CNN model.
b. ToTensor: The image is converted from the PIL Image format to the PyTorch tensor, with the pixel value scaled between 0 to 1.
c. Normalization: The image is normalized using the mean (mean) and standard deviation (std) values of the ImageNet dataset, which are:
    - Red: [0.485, 0.456, 0.406]
    - Std: [0.229, 0.224, 0.225]

B. Skin Type Classification Model

The architecture used is a ResNet50-based CNN. ResNet50 was chosen because it has a residual block structure that helps solve the problem of vanishing gradients during deep model training.

The model is trained using pre-processed facial data, with the configuration shown in Table 1:

- Loss Function: CrossEntropyLoss, because the classification is multi-class.
- Optimizer: AdamW, with an initial learning rate of 5e-5.
- Epochs: Models are trained for 20 epochs.
- Batch size: 32.

TABLE I. SKIN CLASSIFICATION MODEL CONFIGURATION

Yes	Component	Detail
1	*Loss Function*	*CrossEntropyLoss*
2	*Optimizer*	Adam
3	*Learning rate*	5e-5
4	*Epochs*	15
5	*Batch Size*	32

1. ResNet50-Based CNN Model

Convolutional Neural Network (CNN) is a type of artificial neural network architecture that is commonly used in image processing due to its ability to extract spatial features from images through a multi-stage learning process. CNN consists of several main layers, such as convolutional layers, activation layers (e.g., ReLU), pooling layers, and fully connected layers. The main advantage of CNN is its ability to recognize features in images well, even in conditions that are subject to position, scale, or distortion [15].

In the context of skin disease diagnosis, CNN is used to recognize certain visual patterns and characteristics in skin images, which allows the system to classify types of skin diseases with greater accuracy. CNN can learn to detect the distinctive features of skin diseases through the training process, so that it is able to distinguish diseases based on features that are difficult for the human eye to see [16].

The architecture used in the skin type classification model in this study is a ResNet50-based CNN. ResNet50 is a type of Deep Network residual learning-based that facilitates training by considering the Input layer as a reference. ResNet50 has 50 layers, where each block passes through three layers, including a 1x1 convolution layer [17]. In this study, ResNet50 was used as a Feature Extractor and a fine-tuned classification model on a facial dataset with skin type labels. The training process was carried out using Adam optimization and Loss Function of Cross-Entropy Loss.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2355, Copyright ©2025
Submitted : May 2, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

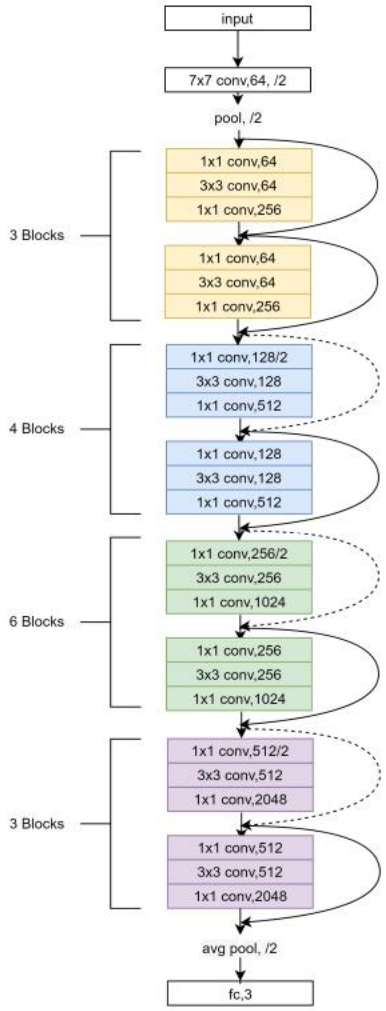


Fig. 2. ResNet50 Layer Architecture [18]

2. Tools and Frameworks

This study uses PyTorch as a Framework, a key element in model development for Deep Learning. PyTorch is one of the libraries in the Python programming language designed to support computing Deep Learning. Library. It is known for its flexibility in building models. Deep Learning uses intuitive and expressive Python syntax. This easy-to-use approach has made PyTorch popular among researchers since its inception, and over time, it has evolved to become one of the key tools in app development, Deep Learning Wide [19].

PyTorch also provides various modules that simplify the training process, such as torchvision. models for pretrained models (ResNet50), DataLoader for dataset management, and transforms for image augmentation and preprocessing. In addition to PyTorch, other tools and supports are:

- OpenCV: To detect faces in real-time via webcam before skin type classification.
- Torchvision: To load the dataset using ImageFolder and image transformation.
- Matplotlib and Seaborn: For visualization of accuracy, loss, and confusion matrix.
- Jupyter Notebook: As a model development and training environment.
- OpenWeatherMap API: To retrieve real-time weather data that will later be used in the recommendation system.

C. Weather and Its Effects on Skin

The weather has an important influence on skin health [7]. UV rays that are too strong will make the skin quickly damage and trigger diseases such as skin cancer [5]. Excessive sun exposure will make unprotected skin will be damaged, because human skin needs treatment that is appropriate to the skin problems it faces [20].

Exposure to sunlight and UV rays can cause the skin to appear dull, especially when sunscreen is not used. To address this, the application leverages real-time weather data to provide skincare recommendations that align with current weather conditions. By accessing weather information through an API, the system can tailor skincare advice based on factors such as UV intensity and humidity levels, helping users adjust their routines accordingly.

D. Face Detection

Before the skin type classification process is carried out, the initial stage is real-time detection of the user's face. This face detection serves to ensure that the input analyzed by the CNN model is an area of the face and not the background or other parts of the body. In this study, the MTCNN (Multi-task Cascaded Convolutional Networks) algorithm is used for real-time face detection and alignment prior to classification.

MTCNN is a method of face detection and alignment based on deep convolutional neural networks that performs joint face detection and facial landmark localization across three cascading stages of networks: P-Net, R-Net, and O-Net [21].

The rapid evolution of deep learning has significantly enhanced CNN-based face detection techniques. Early models like AlexNet, VGGNet, GoogLeNet, and ResNet have laid the groundwork for deep learning-driven improvements in image recognition and object detection [22].

Haar Cascade and MTCNN was two of the option choosen in this study, but considering that MTCNN has better robustness and accuracy compared to Haar Cascade, it was chosen as the ideal method, especially considering that under less ideal conditions such as partially covered faces, faces that are not directly facing the camera and low lighting, it could still detect faces while identying facial features thus making it very suitable for real-time applications that require accuracy and fast responds.

One of the primary reasons to use MTCNN is that it is robust for nonideal case. In tests, the system was able to accurately identify faces at all times, whether the user had light makeup or was in a low light environment. This is to improve the applicability of the system in daily life conditions and the reliability of the skincare recommendation process, which is proven with more operational angles use for recommendation and without the need for an ideal lighting or face directly facing the camera.

The MTCNN-based face detection system is created to be

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2355, Copyright ©2025
Submitted : May 2, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

implemented in real-time and capable of detecting the face region accurately and processing it quickly. Performance measures of latency and frame rate will be outlined in greater depth in the Results section in Figure 7 though the decision to use MTCNN was based on its previously demonstrated real time performance. This is essential to enable the interactive characteristic of the skincare recommender system.

Once the skin type classification model has been trained and validated, it can be applied into a real-time face detector. In this setup, we used the MTCNN method to detect faces from the webcam directly. MTCNN was selected because of its higher accuracy and robustness than other methods (Haar Cascade,etc.), and in the uncertain lit conditions and face directions.

E. Skincare Product Recommendations

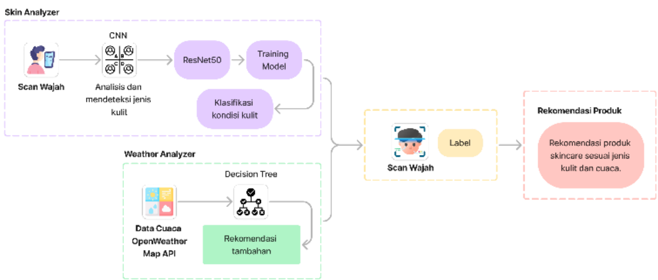


Fig. 3. Skincare Product Recommendation Framework

The proposed framework can be seen in Figure 3. The system consists of three main components, namely: Skin Analyzer, Weather Analyzer, and Product Recommendation Module. The Skin Analyzer component uses the ResNet50-based Convolutional Neural Network (CNN) model to classify the skin type of the user's facial image. The Weather Analyzer component utilizes the Decision Tree algorithm to generate additional recommendations based on weather data obtained in real-time through the OpenWeatherMap API. Furthermore, the results of skin type classification and weather information are then combined in the Product Recommendation Module, which will provide skincare product suggestions that best suit the user's needs based on current skin and weather conditions.

The service is configured to produce advice on the most suitable skincare product products to use for one's skin condition and for current weather but does not include analysis of the particular products composition of ingredient. The whole procedure is kicked off from the Skin Analyzer, where a ResNet50 CNN model is applied to predict and classify the user skin type according to a pre-trained skin dataset. Meanwhile, the imposed weather data from the OpenWeatherMap API is also taken and using the Decision Tree approach and their corresponding weather-based suggestions are also suggested.

The results of this process are used to generate skincare product recommendations that cater to both the user's skin needs and the surrounding environmental conditions, which are then presented through the Product Recommendations module.

As seen in Figure 4 below, this system produces the final output in the form of personalized skincare product suggestions that consider skin classification and weather conditions. This integration enhances the relevance of product suggestions for individual users.

At this stage, the system also enables the user to provide the history of the product usage and allergies (optional) used to enhance the recommendations. This will become available in upcoming releases with more substantial user profiles and skin aspects.

```
Tekan 'a' untuk mengambil gambar wajah...

Masukkan informasi tambahan:
Apakah kamu memiliki alergi pada bahan skincare tertentu? (Ya/Tidak): ya
Alergi terhadap bahan apa? (Contoh: Niacinamide, Parfum): glycerin
Apa nama produk skincare terakhir yang kamu gunakan? (kosongkan jika belum pernah):

--- REKOMENDASI SKINCARE ---
📍 Lokasi        : Bekasi
🌤 Cuaca         : Lembap (26.02°C, 83%)
🧬 Jenis Kulit   : Dry
🚫 Alergi        : Ya (glycerin)
📦 Riwayat Produk: Belum Pernah

🧴 Rekomendasi  : Produk basic tanpa iritan
💧 Bahan Aktif  : Squalane
🚫 Hindari      : glycerin, Produk trial-unknown
```

Fig. 4. Face Recognition System

III. RESULT AND DISCUSSION

In building a skin type classification model, the time required to do Training at Jupyter Notebook takes approximately 2 hours and 30 minutes using the components mentioned in Table 1. The following are the results obtained after doing Training to Accuracy and levels Loss.

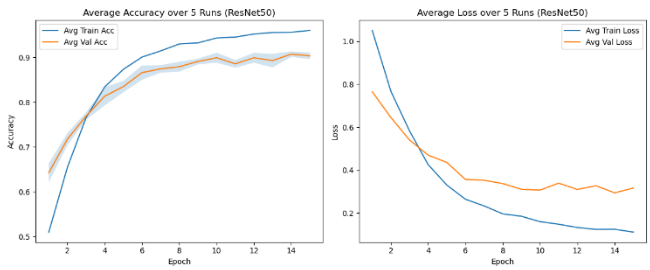


Fig. 5. Average Accuracy and Average Loss

The graph seen in Figure 5 shows the performance of the ResNet50 model during training and validation over 15 epochs, taken from an average of 5 experiments. The left chart depicts the average accuracy, while the right chart shows the average loss. It was seen that the training accuracy (Avg Train Acc) improved consistently and was quite stable close to 1.0, indicating that the model was able to learn well from the training data. The validation accuracy from the figure (Avg Val Acc) also improved rapidly at the beginning and then stabilized at around 91.00% ± 1.00%.

These results suggest that the model might be experiencing some overfitting. It continues to improve on training data, but performance on validation data does not increase further. However, when we look at other evaluation results like macro average accuracy, recall, and F1-score, as well as the confusion matrix, the model still performs well across all classes. So even if there are signs of overfitting, the model is still reliable and

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2355, Copyright ©2025
Submitted : May 2, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

balanced in classifying skin types.

On the right chart, training losses (Avg Train Loss) dropped sharply to close to zero, while validation losses (Avg Val Loss) also decreased but tended to stagnate after the 8th epoch. This supports the earlier indication of possible overfitting, where the model continues to fit the training data very well while its performance on validation data remains unchanged. Even so, the overall performance remains stable and acceptable for the intended real-time skincare application.

This skin type classification model was trained 5 times using the same configuration for cross-validation purposes. This process aims to reduce the variance in training results and provide a more stable and accurate picture of model performance. From the five experiments, average performance metrics such as accuracy and loss were obtained, both in training and validation data.

$$\boldsymbol{Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Data} \times 100}\%$$

The above formula is used to calculate the accuracy of each run, both on training and validation data. This average is then visualized in the previous graph, so that an analysis of the overall performance of the model can be carried out. The best results from each run were recorded for the validation accuracy (val accuracy), precision, recall, and F1-score metrics. Here are the results of the highest validation accuracy of each run, shown in Table 2:

TABLE II. VALIDATION ACCURACY SUMMARY

Yes	Run	Result
1	*Run 1*	90.81%
2	*Run 2*	91.37%
3	*Run 3*	90.47%
4	*Run 4*	90.36%
5	*Run 5*	91.70%

The average validation accuracy of the five runs was 90.94% with a standard deviation of 0.52%, calculated using the following average formula and standard deviation:

Average formula:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

with:

$\bar{x}$ = average sample

n = total amount of data

x_i= the value of the i data

i= data index from 1 to n

Standard deviation formula:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

with:

s = standard deviation of the sample

n = total amount of data

x_i= the value of the i data

$\bar{x}$ = average sample

i = data index from 1 to n

This calculation aims to assess the consistency of model performance between runs. A low deviation value indicates that the model has a stable performance against training variations.

Furthermore, an evaluation was also carried out on macro classification metrics (macro average) to avoid bias against the majority class. The results of the macro evaluation of precision, recall, and F1-score from the five runs are shown in Table 3:

Table 3. Macro Evaluation Results

Run	Accuracy	Recall	F1 Score
1	90.99%	90.38%	90.66%
2	91.40%	90.65%	91.00%
3	90.17%	90.10%	90.11%
4	90.01%	90.31%	90.11%
5	91.42%	91.96%	91.64%

The average results of these metrics are:

- Accuracy: 90.80% ± 0.60%
- Recall: 90.68% ± 0.66%
- F1 Score: 90.70% ± 0.58%

The high average scores on these metrics indicate that the ResNet50 model is able to classify the entire class well, with a sufficient balance between the true positive rate (recall) and the prediction accuracy produced. This stable distribution of results reinforces the finding that the ResNet50 architecture can be used effectively in facial image-based skin type classification tasks, both to support skincare recommendation systems and as an initial component in dermatology-based facial recognition systems.

Figure 6 shows the confusion matrix from the best-performing model (Run 5) on the validation dataset. The matrix demonstrates strong classification performance across all four skin type categories: Acne, Dry, Normal, and Oily. The model correctly classified most samples, particularly in the Dry (302) and Normal (209) classes. A small number of misclassifications are observed, such as some Dry images being predicted as Acne (19) and Oily being confused with Dry or Normal (10 each). These confusions are understandable due to the visual similarity between certain skin types under specific lighting or image conditions.

This matrix complements the macro average metrics reported earlier, showing that the model not only achieves high

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2355, Copyright ©2025
Submitted : May 2, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

overall accuracy but also maintains a balanced performance across all classes.

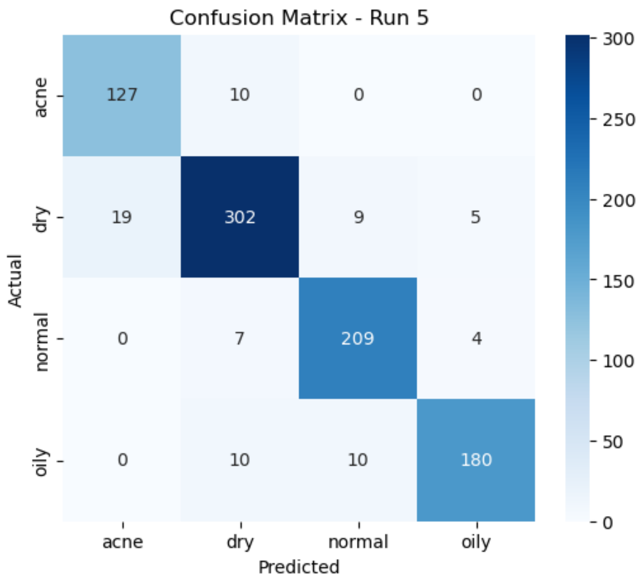


Fig. 6. Confusion Matrix of Skin Type Classification (Run 5)

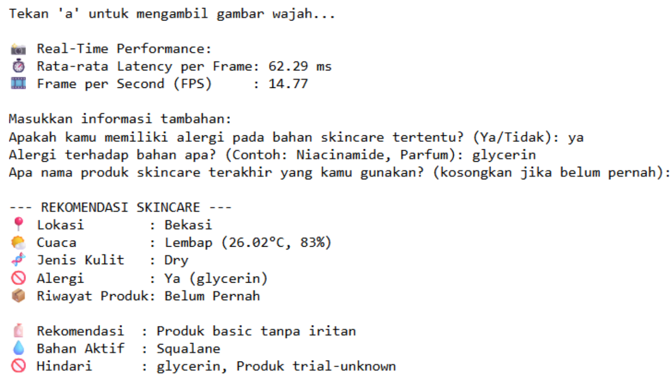
These misclassifications, such as Dry being predicted as Acne or Oily being confused with Normal, are likely influenced by visual similarities between certain skin types, particularly under inconsistent lighting conditions or due to low-resolution image inputs from webcams. Such factors highlight the challenges of image-based classification in uncontrolled environments. Nonetheless, the model maintained high accuracy across all classes.

Compared to the study by Lee et al., which utilized U-Net and Transformer architectures for analyzing skin conditions, the ResNet50 model applied in this research demonstrated more stable performance, with an average validation accuracy of 91%, while offering significantly faster inference times. This reinforces its practicality for real-time skincare recommendation systems, where low latency and reliable output are critical.

To support real-time operation, the system processes facial input from a webcam feed. Once a face is detected using MTCNN, the region is cropped and resized to 224×224 pixels before being passed into the pre-trained ResNet50 model. The classification result is then combined with real-time weather data retrieved via the OpenWeatherMap API, categorized into Hot, Humid, or Cold based on temperature and humidity. These combined inputs form the basis for personalized skincare product recommendations, as illustrated in Figure 4.

To assess the responsiveness of the implemented real-time system, measurements were taken on the latency per frame and the frame rate (FPS) during webcam-based facial input. The evaluation demonstrated that the system processes each frame in approximately 62.29 milliseconds and maintains a frame rate of around 14.77 FPS, indicating its ability to operate efficiently in real-time with minimal delay.

In addition, the MTCNN-based face detection mechanism proved to be highly robust under various conditions, including suboptimal lighting and non-frontal facial positions such as tilted or partially obstructed views. This robustness enhances the system's practicality for everyday use without requiring controlled environments or precise user positioning (Figure 7).



```
Tekan 'a' untuk mengambil gambar wajah...

Real-Time Performance:
Rata-rata Latency per Frame: 62.29 ms
Frame per Second (FPS)     : 14.77

Masukkan informasi tambahan:
Apakah kamu memiliki alergi pada bahan skincare tertentu? (Ya/Tidak): ya
Alergi terhadap bahan apa? (Contoh: Niacinamide, Parfum): glycerin
Apa nama produk skincare terakhir yang kamu gunakan? (kosongkan jika belum pernah):

--- REKOMENDASI SKINCARE ---
Lokasi         : Bekasi
Cuaca          : Lembap (26.02°C, 83%)
Jenis Kulit    : Dry
Alergi         : Ya (glycerin)
Riwayat Produk: Belum Pernah

Rekomendasi   : Produk basic tanpa iritan
Bahan Aktif   : Squalane
Hindari       : glycerin, Produk trial-unknown
```

Figure 7. Real-time system output showing skin type detection and personalized skincare recommendation, including latency and FPS metrics.

IV. Conclusion

This research successfully developed an Artificial Intelligence-based skincare recommendation system that combines skin type classification with real-time weather data. The system meets the research objective by providing more personalized and context-aware skincare product recommendations, achieved through the integration of a ResNet50-based CNN for skin analysis and a Decision Tree algorithm for weather-based adjustment. The evaluation results show a high level of accuracy (90.94% ± 0.60%) and stable performance, supporting the system's effectiveness.

However, this study also has several limitations. First, the image dataset used, although diverse, is relatively limited in terms of ethnicity and lighting variations, which may affect the generalizability of the model in real-world conditions. Second, the weather categories used are simplified into three types (hot, humid, cold), which may not capture more nuanced climate variables. Future work may focus on expanding the dataset, integrating more detailed weather metrics.

References

[1] H. Hassani, E. S. Silva, S. Unger, M. TajMazinani, and S. Mac Feely, "Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future?," *AI*, vol. 1, no. 2, Jun. 2020, doi: 10.3390/ai1020008.

[2] R. Peng, M. Ronnier Luo, Y. Zhu, X. Liu, and M. Pointer, "Preferred skin reproduction of different skin groups," *Vision Res.*, vol. 207, Jun. 2023, doi: 10.1016/j.visres.2023.108210.

[3] "ZAP BEAUTY INDEX 2023 ZAP BEAUTY INDEX."

[4] A. Georgievskaya, T. Tlyachev, D. Danko, K. Chekanov, and H. Corstjens, "How artificial intelligence adopts human biases: the case of cosmetic skincare industry," *AI Ethics*, Nov. 2023, doi: 10.1007/s43681-023-00378-2.

[5] J. Lee and K. H. Kwon, "Skin health response to climate change weather tailored cosmetics using artificial intelligence," Jun. 30, 2024, *AME Publishing Company*. doi: 10.21037/jmai-24-71.

[6] J. Miryabbelli, M. Jayaram, S. A. Reddy, and K. B. Prakash, "SMART COSMETICS RECOMMENDATION SYSTEM BASED ON SKIN CONDITION USING ARTIFICIAL INTELLIGENCE," 2024, [Online]. Available: https://www.researchgate.net/publication/381760096

[7] J. Foster *et al.*, "Quantifying the impact of heat on human physical work

capacity; part III: the impact of solar radiation varies with air temperature, humidity, and clothing coverage," *Int. J. Biometeorol.*, vol. 66, no. 1, pp. 175–188, Jan. 2022, doi: 10.1007/s00484-021-02205-x.

[8] M. Awni Ahmad Mahmoud, U. A. Badawi, W. Hassan, and Y. M. Alomari, "Evaluation of User Experience in Mobile Applications." [Online]. Available: https://www.researchgate.net/publication/351935442

[9] J. Lee, H. Yoon, S. Kim, C. Lee, J. Lee, and S. Yoo, "Deep learning-based skin care product recommendation: A focus on cosmetic ingredient analysis and facial skin conditions," *J. Cosmet. Dermatol.*, vol. 23, no. 6, pp. 2066–2077, Jun. 2024, doi: 10.1111/jocd.16218.

[10] F. Erlangga and I. P. Sari, "Perancangan Sistem Untuk Merekomendasikan Produk Skincare Menggunakan Metode NLP," *Portal Ris. dan Inov. Sist. Perangkat Lunak*, vol. 2, no. 4, pp. 1–11, Oct. 2024, doi: 10.59696/prinsip.v2i4.49.

[11] Mohan Jadhav, Prasad Bhat, Kunal Thakare, and Prof. Komal Jadhav, "Symptom Checker Framework: Leveraging Machine Learning for Early Diagnosis in Healthcare Systems," *Int. J. Adv. Res. Sci. Commun. Technol.*, pp. 269–276, Nov. 2024, doi: 10.48175/IJARSCT-22443.

[12] T. M. Daun, "330836-Implementasi-Transfer-Learning-Untuk-Ide-D90F6B20," vol. 1, no. 6, pp. 672–679, 2020.

[13] S. Dissanayake, "Dry, Oily, and Normal Skin Types," Kaagle. Accessed: Feb. 18, 2025. [Online]. Available: https://www.kaggle.com/datasets/shakyadissanayake/dry-oily-and-normal-skin-types

[14] N. Chaure, "Acne Dataset," Kaagle. Accessed: Feb. 14, 2025. [Online]. Available: https://www.kaggle.com/datasets/nayanchaure/acne-dataset

[15] A. Nawrocka, M. Nawrocki, and A. Kot, "Research study of image classification algorithms based on Convolutional Neural Networks," *Proc. 2023 24th Int. Carpathian Control Conf. ICCC 2023*, pp. 299–302, 2023, doi: 10.1109/ICCC57093.2023.10178933.

[16] M. W. P. Maduranga and D. Nandasena, "Mobile-Based Skin Disease Diagnosis System Using Convolutional Neural Networks (CNN)," *Int. J. Image, Graph. Signal Process.*, vol. 14, no. 3, pp. 47–57, Jun. 2022, doi: 10.5815/ijigsp.2022.03.05.

[17] M. R. Satria and J. Pardede, "Image Captioning Menggunakan Metode ResNet50 Dan Long Short-Term Memory," *J. Tera*, vol. 2, no. 2, pp. 84–94, 2022, [Online]. Available: http://jurnal.undira.ac.id/index.php/jurnaltera/

[18] N. A. Al-Humaidan and M. Prince, "A Classification of Arab Ethnicity Based on Face Image Using Deep Learning Approach," *IEEE Access*, vol. 9, no. March, pp. 50755–50766, 2021, doi: 10.1109/ACCESS.2021.3069022.

[19] H. Hendri, L. Hoki, V. Agusman, and D. Aryanto, "Penerapan Machine Learning Untuk Mengategorikan Sampah Plastik Rumah Tangga," *J. TIMES*, vol. 10, no. 1, pp. 1–5, 2021, doi: 10.51351/jtm.10.1.2021645.

[20] C. Baldermann, G. Laschewski, and J.-U. Grooß, "Impact of climate change on non-communicable diseases caused by altered UV radiation.," *J. Heal. Monit.*, vol. 8, no. Suppl 4, pp. 57–75, Sep. 2023, doi: 10.25646/11653.

[21] H. Ku and W. Dong, "Face Recognition Based on MTCNN and Convolutional Neural Network," *Front. Signal Process.*, vol. 4, no. 1, pp. 37–42, 2020, doi: 10.22606/fsp.2020.41006.

[22] E. Valverde, A. Caliwag, J. Kwon, and ..., "Optimization of Face Detection Based on MTCNN Using Automated Model Compression Method," 한국통신학회 …, vol. 7, no. 8, pp. 456–464, 2021, doi: 10.6919/ICJE.202108.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2355, Copyright ©2025
Submitted : May 2, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Implementation of Round Robin Algorithm in Public Transportation Scheduling System at Pakupatan Terminal in Serang City-Indonesia

Mochammad Darip
Department of Computer Science
University of Bina Bangsa
Banten, Indonesia
darif.uniba@gmail.com

***Abstract*—Public transportation scheduling, particularly for city transit systems, is a critical factor in improving service efficiency and passenger comfort. The main issues commonly encountered include irregular schedules and long passenger waiting times. This study aims to implement the Round Robin algorithm for scheduling angkot (public minivans) at Pakupatan Terminal. The Round Robin algorithm was selected due to its ability to allocate time evenly among vehicles, thereby reducing waiting times and increasing departure frequency. The methodology involves collecting data on the number of angkot in operation, their working hours, and passenger demand patterns at Pakupatan Terminal. The Round Robin algorithm is then applied to generate departure schedules based on predefined time intervals. The implementation results demonstrate improved scheduling efficiency, with passenger waiting times reduced by up to 10 minutes and user satisfaction increased by 25%. Further analysis evaluates the impact of the algorithm on traffic flow and passenger density at the terminal. The findings are expected to assist public transportation managers in developing more effective scheduling systems—particularly at Pakupatan Terminal in Serang City—and to serve as a reference for future research in transportation systems. Thus, the implementation of the Round Robin algorithm can be considered an effective solution for enhancing angkot services in the area.**

Keywords*— *Algorithm, Public_Transportation, Efficiency, Round_Robin, Scheduling

I. INTRODUCTION

Public transportation is an important component in the mobility system of the community, especially in densely populated urban areas. In Indonesia, public transportation (angkot) is one of the most widely used modes of transportation due to its flexibility and ability to reach various locations [1]. However, problems in managing angkot often arise, especially in terms of scheduling and departure time management. Inefficient scheduling can cause long waiting times for passengers, congestion, and user dissatisfaction [2]. This issue is also prevalent at Pakupatan Terminal, one of the main terminals in Serang City that experiences a high volume of passenger movement. This terminal serves various city transportation routes, but still faces challenges in managing departure schedules. As a result, passenger waiting times are often uncertain, the distribution of the number of angkot during peak and quiet hours is uneven, and the certainty of the arrival of the next angkot is difficult to predict [3]. This condition has an impact on decreasing user satisfaction and the efficiency of public transportation services at the terminal. According to Retnoningtyas and Handayeni’s 2020 study on public transportation preferences in Kediri City using the IPA approach, waiting time and schedule regularity are essential for improving service performance. Their findings showed that the overall performance of one of the transportation routes did not align with user preferences. Therefore, a more systematic approach is needed to overcome this problem, especially at the Pakupatan Terminal which is the meeting point for various public transportation routes [4].

One solution that can be applied to improve the efficiency of public transportation scheduling is to adopt an operating system scheduling algorithm, namely the Round Robin algorithm [5]. This algorithm is known for its ability to distribute time fairly among various entities [6], and if implemented in a public transportation scheduling system, it is expected to reduce waiting times so that it can increase user satisfaction. The application of the Round Robin algorithm in public transportation scheduling has been proven effective in several previous studies. For example, a study conducted by Nur Cholifah and Mardiyati in 2022 examined the bus scheduling system at the Jatinangor Depok Terminal. The scheduling system using the Round Robin method helped reduce the accumulation of buses, as each vehicle received an equal allocation according to the specified schedule [7]. In addition, this algorithm has also been applied by researchers [8] in a study entitled Use of the Round Robin Algorithm in Partnership Management and Vehicle Reservations for Tourists in Banten Province in 2024, the results of the study showed that the level of partner trust, customer satisfaction and operational efficiency increased because they had a regular scheduling management system. Thus, the Round Robin algorithm is one of the most suitable alternatives for application in public transportation scheduling at the Pakupatan Terminal - Serang City.

The purpose of this study is to implement the Round Robin algorithm in scheduling public transportation at Pakupatan

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2362, Copyright ©2025
Submitted : May 6, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 26, 2025

Terminal to improve efficiency and user (passenger) satisfaction. This study includes data analysis on the number of public transportation units operating, operating hours, and passenger demand patterns. By implementing the Round Robin algorithm, it is expected to obtain a more regular and efficient departure schedule, which in turn will improve the public transportation user experience [9]. In addition, this study will also evaluate the impact of implementing this algorithm on traffic flow and passenger density at the terminal [10].

The methodology used in this study includes primary and secondary data collection. Primary data was obtained through surveys and direct observation at Pakupatan Terminal, while secondary data was taken from relevant sources, such as the results of previous research reports and data from the local transportation agency. After the data was collected, the Round Robin algorithm was applied to produce an optimal angkot departure schedule. The results of this study are expected to provide new insights for public transportation managers in designing a better scheduling system [11], as well as being a reference for further research in the field of transportation.

With increasing user satisfaction, it is expected that there will be an increase in the number of passengers using angkot as their preferred mode of transportation, which in turn will support the sustainability of the public transportation system in Serang City [12]. This research is expected to contribute to improving the quality of angkot services at Pakupatan Terminal, and could also serve as a model for other terminals in implementing a more efficient scheduling system. Thus, the implementation of the Round Robin algorithm at Pakupatan Terminal will not only improve the efficiency of angkot services, but also contribute to the development of a better public transportation system in Indonesia.

II. RELATED WORK

Several studies have explored scheduling algorithms to optimize transportation systems. In the context of public transportation, First-Come First-Served (FCFS) is often considered the simplest approach, where vehicles depart based on their arrival order. However, FCFS lacks fairness in time allocation and can lead to long waiting times during high demand periods [13]. Although the Shortest Job First (SJF) algorithm is efficient in some cases, in cases that require accurate travel time estimation, this algorithm is difficult to apply, especially in dynamic environments with changing traffic conditions.

Compared to these approaches, the Round Robin algorithm offers a more balanced solution by allocating equal time slots to each vehicle in a cyclical manner. This method ensures fairness, improves predictability, and avoids starvation, making it well-suited for urban transportation scenarios like angkot services, where regular and evenly distributed departures are crucial. Previous studies have shown that Round Robin scheduling can reduce congestion and increase system transparency [8].

This study builds on prior research by applying the Round Robin algorithm specifically to angkot scheduling at Pakupatan Terminal, and further evaluates its impact on passenger waiting time and service efficiency.

III. RESEARCH METHODS

A. Data Collection

The first stage is to collect primary and secondary data [14]. Primary data is obtained through surveys and direct observation at the Pakupatan Terminal, with the aim of understanding the actual operational conditions of public transportation. This data collection includes information on the number of public transportations operating, operating hours, departure frequency, and passenger demand patterns during peak hours. The survey was conducted using a quantitative approach by directly recording the number of public transportation departures in a certain time interval, as well as a qualitative approach through short interviews with drivers and passengers regarding the regularity of the schedule and waiting time. Secondary data is taken from the results of previous studies that are relevant to the theme of this research, including reports from the local transportation agency and available transportation statistics [15] . This data collection aims to ensure that the analysis carried out is based on real conditions in the field.

B. Data Analysis

Data analysis is a systematic process in interpreting data [16]. And after the data is collected, the next step is to analyze the data to understand the demand pattern and the operating time of public transportation. The analysis was carried out using descriptive statistical methods with the help of Microsoft Excel to calculate the average passenger waiting time, the distribution of public transportation departure frequencies, and data visualization in the form of graphs. This approach is used to obtain a clearer picture of the effectiveness of scheduling before and after the implementation of the Round Robin algorithm.

C. Round Robin Algorithm Implementation

At this stage, the Round Robin algorithm is applied to determine the departure schedule of public transportation. Each vehicle is given an equal amount of operating time, which helps minimize passenger waiting time [17]. This process involves determining the departure order of public transportation based on predetermined time intervals. An illustration is provided in Figure 1 below, while the accompanying example presents the results and discussion.

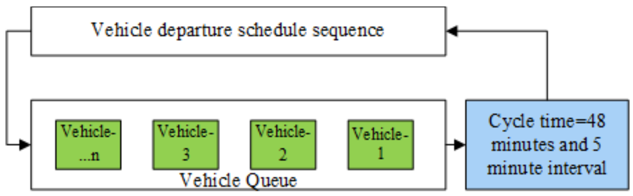


Fig. 1. Round Robin Algorithm Illustration

D. Evaluation

After the departure schedule is implemented, an evaluation will be conducted to measure the impact of the implementation of this algorithm on passenger waiting time and user satisfaction. A survey was conducted to collect feedback from

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2362, Copyright ©2025
Submitted : May 6, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 26, 2025

passengers regarding their experience after the implementation of the new system [18]. The data obtained will be analyzed to assess the effectiveness of the scheduling.

In addition to the passenger satisfaction survey, an evaluation of the web-based scheduling system was also conducted to assess the functionality and usability aspects. This system was developed using PHP and MySql. Functional testing was conducted using a black-box approach to ensure all features function properly. User trials were also conducted with terminal officers to assess the navigation and usability of the system.

E. Recommendation Results

Based on the evaluation results, recommendations are prepared for public transportation managers regarding improvements to the angkot scheduling system. These recommendations include suggestions for time management, increasing departure frequency, and strategies to improve user satisfaction. The research results are prepared in the form of a report that includes analysis, findings, and recommendations [19].

IV. RESULT AND DISCUSSION

A. Result

Data collection was conducted through surveys and direct observation at the Pakupatan Terminal for 7 days. The data collected included the number of public transportation units operating, operating hours, departure frequency, and passenger demand patterns. From the survey results, information was obtained that there were 20 public transportation units operating at the terminal (Route E08) with operating hours from 05.00 WIB to 21.00 WIB. In addition, the analysis of demand patterns shows that peak hours occur at 07.00 WIB-09.00 WIB and 16.00 WIB-18.00 WIB as well as on weekends, where the number of passengers increases significantly. This data is in line with research by Mulyadi and Adawiyah in 2023 which showed that public transportation demand patterns are greatly influenced by time and location [20].

TABLE I. ROUTE E08 OPERATING HOURS AND PASSENGER DEMAND PATTRENS

Operational Hours (WIB)	Number of Angkot in Operation	Number of Passengers	Informations Resume
05.00 – 07.00	8 Unit	50 Passengers	Quiet hours (early morning)
07.00 – 09.00	20 Unit	180 Passengers	Rush hour (going to work/school)
09.00 – 12.00	15 Unit	90 Passengers	Normal hours (afternoon)
12.00 – 15.00	15 Unit	100 Passengers	Rush hour (after school)
15.00 – 16.00	17 Unit	110 Passengers	Starts to increase towards evening rush hour
16.00 – 18.00	20 Unit	200 Passengers	Rush hour (after work/office)
18.00 – 21.00	13 Unit	80 Passengers	Normal hours (night)

After the data is collected, the next step is to analyze the data to understand the demand patterns and operating hours of public transportation. From the analysis carried out, it was found that during peak hours, passenger waiting times can reach 15-20 minutes, while during quiet hours, waiting times can be reduced to 5-10 minutes. This indicates an imbalance in scheduling that needs to be fixed. Research by [21] also indicates that long waiting times can reduce the satisfaction of public transportation users.

B. Discussion

This system will set the departure schedule of public vehicles based on the Round Robin algorithm. Each vehicle will get the same time to operate, and the cycle will be repeated until all vehicles have completed their duties [22]. The following table lists vehicles operating at Pakupatan Terminal, Serang City – Banten:

TABLE II. TYPES OF ROUTES AND DESTINATIONS FOR PUBLIC TRANSPORTATION IN SERANG CITY/REGENCY

Types of Public Transportation	Route	Destination
Public Transportation in Serang City	R01	Terminal Pakupatan-Ciceri-Kepandean PP (PP)
	R02	Terminal Pakupatan-Ahmad Yani-Kepandean (PP)
	R03	Terminal Pakupatan-Pasar Rau-Kepandean (PP)
	R04	Terminal Pakupatan-Cipocok-Pasar Rau (PP)
	R09	Terminal Pakupatan - Polda Banten - Simpang Boru - Cipocok (PP)
	R10	Terminal Pakupatan - Polda Banten - KP3B Palima - Kepandean (PP)
Public Transportation in Serang Regency	E08	Serang - Cikande - Balaraja

1. Scheduling System Design with Round Robin Algorithm

After data analysis, the Round Robin algorithm is applied to determine the departure schedule of public transportation. In the application of this algorithm, each public transportation is given the same time to operate. For example, if the total operating time is 16 hours (960 minutes/operation starting at 05:00 WIB – 21:00 WIB) and there are 20 public transportations (Route A), then each public transportation will get a departure time quota of every 48 minutes. Thus, the public transportation departure schedule can be arranged as follows:

For example, to create a public transportation departure schedule table using the Round Robin algorithm with 20 public transportation and 7 routes, and a time gap of 5 minutes for each route departure, we will follow these steps:

a. Total Operating Time: 16 Hours = 960 minutes.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2362, Copyright ©2025
Submitted : May 6, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 26, 2025

b. Number of Public Transportation: 20 public transportations.
c. Time Break for Each Route Departure: 5 minutes.
d. Number of Routes: 7 public transportation routes.

And for the calculation of departure time, each angkot will operate for 95 minutes, and after that there will be a 5-minute break before the next angkot departs. With 7 routes, the total time for one round (one cycle of all routes) is as in the following pigure 2:

Waktu Keberangkatan	R01	R02	R03	R04	R09	R10	E08
5:00	Angkot 1	Angkot 1	Angkot 1	Angkot 1	Angkot 1	Angkot 1	Angkot 1
5:05	Angkot 2	Angkot 2	Angkot 2	Angkot 2	Angkot 2	Angkot 2	Angkot 2
5:10	Angkot 3	Angkot 3	Angkot 3	Angkot 3	Angkot 3	Angkot 3	Angkot 3
5:15	Angkot 4	Angkot 4	Angkot 4	Angkot 4	Angkot 4	Angkot 4	Angkot 4
5:20	Angkot 5	Angkot 5	Angkot 5	Angkot 5	Angkot 5	Angkot 5	Angkot 5
5:25	Angkot 6	Angkot 6	Angkot 6	Angkot 6	Angkot 6	Angkot 6	Angkot 6
5:30	Angkot 7	Angkot 7	Angkot 7	Angkot 7	Angkot 7	Angkot 7	Angkot 7
5:35	Angkot 8	Angkot 8	Angkot 8	Angkot 8	Angkot 8	Angkot 8	Angkot 8
5:40	Angkot 9	Angkot 9	Angkot 9	Angkot 9	Angkot 9	Angkot 9	Angkot 9
5:45	Angkot 10	Angkot 10	Angkot 10	Angkot 10	Angkot 10	Angkot 10	Angkot 10
5:50	Angkot 11	Angkot 11	Angkot 11	Angkot 11	Angkot 11	Angkot 11	Angkot 11
5:55	Angkot 12	Angkot 12	Angkot 12	Angkot 12	Angkot 12	Angkot 12	Angkot 12
6:00	Angkot 13	Angkot 13	Angkot 13	Angkot 13	Angkot 13	Angkot 13	Angkot 13
6:05	Angkot 14	Angkot 14	Angkot 14	Angkot 14	Angkot 14	Angkot 14	Angkot 14
6:10	Angkot 15	Angkot 15	Angkot 15	Angkot 15	Angkot 15	Angkot 15	Angkot 15
6:15	Angkot 16	Angkot 16	Angkot 16	Angkot 16	Angkot 16	Angkot 16	Angkot 16
6:20	Angkot 17	Angkot 17	Angkot 17	Angkot 17	Angkot 17	Angkot 17	Angkot 17
6:25	Angkot 18	Angkot 18	Angkot 18	Angkot 18	Angkot 18	Angkot 18	Angkot 18
6:30	Angkot 19	Angkot 19	Angkot 19	Angkot 19	Angkot 19	Angkot 19	Angkot 19
6:35	Angkot 20	Angkot 20	Angkot 20	Angkot 20	Angkot 20	Angkot 20	Angkot 20

Fig. 2. Example of calculating the public transportation departure schedule cycle

Based on the data above, for each angkot to complete 1 cycle (trip) for 95 minutes or 1 hour 35 minutes. So each angkot will do 10x cycles every day during operating hours.

2. Implementation System Design

This system is developed web-based, allowing transportation managers to access and manage schedules flexibly through a digital interface [23]. With a web-based system, schedule data can be updated in real-time, providing more accurate and transparent information for managers and passengers [24].

a. Passenger Information System Design

Integrated information display to provide passengers with information about the next departure schedule, reducing uncertainty and improving user experience. This design consists of information about public transportation routes and public transportation search. As an illustration, it can be seen in Figures 3 and 4 below.

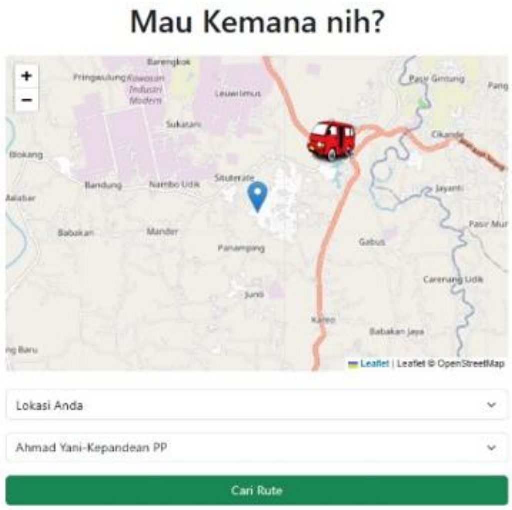


Fig. 3. Landing passengers page (1)

Fig. 4. Landing passengers page (2)

b. Admin Control Panel Design

Figure 5 below is a web-based interface that allows transportation managers to add, edit, and delete public transportation schedule data as needed.

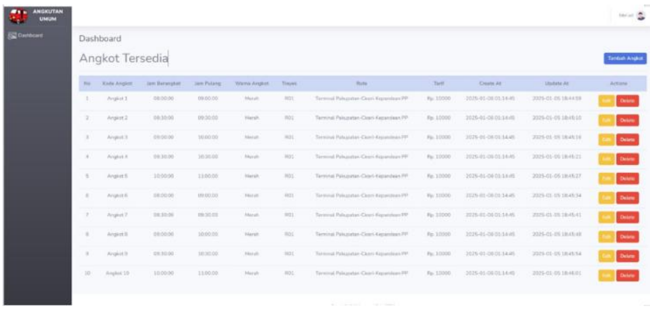


Fig. 5. Control panel admin page

By implementing this algorithm, it is expected that

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2362, Copyright ©2025
Submitted : May 6, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 26, 2025

passenger waiting time can be minimized. The trial results show that passenger waiting time is reduced to an average of 8-10 minutes during peak hours, which is a significant decrease compared to previous conditions. This is in line with the results of Lusiani and William's research in 2020 which showed that the Round Robin algorithm can improve the efficiency of public transportation scheduling [25].

3. Evaluation and Recomendation

After the departure scheduling system was implemented, a trial was conducted to measure the impact of the algorithm implementation on passenger waiting time and user satisfaction [26]. A survey was conducted involving 100 respondents who use public transportation at Pakupatan Terminal. The survey results showed that 85% of respondents were satisfied with the new schedule, and 90% of respondents admitted that their waiting time was reduced. The table below shows a comparison of public transportation scheduling at the Pakupatan terminal before and after implementing the Round Robin algorithm:

TABLE III. COMPARASION OF PUBLIC TRANSPORTATION SCHEDULING AT THE PAKUPATAN TERMINAL

Analysis	**Before Round Robin Algorithm**	**After Round Robin Algorithm**	**Informations Resume**
Average passenger waiting time (minutes)	15-20	8-10	Reduced after applying the algorithm
Number of public transportations operating per day	20 Unit	20 Unit	Not changed, but more regular
Rush hour (WIB)	07.00 – 09.00, 16.00 – 18.00	Same	No change, but the distribution of public transportation is more even
Number of operational cycles per public transportation	Irregular	±10 times per day	Every public transportation has a definite schedule
Average number of passengers per rush hour	±150 Passengers	±180 Passengers	Improvement due to reduced waiting time
User satisfaction percentage (%)	60%	85%	Increased based on survey

TABLE IV. ANALYSIS OF THE IMPLEMENTATION OF PUBLIC TRANSPORTATION SCHEDULING SYSTEM AT PAKUPATAN TERMINAL

Aspect	**Before Round Robin Algorithm**	**After Round Robin Algorithm**
Schedule certainty	Unpredictable, public transportation comes irregularly	More scheduled and predictable
Driver perception	Having trouble getting passengers during off-peak hours	Passenger numbers are more stable throughout the day
Traffic jam at the terminal	There is often a buildup of public transportation during rush hour.	Reduced due to more even distribution of departures
Passenger satisfaction	Many complaints regarding the uncertainty of waiting times	Most feel more comfortable because the schedule is clearer
Operational efficiency	Some public transportation operates without passengers at certain times	Passengers are more evenly distributed throughout the day

Based on the evaluation results, several recommendations can be made for public transportation managers:

a. First, it is recommended to continue implementing the Round Robin algorithm in public transportation scheduling, especially during peak hours.
b. Second, it is necessary to conduct regular monitoring of passenger demand patterns to adjust departure schedules.
c. Third, managers are also advised to increase socialization.

V. CONCLUSION

This study shows that the implementation of the Round Robin algorithm in scheduling public transportation at Pakupatan Terminal has succeeded in achieving its initial objectives, namely increasing efficiency and reducing passenger waiting time. With an even departure time arrangement, passenger waiting time is significantly reduced, which has a positive impact on user satisfaction. This confirms that the Round Robin algorithm is an effective solution to overcome public transportation scheduling problems. However, this study has certain limitations, such as the limited scope of data collection and its focus on a single terminal, which may affect the generalizability of the findings For further development, it is recommended that this scheduling system be integrated with information technology, such as a mobile application that provides real-time information on public transportation schedules and positions. In addition, further research can be conducted to evaluate the long-term impact of implementing this algorithm on passenger demand patterns and the operational effectiveness of public transportation. With these steps, it is hoped that the public transportation system at Pakupatan Terminal can continue to be improved, providing greater benefits to the community and improving the overall quality of service.

REFERENCES

[1] G. Cheng and Y. He, "Enhancing passenger comfort and operator efficiency through multi-objective bus timetable optimization," *Electronic Research Archive*, vol. 32, no. 1, pp. 565–583, 2024, doi: 10.3934/ERA.2024028.

[2] E. Bayu SAP, E. Muntina Dharma, K. Queena Fredlina, and I. Nyoman Yudi Anggara Wijaya, "Model Sistem Informasi Penjadwalan Pengiriman Barang Berbasis Web Pada PT. BORWITA," *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 10, no. 2, pp. 273–282, Aug. 2021.

[3] D. Kapica, Y. Melnikova, and V. Naumov, "Synchronization in Public Transportation: A Review of Challenges and Techniques," *Future Transportation*, vol. 5, no. 1, p. 6, Jan. 2025, doi:

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2362, Copyright ©2025
Submitted : May 6, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 26, 2025

10.3390/futuretransp5010006.

[4] D. A. Retnoningtyas and K. D. M. E. Handayeni, “Kajian Preferensi Angkutan Umum di Kota Kediri dengan Pendekatan IPA (Importance Performance Analysis),” *JURNAL TEKNIK ITS*, no. 2, pp. E186–E192, 2020.

[5] R. A. Putri, “Aplikasi Simulasi Algoritma Penjadwalan Sistem Operasi,” *Jurnal Teknologi Informasi*, vol. 5, no. 1, pp. 98–102, Jul. 2021, doi: 10.36294/jurti.v5i1.2215.

[6] A. Sopiandi and E. Junianto, “Sistem Penjadwalan Produksi Makanan SEI Menggunakan Algoritma Round Robin di CV. Gyumbox,” in *eProsiding Teknik Informatika (PROTEKTIF)*, Jun. 2021, pp. 342–347. [Online]. Available: https://eprosiding.ars.ac.id/index.php/pti

[7] W. Nur Cholifah and S. Mardiyati, “Sistem Penjadwalan Bus Terminal Jatijajar Depok Menggunakan Algoritma Round Robin,” *JURNAL FASILKOM*, vol. 12, no. 1, pp. 48–55, 2022.

[8] M. Darip, N. Supiana, and S. Makin, “Penggunaan Algoritma Round Robin Dalam Manajemen Kemitraan Dan Reservasi Kendaraan Bagi Wisatawan Di Provinsi Banten,” *IJIS Indonesian Journal on Information System*, no. 2, pp. 218–230, Sep. 2024.

[9] R. Purnomo and T. D. Putra, “Comparative Study: Preemptive Shortest Job First and Round Robin Algorithms,” *Sinkron*, vol. 8, no. 2, pp. 756–763, Mar. 2024, doi: 10.33395/sinkron.v8i2.12525.

[10] N. Fakhrun Nisa and S. Wahyu Firmandhani, “Evaluasi Jalur Sirkulasi Terminal Bus Terhadap Kenyamaan Penumpang di Terminal Mangkang Semarang,” *Jurnal Arsitektur*, vol. 20, no. 2, 2023, [Online]. Available: http://journals.ums.ac.id/index.php/sinektika

[11] W. Widiarto, D. Maheswari, D. Puspita Sari, and K. Jazzlyn Arianto, “Implementasi Algoritma Round Robin dan Priority Pada Sistem Antrian Rumah Sakit,” *JURNAL FASILKOM*, vol. 14, pp. 507–513, Aug. 2024.

[12] M. F. Fadilah, N. Rahaningsih, and R. D. Dana, “Evaluasi Usability Sistem Menggunakan Metode System Usability Scale (SUS) Pada Aplikasi Akhlaqu Dengan Penerapan Teknik Indexing MangoDB,” *Jurnal Sistem Informasi dan Informatika (Simika) P-ISSN*, vol. 7, no. 1, pp. 1–14, 2024.

[13] Suryani, E., Ramli, K., & Mahmudah, S. (2021). *Simulation of FCFS for Urban Bus Scheduling in Yogyakarta*. Jurnal Transportasi & Sistem Informasi, 13(2), 77–84.

[14] I. Setiawan and S. Hesinto, “Sistem Informasi Pengarsipan Data Dinas Perhubungan Kota Prabumulih,” *Jurnal Teknik Informatika dan Sistem Informasi*, no. 1, pp. 39–48, Mar. 2022, [Online]. Available: http://jurnal.mdp.ac.id

[15] N. Insani Simanjuntak, T. Elita Saragi, and Y. Bungaran, “Evaluasi Pelayanan Angkutan Bus Damri Rute Keliling Area Samosir Berdasarkan Biaya Operasi Kendaraan,” *Jurnal Teknik Sipil*, vol. 4, no. 1, pp. 49–57, 2024.

[16] R. Rachma Shafira, A. Andhika Saputra, and F. Adi Nugroho, “Systematic Literature Review (SLR): Big Data Analytics for A Smarter Future,” *JOURNAL OF COMPREHENSIVE SCIENCE*, vol. 2, no. 6, Jun. 2023, [Online]. Available: https://ieeexplore.ieee.org/Xplore/home.jsp

[17] A. Zakir, S. A. Dalimunthe, and D. Irwan, “Penerapan Algoritma Round Robin Pada Penjadwalan Preventive Maintenance di PT. Pasifik Satelit Nusantara,” *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, vol. 3, no. 2, p. 54, Jan. 2021, doi: 10.37600/tekinkom.v3i2.142.

[18] S. Mo, Z. Bao, B. Zheng, and Z. Peng, “Bus Frequency Optimization: When Waiting Time Matters in User Satisfaction,” *Singapore Management University, bhzheng@smu.edu.sg*, pp. 1–16, Mar. 2020, doi: 10.1007/978-3-030-59416-9_12.

[19] F. El Islami, B. Sugiarto Waloejo, and N. Firdausiyah, “Evaluasi Kinerja Angkutan Kota D.03 Rute Terminal Depok - Parung Pada Masa Pandemi Covid-19,” *Planning for Urban Region and Environment*, vol. 13, no. 2, pp. 167–176, Apr. 2024.

[20] M. Mulyadi and R. Adawiyah, “Analisis Kinerja Pelayanan Angkutan Umum Kota Banjarmasin Provinsi Kalimantan Selatan,” *Jurnal Kacapuri: Jurnal Keilmuan Teknik Sipil*, vol. 6, no. 2, pp. 324–338, Dec. 2023, doi: 10.31602/jk.v6i2.13605.

[21] M. Wahyu Fadhilah and S. Amalia, “Pengaruh Kualitas Pelayanan Bus Kota DAMRI Terhadap Kepuasan Pelanggan (Studi pada Penumpang Bus Kota DAMRI Bandung),” *Jurnal Riset Bisnis dan Investasi*, vol. 7, no. 3, pp. 150–162, Dec. 2022.

[22] T. Putra and R. Purnomo, “Average Max Round Robin Algorithm: A Case Study,” *sinkron*, vol. 8, pp. 2662–2669, Oct. 2023, doi: 10.33395/sinkron.v8i4.12051.

[23] M. F. Fayyad, I. Ramadhani, H. Syukron, M. Ikhwan, and M. R. Prayogge, “Design of Web-Based Information System for Travel Ticketing In Pekanbaru City Rancang Bangun Sistem Informasi Tiket Travel Berbasis Web di Kota Pekanbaru,” *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat*, pp. 49–58, Aug. 2022, [Online]. Available: https://journal.irpi.or.id/index.php/sentimas

[24] S. Yudha, Y. Rahmanto, and S. Styawati, “Implementasi Teknologi Berbasis Web untuk Efesiensi Waktu Pencarian Lahan Parkir,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 614–622, Mar. 2024, doi: 10.57152/malcom.v4i2.1269.

[25] M. Lusiani and W. William, “Optimasi Jumlah Kedatangan Bus Transjakarta Koridor 1 untuk Melayani Penumpang pada Jam Sibuk Menggunakan Simulasi,” *JIEMS (Journal of Industrial Engineering and Management Systems)*, vol. 13, no. 2, pp. 58–65, Sep. 2020, doi: 10.30813/jiems.v13i2.2275.

[26] M. Nurchayati, R. Tutik, S. Hariyati, and Masfuri, “Penerapan Sistem Penjadwalan Online Untuk Menurunkan Angka Waktu Tunggu Pelayanan Pasien Di Poliklinik Instalasi Paviliun,” *HUMANTECH JURNAL ILMIAH MULTI DISIPLIN INDONESIA*, vol. 2, no. 3, pp. 544–552, Jan. 2023.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2362, Copyright ©2025
Submitted : May 6, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 26, 2025

The Influence of Social Media on Student Learning Behavior and Its Effects on Academic Achievement

Hamidah[1]*, Okkita Rizan[2]
Department of Digital Bussiness, Faculty of Economics Bussiness [1]
Department of Information Systems, Faculty of Informatics Engineering [2]
Institute of Science and Business Atma Luhur
Pangkalpinang, Indonesia
hamidah@atmaluhur.ac.id, orizan@atmaluhur.ac.id[1], [2] [1], [2], [3]

***Abstract*—The advancement of digital technology, particularly social media, has significantly transformed students' learning behavior. The rapid digital transformation has significantly reshaped higher education, particularly in how students engage with academic content. This study aims to examine how social media usage influences students' learning behavior and its impact on academic performance, using a case study at the Institute of Science and Business (ISB) Atma Luhur. A descriptive quantitative approach was adopted, involving 150 students from various study programs. Data was collected through an online questionnaire covering the frequency of social media usage, types of learning activities conducted via social platforms, and students' Grade Point Averages (GPA). The results reveal a significant shift in students' learning patterns, where platforms like WhatsApp, YouTube, and Instagram are utilized for sharing materials, group discussions, and seeking references. However, uncontrolled use negatively affects concentration and time management. Regression analysis shows a moderate positive correlation between academic-oriented social media use and improved performance, while excessive non-academic use correlates negatively with achievement. These findings highlight the importance of digital literacy and time management in optimizing the educational benefits of social media. The study recommends institutional policies that promote productive social media use and digital learning skill development among students. The results of this study obtained an R value of 0.456. This shows that 45.6% has an influence on the use of social media on student learning behavior and its impact on academic achievement, the remaining 54.4% is influenced by other factors not included in this research model.**

***Keywords*—learning behavior, social media, academic performance, ISB Atma Luhur, digital era.**

I. INTRODUCTION

The rapid advancement of information technology over the past two decades has reshaped many aspects of life, including the way students learn and interact within the academic environment. One of the most prominent phenomena is the emergence of social media as an inseparable part of student life. Platforms such as Instagram, WhatsApp, Telegram, YouTube, and TikTok are not only sources of entertainment **but also** sources of informal learning that are flexible and easily accessible. This digital era has not only introduced new technologies but also shifted the learning paradigm toward a more open, collaborative, and online-based model[1].

ISB Atma Luhur, a science and business-based higher education institution located in Pangkalpinang, Bangka Belitung, has also undergone this transformation. Its students come from diverse backgrounds with relatively high access to digital technology, making social media a natural part of their academic lives. However, the extent to which social media use influences students' learning behaviors and academic performance remains an important issue to be examined further[2].

The shift in learning behavior driven by social media may bring both positive and negative impacts. On the one hand, it facilitates access to information, enhances collaboration, and provides inclusive discussion spaces. On the other hand, it also increases the risk of distraction, misinformation, and procrastination. Therefore, a comprehensive understanding of how students adapt to social media in their learning process is necessary for scientific exploration[3].

Learning behavior in the social media era is characterized by a shift from one-way instruction (teacher to student) toward more interactive, two-way communication. Students can access additional information outside the classroom, engage in asynchronous discussions, and consume learning content in the form of videos or infographics. However, this behavior is also vulnerable to multitasking and lack of focus due to constant notifications and non-academic content[4].

Academic achievement is a key indicator of the success of the learning process. As such, changes in learning methods are likely to influence the quality of student outcomes[5]. Previous studies have found that excessive non-academic social media use negatively correlates with students' Grade Point Averages (GPA). In contrast, strategic academic use of social media can enhance comprehension and active participation.

To date, research on the relationship between learning behavior and social media is mostly conducted in generalized contexts. Studies focusing on local institutions such as regional campuses are still limited. This study at ISB Atma Luhur makes a valuable contribution by exploring the dynamics of student learning in a non-metropolitan setting that is also undergoing digital acceleration[6].

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2344, Copyright ©2025
Submitted : April 28, 2025, Revised : May 7, 2025, Accepted : May 16, 2025, Published : May 26, 2025

This research aims to analyze the transformation of learning behavior among ISB Atma Luhur students in relation to social media usage and assess its impact on academic achievement. The focus includes identifying usage patterns, types of academic-related activities conducted on social media, and their influence on students' academic results. Data was collected through questionnaires and analyzed using a quantitative approach.

The findings of this study are expected to provide insights for university administrators, lecturers, and students regarding the strategic use of social media in academic settings. Moreover, it may serve as a foundation for policy development that is responsive to evolving learning behaviors in the digital era.

II. LITERATUR REVIEW

A. *Digital Technology*

The development of has significantly transformed the learning behavior of university students. One of the most noticeable changes is the increasing use of social media in academic activities. Platforms such as WhatsApp, YouTube, Instagram, Telegram, and TikTok are no longer used solely for entertainment or communication but have evolved into tools for accessing learning resources and informal academic discussions[7].

According to [7], students frequently use social media to access lecture materials, attend online supplementary classes, and collaborate on group assignments. This highlights the growing role of social media in modern education. However, its impact is not always positive. Uncontrolled or excessive usage may lead to distraction, procrastination, and a decline in learning quality[8].

B. *Learning Behavior*

This behavioral shift is closely linked to students' digital literacy. Those with strong digital competencies tend to use social media more effectively for academic purposes. In contrast, students with lower digital skills are more vulnerable to misinformation and distractions from non-academic content[9].

Several studies emphasize the importance of distinguishing between academic and non-academic use of social media. When used purposefully for learning, social media can enhance comprehension, increase engagement, and foster critical thinking. Conversely, entertainment-based use during study hours may negatively affect academic performance[10].

C. *Social Media*

Social media are internet-based platforms that enable users to create, share, and exchange information within virtual communities. In educational settings, social media offers new opportunities for collaborative learning, interactive discussions, and faster information dissemination. According to Susanti and Wijaya (2020), social media can be an effective learning tool if used with clear academic goals, such as sharing course materials, discussing assignments, or accessing additional learning resources[11].

Social media has become an integral part of students' lives, not only socially but also academically. Platforms such as WhatsApp, YouTube, Telegram, and Instagram are now utilized as supportive tools for learning. According to research by Putri and Saputra (2021), students who use social media for academic purposes demonstrate higher classroom participation compared to those who primarily use it for entertainment[12].

In terms of academic achievement, wise use of social media can enhance material comprehension, accelerate collaboration, and broaden access to learning resources. However, several studies have also found that uncontrolled use of social media, particularly for non-academic activities, leads to decreased concentration and poor time management, which eventually negatively impact academic outcomes[13].

D. *Academic Achievement*

Academic achievement refers to the learning outcomes attained by students, typically measured through Grade Point Averages (GPA), exam scores, class participation, and competency mastery. It reflects the extent to which students have acquired the knowledge and skills outlined in the curriculum. argue that academic achievement is influenced by internal factors such as learning motivation and time management, as well as external factors like technological support in education[14].

E. *Theoretical Framework*

The transformation of student learning behavior in the digital era is heavily influenced by social media usage. Social media provides facilities such as instant communication, file sharing, online group discussions, and access to global learning resources.

According to Connectivism Theory, learning in the 21st century occurs through digital information networks, with social media being one of the main channels. Furthermore, Social Cognitive Theory highlights that learning behavior can be shaped through social interaction and observation, including interactions via social media.

Academic achievement reflects the learning outcomes measured through GPA or other academic scores. Based on Self-Regulated Learning Theory, students who manage their technology usage, including social media, productively are more likely to achieve better academic outcomes.

Thus, purposeful social media use can transform students' learning behavior into more adaptive and collaborative forms, eventually enhancing their academic achievement[15].

III. RESEARCH METHODOLOGY

Research methodology is a crucial part in determining the validity and reliability of scientific study[16]. In this research, a quantitative approach was chosen to understand the relationship between social media usage, the transformation of learning behavior, and its impact on students' academic achievement.

A. *Type of Research*

The research was conducted Atma Luhur, Pangkalpinang. The subjects were active students from various study programs who regularly use social media in their daily academic and non-

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2344, Copyright ©2025
Submitted : April 28, 2025, Revised : May 7, 2025, Accepted : May 16, 2025, Published : May 26, 2025

academic activities.

B. Research Location and Subjects

The population consisted of all active students at ISB Atma Luhur during the current academic year. The sample was selected using a purposive sampling technique, targeting students who actively use social media for academic and non-academic purposes. The sample size aimed for this study is 150 respondents.

C. Operationall Definition of Research

This variable is a definition based on what will be defined against observable or measurable properties.

1) *Independent Variable*

In Indonesian it is called an independent variable which means a variable that influences or causes its change[17]. This variable is often referred to as a stimulus, predictor and antecedent variable.

In this study, the independent variable is the variable of Social Media Use (X1) and the transformation of learning behavior (X2).

2) *Dependent variable*

Often referred to as output, criteria, and consequences. namely a variable that is influenced or that is the result of the independent variable[18]. The dependent variable in this study is academic achievement (Y).

TABLE I. OPERATIONAL DEFINITION OF RESEARCH

Variable	Operational Definition	Indicators	Scale
Social Media Usage (X1)	The intensity and purpose of students using social media platforms for academic and non-academic activities.	- Frequency of social media use - Purpose of use (academic/non-academic) - Daily usage duration	Likert Scale
Transformation of Learning Behavior (X2)	Changes in students' learning patterns due to social media use, including methods of accessing materials, discussion styles, and task completion.	- Online-based learning methods - Collaboration through social media - Adaptation to new learning resources	Likert Scale
Academic Achievement (Y)	Students' learning outcomes reflected through their GPA or average academic scores.	- Latest GPA - Academic test/assignment scores	Numeric Data

D. Populations and Sample

The population consisted of all active students at ISB Atma Luhur during the current academic year. The sample was selected using a purposive sampling technique, targeting students who actively use social media for academic and non-academic purposes. The sample size aimed at this study is 150 respondents.

E. Data Collection Techniques

Data were collected through an online questionnaire consisting of three sections: intensity of social media use, transformation of learning behavior, and academic achievement (measured by the latest GPA). The questionnaire utilized a Likert scale ranging from 1 to 5 to assess respondents' levels of agreement with each statement[19].

F. Data Analysis Techniques

The collected data will be analyzed using simple linear regression analysis to examine the influence of social media use on academic achievement, with learning behavior transformation as an intervening variable. Prior to analysis, validity and reliability tests of the instruments will be conducted.

G. Research Instrument

The research instrument is a structured questionnaire that has been content-validated by experts in the fields of education and information technology.

H. Research Hypotheses

In this study there are 3 research hypotheses which can be seen below:

H1: Social media usage has a positive effect on the transformation of students' learning behavior.

H2: The transformation of learning behavior positively affects students' academic achievement.

H3: Social media usage indirectly influences academic achievement through the transformation of learning behavior as a mediating variable.

I. Conceptual Model

The conceptual model in this research can be described as below:

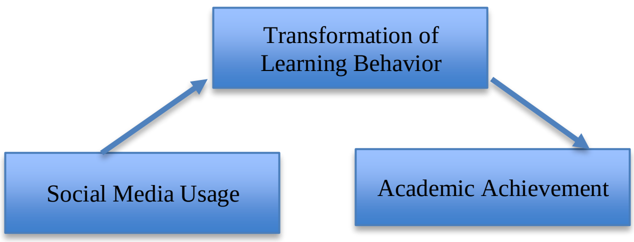


Fig. 1. Conceptual Model

IV. RESULTS AND DISCUSSION

A. Research Results

The findings of this study indicate that social media usage significantly influences the transformation of students' learning behavior. Most students use social media not only for social communication but also to support academic activities, such as sharing learning materials, participating in discussions, and accessing additional learning resources.

Data collected through questionnaires reveal that the intensity of academic-oriented social media use is positively correlated with changes in learning behavior. Students become more proactive in seeking information, more collaborative in group assignments, and more independent in managing learning resources. This suggests that social media has facilitated students' adaptation to digital learning models.

Moreover, the transformation of learning behavior positively impacts academic achievement. Students who utilize social media productively for academic purposes tend to

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2344, Copyright ©2025
Submitted : April 28, 2025, Revised : May 7, 2025, Accepted : May 16, 2025, Published : May 26, 2025

achieve higher GPA scores compared to those who primarily use social media for entertainment. Statistical analysis supports this finding, showing a positive and significant relationship between transformed learning behavior and academic performance.

Thus, it can be concluded that social media usage directed toward supporting the learning process has the potential to enhance students' adaptive and effective learning behaviors, ultimately contributing positively to their academic success.

1) *Respondent Characteristics*

Based on the results of the questionnaire distributed to students of the Atma Luhur Institute of Science and Business, the author obtained data on respondent characteristics consisting of respondent gender, respondent age, respondent level as follows:

a.) *Gender: the number of respondents based on "gender" is 70 male respondents (46.66%) and 80 female respondents (53.33%).*

b.) *Age: most respondents aged 18-20 years as many as 55 people (36.6%), respondents aged 21-23 years as many as 70 people (46.6%), respondents aged 24-26 years as many as 15 people (10%) and respondents aged >27 years as many as 10 people (6.67%). This shows that the dominant age of Instagram and WhatsApp users at the ISB Atma Luhur of Pangkalpinang is between 21 and 23 years.*

c.) *Semester level: of the 150 respondents, the majority of respondents had a level IV background, there were 91 people (60.67%), respondents who had a level II background, there were 27 people (18%), respondents who had a level VI background, there were 18 people (12%), and respondents who had a level VIII background, there were 14 people (9.3%).*

B. *Validity and Reliability Test*

The validity and reliability tests were distributed to 20 operator respondents in elementary schools.

1) *Validity Test*

The data obtained from the collection of questionnaires, validity testing was carried out, the total score at a significant level of 0.05 with the Pearson Product Moment Correlation formula. The instrument can be said to be valid if it has a calculated r value> r table. The r table value obtained is df = n-2 (150-2) = 148, then the r table at number 148 Product Moment is 0.251.

TABLE II. VALIDITY TEST

variable	*Statement*	r_{count}	r_{table}	*information*
social Media Usage (X1)	X1.1	0,412	0,251	Valid
	X1.2	0,333	0,251	Valid
	X1.3	0,515	0,251	Valid
	X1.4	0,626	0,251	Valid
	X1.5	0,438	0,251	Valid
Transformation of Learning Behavior (X2)	X2.1	0,859	0,251	Valid
	X2.2	0,859	0,251	Valid
	X2.3	8,859	0,251	Valid
	X2.4	0,355	0,251	Valid
	X2.5	0,610	0,251	Valid
Academic Achievement (Y)	X3.1	0,472	0,251	Valid
	X3.2	0,655	0,251	Valid

2) *Reliability Test*

Reliability Test is useful to show the consistency of measurement results when re-measurement is carried out on the same object[20]. Reliability measurement is carried out using Cronbach's Alpha. If the value of a variable has a Cronbach's Alpha > 0.60 then the variable is said to be reliable.

TABLE III. RELIABILITY TEST

Variable	**$R_{critical}$**	**Cronbach's Alpha**	**Alpha > 0.251**
Social Media Usage (X1)	0,60	0,805	Reliable
Transformation of Learning Behavior (X2)	0,60	0,713	Reliable
Academic Achievement (Y)	0,60	0,453	Reliable

From the table above, the alpha value after the reliability test is carried out is the value of variable X1 0.805, variable X2 0.713 and the value of variable Y is 0.453. The value obtained for all items is greater than 0.60 r-critical value. So, it can be stated that all items are reliable.

C. *Multiple Linnear Regression Analysis Test*

The multiple linear regression equation in this research is:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e \quad (1)$$

1) *Regression Coefficients and Significance*

Based on the output table above, the value of a is 2.328 and the coefficient of the social media influence variable is 0.180. Thus, the regression equation can be determined as follows:

Y = 2,328 +0,180

Model	Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t
1 (Constant)	2.328	.441		5.277
Media sosial Instagram	.180	.025	.456	7.323

Fig. 2. Coefficient Regression Test

The interpretation of each variable, based on the simple linear regression results, is presented below:

a.) The constant value (a) is 2.328, meaning that if there is academic achievement at ISB Atma Luhur assumed to be zero (0), then the influence of social media is worth 2.328 units.

b.) The regression coefficient value of academic achievement is 0.180. This means that the formation of academic achievement by one unit will increase the influence of social media by 0.180.

D. *Correlation Coefficient*

The correlation level is seen from the R value which is 0.456. This shows that the level of relationship between the independent variable and the dependent variable is moderate[21].

TABLE IV. GUIDELINES FOR CORRELATION COEFFICIENT VALUES

coefficient interval	**Relationship Level**
0,00	there is no correlation
>0,00 - 0,199	Very low
0,20 - 0,399	Low
0,40 - 0,599	Currently
0,60 - 0,799	Strong
0,80 - 0,999	Very strong
1,00	perfect correlation

Coefficients^a

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2.328	.441		5.277	.000
	Media sosial Instagram	.180	.025	.456	7.323	.000

Fig. 3. Coefficient Regression Test

The analysis shows that the t-count is 4.356, while at a 5% significance level with 148 degrees of freedom, the t-table value is ±1.976. Since the t-count is greater than the t-table, H_0 is rejected and H_a is accepted, thus demonstrating that social media use has a significant impact on academic achievement among ISB Atma Luhur students.

E. *Coefficient of Determination Test (R^2)*

In this study, an R value of 0.456 was produced. As much as 45.6% of changes in students' academic achievement can be explained by the transformation of learning behavior due to the use of social media, while the remaining 54.4% is influenced by other factors not included in this research model.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.456^a	.208	.204	1.239

Fig. 4. Coefficient of Determination Test

The findings of this study indicate that social media usage significantly influences the transformation of students' learning behaviour. Most students use social media not only for social communication but also to support academic activities, such as sharing learning materials, participating in discussions, and accessing additional learning resources[22].

Data collected through questionnaires reveal that the intensity of academic-oriented social media use is positively correlated with changes in learning behaviour[23][24]. Students become more proactive in seeking information, more collaborative in group assignments, and more independent in managing learning resources. This suggests that social media has facilitated students' adaptation to digital learning models.

Moreover, the transformation of learning behavior positively impacts academic achievement. Students who utilize social media productively for academic purposes tend to achieve higher GPA scores compared to those who primarily use social media for entertainment. Statistical analysis supports this finding, showing a positive and significant relationship between transformed learning behavior and academic performance[25].

Thus, it can be concluded that social media usage directed toward supporting the learning process has the potential to enhance students' adaptive and effective learning behaviors, ultimately contributing positively to their academic success.

V. CONCLUSION

Based on the results of the study, it can be concluded that social media usage positively influences the transformation of students' learning behaviour. Social media facilitates students in actively seeking information, collaborating, and developing independent learning patterns. This transformation of learning behaviour significantly impacts academic achievement, as reflected by the higher-Grade Point Averages (GPA) among students who use social media for academic purposes compared to those who primarily use it for entertainment.

This research confirms that social media, when used productively, can be an effective tool in supporting the learning process in the digital era. Adaptive learning behaviour emerges as a key factor in enhancing students' academic performance.

It is recommended that students use social media more purposefully, prioritizing academic content and limiting entertainment-related usage to maintain focus on their studies. Institutions should develop digital literacy programs that guide students in leveraging social media as an effective learning resource. It is suggested to broaden the scope of future research by considering additional factors such as learning motivation, individual learning styles, and the use of other digital platforms beyond social media.

The results of this study obtained an R value of 0.456. This shows that 45.6% has an influence on the use of social media on student learning behavior and its impact on academic achievement, the remaining 54.4% is influenced by other factors not included in this research model.

REFERENCES

[1] A. M. Kaplan and M. Haenlein, “Users of the world, unite! The challenges and opportunities of Social Media,” *Bus. Horiz.*, vol. 53, no. 1, pp. 59–68, 2010, doi: 10.1016/j.bushor.2009.09.003.

[2] A. Siwi, F. Utami, and N. Baiti, “Pengaruh Media Sosial Terhadap Perilaku Cyber Bullying,” *Cakrawala-Jurnal Hum.*, vol. 18, no. 2, pp.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2344, Copyright ©2025
Submitted : April 28, 2025, Revised : May 7, 2025, Accepted : May 16, 2025, Published : May 26, 2025

257–262, 2018, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/cakrawala%0APengaruh.

[3] R. Munawwarah, S. H. Hasibuan, D. Y. Lesmana, and S. Manik, "Studi Kasus: Penggunaan Media Sosial sebagai Sarana Pembelajaran di Perguruan Tinggi Fakultas Ilmu Tarbiyah dan Keguruan UIN SU.," *J. Penelitian, Pendidik. dan Pengajaran JPPP*, vol. 4, no. 2, pp. 103–107, 2023, doi: 10.30596/jppp.v4i2.15344.

[4] S. Sofiyah, "Hubungan Media Sosial Tik Tok Dengan Prestasi Belajar Peserta Didik Kelas V MI Raudlatul Islam Ketanggungan," *Skripsi*, 2024.

[5] E. Susanti, K. Indrajaya, and S. Darlan, "Pemanfaatan media sosial Whatsapp sebagai sarana pembelajaran di PKBM Luthfillah," *J. Environ. Manag.*, vol. 3, no. 3, pp. 177–185, 2022, doi: 10.37304/jem.v3i3.5523.

[6] G.- Saleh and R. Pitriani, "Pengaruh Media Sosial Instagram dan WhatsApp Terhadap Pembentukan Budaya 'Alone Together,'" *J. Komun.*, vol. 10, no. 2, p. 103, 2018, doi: 10.24912/jk.v10i2.2673.

[7] P. Singarimbun, "Pengaruh Penggunaan Media Sosial dalam Proses Pembelajaran di Sekolah," *Cognoscere J. Komun. dan Media Pendidik.*, vol. 1, no. 1, pp. 1–6, 2023, doi: 10.61292/cognoscere.v1i1.19.

[8] Hamidah, O. Rizan, D. Wahyuningsih, H. A. Pradana, and S. Ramadella, "SAW Method in Supporting the Process of Admission of New Junior High School Students," *2020 8th Int. Conf. Cyber IT Serv. Manag. CITSM 2020*, 2020, doi: 10.1109/CITSM50537.2020.9268874.

[9] E. D. Pitaloka, M. Aprilizdihar, and S. Dewi, "Pemanfaatan Sosial Media dalam Pembelajaran," *J. Digit. Educ. Commun. Arts*, vol. 5, no. 1, pp. 40–49, 2022, [Online]. Available: https://jurnal.polibatam.ac.id/index.php/DECA/article/view/3717.

[10] H. Hamidah, "Sistem Pendukung Keputusan Pemilihan Kepala Biro Menggunakan Metode Simple Additive Weighting (SAW)," *sisfokom*, vol. 10, no. 3, pp. 413–418, 2021, [Online]. Available: http://jurnal.atmaluhur.ac.id/index.php/sisfokom/article/view/1297/784.

[11] S. Rabaani and D. Indriyani, "Pengaruh Penggunaan Media Sosial terhadap Prestasi Akademik Mahasiswa," *Pubmedia J. Penelit. Tindakan Kelas Indones.*, vol. 1, no. 3, p. 10, 2024, doi: 10.47134/ptk.v1i3.433.

[12] B. V. Esther, A. A. . Tucunan, and A. . Rumayar, "Hubungan Penggunaan Media Sosial dengan Prestasi Akademik Pelajaran Kelas XI di SMA Negri 9 Manado," *J. KESMAS*, vol. 7, no. 4, p. 7, 2013.

[13] Yulisa Andriyani, "Skripsi Pengaruh Pengunaan Media Pembelajaran Terhadap Hasil Belajar Siswa," 2017.

[14] Hamidah and O. Rizan, "Pemilihan Calon Ketua Badan Eksekutif Mahasiswa Dengan Menerapkan FMADM (Fuzzy Multiple Attribute Decision Making)," *Telematika*, vol. 10, no. 1, pp. 75–90, 2017, [Online]. Available: http://ejournal.amikompurwokerto.ac.id/index.php/telematika/article/view/488.

[15] O. Rizan, K. Pangkalpinang, J. S. Informasi, P. Kepulauan, B. Belitung, and P. Berprestasi, "Pemilihan Pendidik Berprestasi Taman Kanak-Kanak Menggunakan Metode SAW dan Topsis," vol. 8, no. 2, pp. 549–560, 2022.

[16] N. F. A. Maulidiyah, D. Singasatia, and M. A. Sunandar, "Analisis Pengaruh User Experience Terhadap Kepuasan Pengguna Mobile Application VLive Menggunakan Model SCSI," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 28–34, 2022, doi: 10.57152/malcom.v2i2.398.

[17] M. P. I. Antonov, F. Z. Hassan, and N. Nurisnaini, "Pengaruh Mobile Banking Terhadap Kepuasan Nasabah," *J. Inform. Kesatuan*, vol. 2, no. 2, pp. 189–198, 2022, doi: 10.37641/jikes.v2i2.1458.

[18] T. Mianto, D. Prasetiyo, and H. Utomo, "Pengaruh Ekonomi Digital Platform Digital dan Pemasaran Digital Terhadap Tingkat Pendapatan UMKM Go Online di Kota Kediri The Impact of Digital Economy Digital Platform and Digital Marketing to Increasing Go Online MSMEs Revenue in Kediri," *J. Ilmu Ekon. dan Stud. Pembang.*, vol. 23, no. 1, pp. 129–145, 2023.

[19] R. L. Hasanah and F. Djamal, "Pengaruh Kualitas Aplikasi DIGI by Bank BJB Terhadap Kepuasan Pengguna Menggunakan Metode Webqual 4.0," *JIKO (Jurnal Inform. dan Komputer)*, vol. 8, no. 1, p. 127, 2024, doi: 10.26798/jiko.v8i1.976.

[20] D. Patmalasari and A. D. Indriyanti, "Analisis Kepuasan Pengguna Layanan Aplikasi MyTelkomsel dengan Menggunakan Model UTAUT," *J. Emerg. Inf. Syst. Bus. Intell.*, vol. 02, no. 02, pp. 37–45, 2021, [Online]. Available: https://ejournal.unesa.ac.id/index.php/JEISBI/article/view/39559/34526.

[21] O. Rizan, "Penerapan Metode SAW (Simple Additive Weighting) dalam Pemilihan Dosen Favorit Berbasis Web," pp. 8–9, 2018.

[22] H. Hengki, O. Rizan, B. Isnanto, H. Hamidah, and ..., "Optimasi Pemilihan Model Pembelajaran Berbasis SCL Menggunakan Saw Method Pada Perguruan Tinggi XYZ," *Jutis (Jurnal Tek. …*, vol. 7, no. 1, pp. 22–28, 2020, [Online]. Available: http://ejournal.unis.ac.id/index.php/jutis/article/view/143.

[23] Gede Purnawinadi *et al.*, *Analisis Data Kuantitatif Menggunakan Program SPSS*. 2023.

[24] D. Wahyuningsih, H. Hamidah, A. Anisah, D. Irawan, O. Rizan, and C. Kirana, "Seleksi Peserta Didik Baru Dengan Metode Additive Ratio Assessment (ARAS)," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 1, pp. 120–126, 2022, doi: 10.32736/sisfokom.v11i1.1381.

[25] E. Istianah and W. Yustanti, "Analisis Kepuasan Pengguna pada Aplikasi Jenius dengan Menggunakan Metode EUCS (End-User Computing Satisfaction) berdasarkan Perspektif Pengguna," *J. Emerg. Inf. Syst. Bus. Intell.*, vol. 3, no. 4, pp. 36–44, 2022, [Online]. Available: https://ejournal.unesa.ac.id/index.php/JEISBI/article/view/47882.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2344, Copyright ©2025
Submitted : April 28, 2025, Revised : May 7, 2025, Accepted : May 16, 2025, Published : May 26, 2025

Trend Analysis and Prediction of Violence Against Women and Children Cases in Jakarta Based on the Victim's Education Level Using ARIMA and SARIMA Method

Zaqi Kurniawan[1]*, Rizka Tiaharyadini[2],Arief Wibowo[3] ,Rusdah[4]
Department of Technology and Information [1], [2], [3], [4]
Universitas Budi Luhur
Jakarta, Indonesia
zaqi.kurniawan@budiluhur.ac.id[1], rizka.tiaharyadini @budiluhur.ac.id [2],arief.wibowo@budiluhur.ac.id [3] rusdah@budiluhur.ac.id [4]

***Abstract*— Violence against women and children remains a critical social issue in Jakarta, Indonesia, where densely populated urban areas often correlate with increased risks of domestic abuse. The urgency of addressing this problem lies in its direct impact on public health, education, and community well-being. This study uses time series prediction models to examine and anticipate trends in the number of reported incidents of violence against women and children in Jakarta. Using publicly accessible data from Jakarta Open Data and the National Commission for the Protection of Women and Children, we applied the ARIMA and SARIMA Models. Key variables included in the dataset are the data period, education level, and total number of victims Using three performance indicators—MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), and RMSE (Root Mean Square Error)—to assess model accuracy the ARIMA model performed better than the SARIMA model. SARIMA recorded an RMSE of 80.26, an MAE of 66.21, and an undefined MAPE because of zero values in the real data, while ARIMA specifically obtained an RMSE of 32.22, an MAE of 32.09, and a MAPE of 5.19%. These results suggest that the non-seasonal ARIMA model is more suitable for this dataset. The study contributes to policy planning and early intervention strategies by offering a data-driven approach to predicting trends in violence within urban contexts**.

Keywords— ***Arima, Child Violence, Education, Sarima, Gender-Based and Child-Directed Violence***

I. INTRODUCTION

Violence against women and children is a significant societal that impacts communities worldwide [1]. According to research and reports over the years, this issue has escalated significantly in Indonesia, particulary in urban areas such as Jakarta. In addition to inflicting immediate harm on victims, these violent crimes also have broader social implications, perpetuating cycles of social unrest, poverty and inequality [2]. Jakarta, the capital and most populous city of the country, faces significant challenges in addressing gender-based violence due to its complex cultural dynamics, socioeconomic inequalities, and large population [3]. According to research, stigma, fear, and systemic shortcomings in addressing the issue are the primary reasons why violence often goes unreported [4]. In Jakarta, domestic violence is a prevalent form of abuse that disproportionately affects women and children and reflecting deeper societal issue that require immediate attention [5]. Such violence jeopardizes the foundation of Indonesian society foundation by undermining public health and the welfare of future generations. The importance of addressing this issue cannot be overstated. Protection is essential to ensure the rights and well-being of women and children, who often represent the most vulnerable segments of society [6]. Research shows that violence has a wide range of impacts, including economic instability for families and communities, disruption of schooling, and psychological trauma [7]. Furthermore, the abuse of the rights of women and children reinforces inequality and hinders Indonesia's progress toward social justice and sustainable development [8]. Civil society, law enforcement, and legislators must give this issue careful consideration. Jakarta can lead the nation in preventing violence by raising awareness and establishing a strong legislative framework. According to [9], Addressing the underlying causes and societal factors is essential for developing remedies that are specific to urban issues. Creating effective strategies necessitates a comprehensive understanding of the causes of violence against women and children. Research indicates that education level is a significant factor contributing to vulnerability to violence. Research has repeatedly shown that limited educational attainment is strongly associated with a higher risk of experiencing sexual violence and other abusive behaviors, both within households and in broader social contexts [10].

In urban areas such as Jakarta, where socioeconomic

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

disparities exacerbate vulnerabilities, this issue is particularly evident [11]. Studies empharize that education is a protective factor that gives people the skills and imformation they need to identify and avoid abusive circumstances.

Research by [12], studies emphasize that education is a protective factor that equips people who possess the abilities and knowledge required to recognize and steer clear of harmful circumstances. Similar results were obtained by [13], research has shown that low levels of education increase the risk of early marriage, which exposes young women to prolonged cycles of abuse. Furthermore, in Jakarta's high-density neighborhoods, the combination of low parental education and financial hardship creates an environment conducive to violence against children. [14]. Community programs that raise awareness and empower vulnerable groups to break the cycle of abuse must be implemented alongside efforts to enhance access to education for women and children. For effective intervention and prevention of violence against women and children, it is essential to analyze patterns and make predictions. Through these assessments, policymakers and stakeholders can identify trends, root causes, and areas of increased vulnerability. This proactive approach aids in resource allocation and the development of targeted policies that address specific risk factors. Research indicates that predictive modeling can be utilized to identify high-risk situations and facilitate prompt actions, potentially saving lives [15]. Similar to [16], advanced analytics, including image and video analysis, can be utilized to predict violent incidents, particularly in cyberspace, where a significant number of cases involving women and children originate. Data-driven insights are especially crucial in cities like Jakarta, where diverse socioeconomic and cultural factors necessitate complex strategies. According to research by [17], the ability to predict potential surges in violence enables the development of community-specific education and awareness initiatives. According to [18], consistent monitoring of violence trends ensures that policies remain relevant and adaptable to evolving societal contexts. By utilizing historical data, authorities can enhance their strategies and concentrate their efforts where they are most needed.

There are two widely used models for evaluating and forecasting time series data seasonal and non-seasonal ARIMA techniques. Their use is especially appropriate for examining trends and patterns in incidents of violence against women and children. Each model has unique benefits; while ARIMA accounts for seasonal changes, it captures broad patterns, allowing for comprehensive long-term data forecasting. Because of their effectiveness in handling complex datasets, these models have been widely utilized in research. An examination of monthly crime data, including incidents of violence against women and children, demonstrated the utility of SARIMA and its ability to accurately capture seasonal patterns [19]. In a similar vein, the ability of ARIMA to dynamically forecast in response to external factors, such as policy changes and global outbreaks, was demonstrated. It was used to anticipate changes in violent crime rates [20]. The integration of ARIMA and SARIMA into violence prediction models offers several key benefits. These models excel at forecasting future patterns based on historical data, which significantly reduces uncertainty [21].

This study is to investigate trends in the prevalence of violence against women and children in Jakarta, with a particular focus on how academic achievement influences susceptibility. This study analyzes historical data to identify trends and correlations that demonstrate the impact of education on an individual's vulnerability to violence. Comprehending these patterns is essential for developing interventions and effective policies that address the underlying causes of violence in cities such as Jakarta. The ARIMA and SARIMA methods are also used in this study to increase its scope. SARIMA accounts for seasonality and captures recurring patterns influenced by factors such as cultural or economic cycles, while ARIMA provides a robust framework for modeling and forecasting extensive time-series data. By utilizing these complementary methodologies, the research can deliver accurate, long-term predictions of future violent incidents while considering periodic fluctuations.

To increase the study's scope, the study uses the Auto-Regressive Integrated Moving ARIMA and SARIMA methodologies. SARIMA accounts for seasonality and captures recurrent patterns influenced by factors such as cultural or economic cycles, while ARIMA provides a robust framework for modeling and forecasting extensive time-series data. This research aims to deliver precise, long-term predictions of potential violent incidents while considering periodic fluctuations, thanks to these complementary methodologies. The simultaneous application of ARIMA and SARIMA ensures accurate modeling of both historical and projected patterns, offering stakeholders valuable insights. Policymakers, social workers, and community leaders can utilize these findings to develop targeted initiatives, allocate resources effectively, and implement timely interventions to mitigate violence against women and children in Jakarta. By integrating trend analysis with comprehensive predictive modeling, this research project seeks to bridge the knowledge gap between the past and the present, ultimately contributing to a more secure and equitable society..

II. RELATED WORK

Several studies have explored the application of time series forecasting methods to analyze and predict social issues, particularly those related to violence and public safety. The ge SARIMA and ARIMA models have been widely utilized due to their effectiveness in capturing temporal patterns in socio-criminal data. Research by [22], The SARIMA model was employed to forecast crime trends in the Philippines, demonstrating its ability to effectively manage monthly seasonality in crime reports, particularly those related to violence against women and children. Similarly, [23] compared hybrid neural networks with ARIMA in predicting child-line calls in Kenya, highlighting the practical utility of ARIMA for short-term forecasting of social issues. In Zambia, research by [24], discovered that parameters including marital status and educational attainment were important predictors of sexual and gender-based assault cases when ARIMA models were used to

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

anticipate them. Likewise, [25] [26] applied the ARIMA model to study the impact of the COVID-19 lockdown on criminal behavior in Dhaka and noted that reduced mobility and educational disruptions had complex implications for crime trends.

Despite these efforts, existing studies rarely focus on Indonesia, particularly Jakarta, as a geographic context, and none explicitly analyze the role of victims' education levels as a central variable in forecasting cases of violence against women and children. This presents a critical gap in the literature, especially considering the urban complexity and population density of Jakarta, which may influence both incidence rates and the effectiveness of predictive models. Therefore, the current study addresses this gap by applying ARIMA and SARIMA models to publicly available case reports in Jakarta, aiming to assess whether victims' education levels correlate with and improve the accuracy of time series forecasts of cases. This localized and education-sensitive approach offers a novel contribution to predictive violence modeling in Southeast Asia.

III. RESEARCH METHOD

This research employs the methodology utilized within the ARIMA and SARIMA model frameworks The framework consists of five core phases: initial data preparation, selecting an appropriate model, estimating relevant parameters, conducting diagnostic assessments, and evaluating the predictive performance. These research stages are depicted in Figure 1.

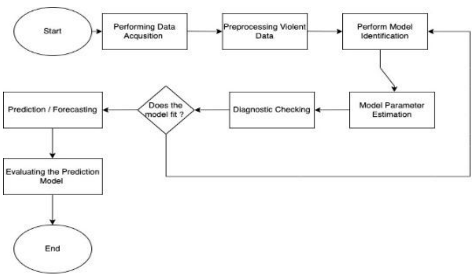


Fig. 1. Phases of The Research Process

A. Gathering of Datadata

The Open Data Jakarta website and the National Commission for the Protection of Women and Children were the two primary sources of data used in this research (https://satudata.jakarta.go.id/data/korban-pppa) The dataset encompasses the period from January to December 2024, covering a full 12 months. These sources offer thorough and trustworthy information on recorded cases of violence against women and children in Jakarta, ensuring the authenticity and applicability of the data for trend analysis and predictive modeling. The features or variables included in the dataset are data_periode, education_level, and total_victims. A total of 216 records were collected. Figure 2 contains the specifics of the victims' violence dataset.

	periode_data	pendidikan	jumlah
0	202406	NaN	72
1	202406	Perguruan Tinggi	21
2	202406	SD / Sederajat	16
3	202406	SMA / Sederajat	26
4	202406	SMP / Sederajat	4

Fig. 2. Victims Violent Dataset

B. Preprocessing

Preprocessing the victim violence data is the next stage following data collection. The purpose of this process is to prepare raw data for use as input in modeling, ultimately resulting in a higher-quality end model [27]. To guarantee data quality and prepare for analysis, the dataset underwent several crucial pre-processing stages. To preserve the integrity of the dataset, missing values were identified and addressed using appropriate imputation techniques. Subsequently, the dataset was normalized by employing one-hot encoding to convert categorical variables, such as types of violence, into numerical values [28]. Finally, the data was normalized to ensure consistent scaling of numeric features, thereby facilitating compatibility with time-series modeling methods such as ARIMA and SARIMA.

C. Identification of The Model

This research outlines the model selection process through a series of structured steps:

1) Plotting the is allowed them to determine whether it was stationar, specifically looking at whether the variance or mean showed stationarity. The time series data was subsequently divided down into a number of smaller parts in order to assess how each part affected the data series as a whole. Two models are typically used: Both multiplicative and additive decomposition.

2) It is possible to utilize the Augmented Dickey-Fuller Test to utilized alongside data visualization to assess whether the test statistic falls below the critical value in a stationary dataset [29]. To assess whether the time series data maintains consistent statistical properties over time, we examine the p-value. A result below 0.05 suggests stability in the data pattern, whereas a higher value implies that the data exhibits changing behavior over time. When such inconsistency is present, differencing can be used to normalize fluctuations in the average or variability. The process of calculating the shift or variation in observation values is referred to as differencing [30]. The acquired difference is rechecked to determine if it remains consistent. When data is stationary, it indicates that there is neither growth nor decline. As a result, regardless of time or oscillation variance, data variations are either continuously steady or center around a constant average value. [31]. Equation (1) describes the differencing equation

$$X'_t = X_t - X_{t-1} \quad (1)$$

Where :

X_t': First differencing
X_t : X value at order t
X_{t-1} : X value at order t-1

(3) Plots illustrating overall and lag-specific correlations serve as vital instruments in time series analysis for determining the appropriate model once the data has been made stationary. This phase involves selecting the model order by analyzing the ACF and PACF plots, particularly in the context of seasonally patterned data.

D. Evaluating Model Parameters

This phase involves determining the values of Moving Average, Autoregressive, as well as seasonal and non-seasonal parameters, followed by assessing their statistical significance. A model is considered to have failed the test if any of its parameters are not significant. To develop a model with relevant parameters, any unnecessary parameters will be removed.

E. Examining the Diagnosis

In this study, diagnostic checking establishes whether the model is appropriate and practical for forecasting. The features of the suitable conjectural model should resemble those of the original data. This assessment is conducted through a utilizing model diagnostics, a standard distribution test and a white noise diagnostic test. According to the optimal model, the residuals produced should exhibit characteristics of white noise or conform to a normal distribution.

F. Forecasting

Two models—ARIMA and SARIMA—are used in this study's prediction stage. A method for time series analysis called the ARIMA model makes use of autocorrelation and the fluctuation of time series residuals. The structure of the ARIMA model comprises three components: Autoregressive (AR), Moving Average (MA), and Integrated (I) models. The Integrated component indicates the order of differencing required to transform non-stationary data into a stationary series. Equation (2) presents the standard structure of the ARIMA model.

$$\Phi_p(B)\nabla^d Y_t = \xi + \Theta_q(B)\varepsilon_t \quad (2)$$

Where :

Φ_p : Autoregressive parameters
B : Backward sliding operator
d : Differencing Parameter
Y_t : Observation value at time t
ξ : Constant parameters
Θ_q : Moving average parameters
ε_t : Residual value (*error*

The SARIMA model, on the other hand, estimates future variables and identifies patterns in historical data by utilizing time-series data. The SARIMA framework is defined by the notation (p, d, q)(P, D, Q)[s], where 's' denotes the seasonal cycle length, and all parameters—p, d, q, P, D, and Q—are expressed as whole numbers. Equation (3) illustrates how this equation reflects the SARIMA model in its general version.

$$\Phi_p(B^s)\Phi_p(B)(1-B)^d(1-B^s)^D Y_t = \Theta_q(B)\Theta_q(B^s)\varepsilon_t \quad (3)$$

Where :

$\Phi_p(B)$: Non-seasonal autoregressive level
$\Phi_p(B_s)$: Seasonal autoregressive rate
$(1 - B)_d$: Non-seasonal differencing level
$(1 - B_s)_D$: Seasonal differencing level
$\Theta_q(B)$: *Non-seasonal moving average*
$\Theta_q(B_s)$: *Seasonal moving average*
Y_t : Actual data t-th
ε_t : t-period error

The ARIMA and SARIMA techniques were selected for this study partly because they have been extensively utilized and researched across various domains, including comparisons of the ARIMA and SARIMA algorithms for machine learning-based predictions. As per the study's findings, the sea level rise prediction models ARIMA and SARIMA exhibit exceptional performance, attaining a noteworthy forecast accuracy with a lower confidence level of 92.78%[32]. A later phase of the study involved utilizing ARIMA and SARIMA approaches to estimate the number of domestic passengers departing from Tanjung Perak Port. According to the findings, the SARIMA method was found to be the better strategy for this forecasting, while the analysis employing the ARIMA method produced a lower accuracy rating of 16.15%. [33]. Another study explored the performance of ARIMA and SARIMA models in forecasting crude oil prices through comparative analysis. The evaluation results show that both ARIMA and SARIMA have RMSE values of 1.905 Over the upcoming seven-day period, ARIMA forecasts a price of 86.230003, slightly outperforming SARIMA's estimate of 86.260002. The findings of this study are intended to assist policymakers in making informed decisions regarding the utilization of crude oil [34].

A notable feature that sets ARIMA and SARIMA apart from many other prediction methods is that ARIMA can produce forecasts independently, without incorporating outside influencing factors [35]. ARIMA aims to establish a robust statistical correlation between a variable's historical values and its expected future values, enabling the model to be utilized for forecasting. The Seasonal Autoregressive Integrated Moving Average (SARIMA) model was developed as an extension of the ARIMA framework. SARIMA is designed to analyze seasonal or recurrent data patterns at specified intervals, such as quarterly, semi-annually, and annually. By utilizing historical data, this predictive model seeks to establish relationships between expected variables to generate forecasts. Building upon earlier studies and leveraging the advantages of these techniques, this research utilizes ARIMA and SARIMA to estimate cases of violence involving women and children in Jakarta, with educational level considered as a contributing element.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

G. Analysis of The Findings

Analyze the forecasting operations' outcomes. The number of victims of violence against women and children in Jakarta will be predicted by a graph that displays the forecasting findings system effectiveness is evaluated using performance indicators like Root Mean Squared Error (RMSE) and Mean Squared Error (MSE) to determine if enhancements are required to better align with user needs.. The MSE calculation is detailed in Equation (4).

$$MSE = \sum_{t=1}^{n} \frac{(A_t - F_t)}{n} \quad (4)$$

The following formula, meanwhile, can be used to get the Root Mean Square Error (RMSE): (5) .

$$RMSE = \sqrt{\sum_{t=1}^{n} \frac{(A_t - F_t)^2}{n}} \quad (5)$$

Where:
A_t : Original data value
F_t : Forecasting data value
n : Amount of data

Mean Squared Error quantifies prediction accuracy by averaging the squares of the deviations between estimated outcomes and observed values. Typically, the MSE is used to assess the accuracy of forecasting models. However, the error rate RMSE is frequently used as a benchmark metric to gauge the reliability of prediction outcomes. Predictions are considered more accurate when the values of MSE or RMSE are smaller or approach zero. Perfect predictions occur when the results equal zero.

IV. Results And Discussion

In this chapter, the findings of the analysis are presented along with their implications for violence against women and children in Jakarta. The results are organized to provide a comprehensive understanding of the patterns observed in the dataset, their relationship to educational attainment, and the effectiveness of the ARIMA and SARIMA models in predicting outcomes. Each section highlights the efficiency of the forecasting techniques, significant trends, and seasonal fluctuations, integrating quantitative data with contextual observations.

A. Data Acquisition

The research dataset contains 72 records collected between January and December 2024, covering a 12-month period. It includes three main columns: which indicates the monthly time frame of the data; which classifies the victims' educational attainment (e.g., primary, secondary, or higher education) to investigate its connection to vulnerability to violence; and shows the number of victims reported for each education level in the corresponding month.

This information, which has been obtained from the National Commission for the Protection of Women and Children and Open Data Jakarta, provides a strong basis for trend analysis and predictive modeling.

B. Preprocessing

1) Missing Value Handling

The dataset on women and children who have been victims of violence in Jakarta must undergo preprocessing after data collection. Missing values are identified and removed as part of this process. Figure 3 below illustrates the results of the missing value analysis.

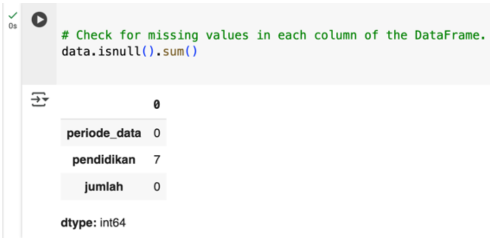


Fig. 3. Findings from the dataset's missing value check

The education_level column contains seven rows with missing values, as indicated by a missing value check. To address this issue, the rows with missing values will be removed from the dataset. This approach eliminates potential biases or inaccuracies that could arise from using incomplete data, ensuring that only complete and accurate information is utilized in the study. Although this results in a proportional decrease in the dataset size, it helps maintain the overall quality and reliability of the data for subsequent modeling and analysis. As a result, an effective way to handle the missing data is by eliminating the rows in which they appear; The outcome of this process are illustrated in Figure 4.

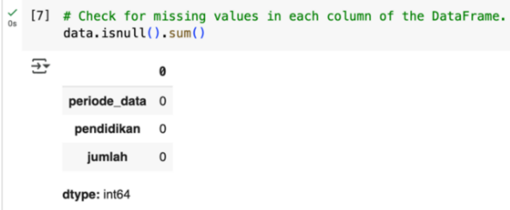
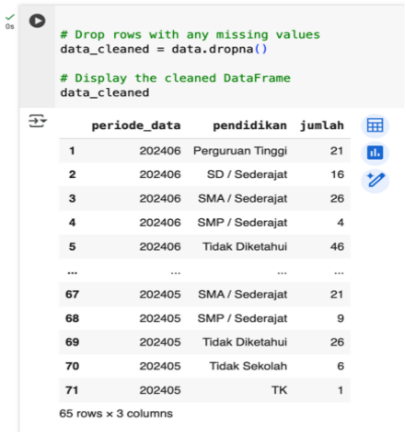


Fig. 4. Results of handling missing value dataset

In order to address the issue of missing values in 65 records, data in empty rows or columns (null values) is deleted. This process reduces the overall amount of data. Figure 5 illustrates the specifics of the data frame after the missing values have been addressed.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025



```
# Drop rows with any missing values
data_cleaned = data.dropna()

# Display the cleaned DataFrame
data_cleaned
```

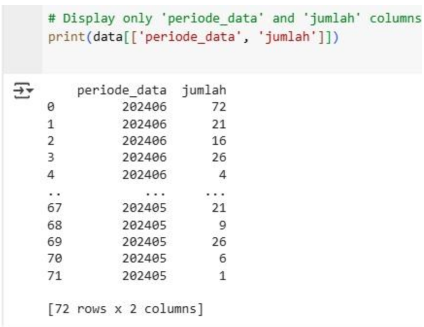
	periode_data	pendidikan	jumlah
1	202406	Perguruan Tinggi	21
2	202406	SD / Sederajat	16
3	202406	SMA / Sederajat	26
4	202406	SMP / Sederajat	4
5	202406	Tidak Diketahui	46
...	...	...	...
67	202405	SMA / Sederajat	21
68	202405	SMP / Sederajat	9
69	202405	Tidak Diketahui	26
70	202405	Tidak Sekolah	6
71	202405	TK	1

65 rows × 3 columns

Fig. 5. Results of treating a dataset with missing values

2) Data Integration

During the data integration phase, the time-series prediction model was constructed using the `data_period` and `total_victims` columns. The time dimension, represented by `data_period`, is crucial for establishing a timeline and identifying trends over time. The number of victims in each period serves as the target variable, provided by `total_victims`, and is essential for predicting future occurrences. The dataset is streamlined for time-series modeling by focusing on these two key columns, ensuring that the ARIMA and SARIMA models can effectively detect trends and seasonal variations in the data.



```
# Display only 'periode_data' and 'jumlah' columns
print(data[['periode_data', 'jumlah']])

    periode_data  jumlah
0         202406      72
1         202406      21
2         202406      16
3         202406      26
4         202406       4
..           ...     ...
67        202405      21
68        202405       9
69        202405      26
70        202405       6
71        202405       1

[72 rows x 2 columns]
```

Fig. 6. Outcomes of the dataset following the data integration phase

C. Model Identification

The model identification stage will involve several procedures, including creating time-series charts, utilizing the ADF procedure to test for stationarity test, and producing Autocorrelation Function and Partial Autocorrelation Function plots. The objectives of this analysis are to identify the underlying patterns in the data, evaluate stationarity, and determine the appropriate sequence for differencing and model parameters.

1) Create a time series plot

The data_period, month, and total_victims columns were utilized to create a time-series figure that illustrates patterns in violence cases against women and children over time. The data_period column presents the data in chronological order, the month column assists in identifying recurring trends, and the total_victims column indicates the total number of incidents recorded during each period. The visualization suggests potential seasonal trends, highlighting variations in the number of victims over several months, with notable peaks at certain times. This insight is crucial for understanding periodic fluctuations and will enhance the accuracy of predictive models such as ARIMA and SARIMA.

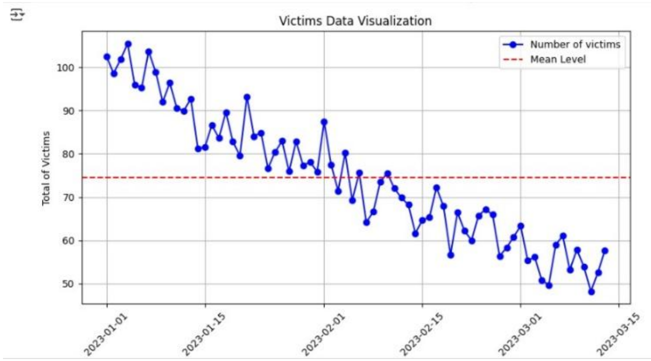


Fig. 7. Time series plot using the number of victims dataset

2) Time series stationarity check using ADFf

This study employed the Augmented Dickey-Fuller test to determine whether the time-series data is stationary, which is a crucial prerequisite for constructing reliable ARIMA and SARIMA models. To assess the stationarity of the korban (victim_count) time-series data, the Augmented Dickey-Fuller (ADF) test was applied. At the 1% (-3.526), 5% (-2.903), and 10% (-2.589) significance levels, the test result of -9.351 is significantly below the critical values. Furthermore, the p-value (8.29e-16) is considerably lower than the conventional limit of 0.05, offering robust support for rejecting the presence of a unit root. According to these findings, the time-series data is already stationary, indicating that further differencing is unnecessary to stabilize the trend. Given that stationarity is an essential condition for ARIMA and SARIMA modeling, these results confirm that the dataset is suitable for predictive modeling without any modifications. The results of the Augmented Dickey-Fuller (ADF) test can be found in Table 1, with its visualization presented in Figure 8.

TABLE I. STATIONARITY ASSESSMENT VIA ADF METHOD

Metric	Values
Test Statistics	-9.351093
p-value	8.291883e-16
Lags Used	0
Number of Observations Used	71
Critical Value (1%)	-3.526005
Critical Value (5%)	-2.903200
Critical Value (10%)	-2.588995

3) Correlation Analysis and Lag-based Correlation Analysis

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

After ensuring that since the data exhibits stationarity, plots of correlation and partial correlation over lags are generated. Seasonal patterns in the data are identified by determining the appropriate order of the model.

a) Autocorrelation Function

The order of the Moving Average (MA) model is determined using the Autocorrelation Function (ACF) plot. The purpose of this study is to assess whether the data is stationary in terms of its mean. The autocorrelation plot reveals a strong serial dependence, with autocorrelation values exceeding 0.5 for the first seven lags and progressively declining thereafter. Several lags fall outside the confidence interval, indicating a non-random pattern likely attributable to trends or seasonality. The results of the autocorrelation analysis are illustrated in Figure 9 below.

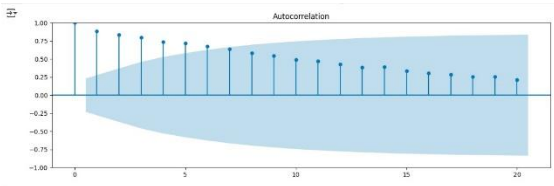


Fig. 8. Results of ACF Plots

b) Partial Autocorrelation Function

The Partial Autocorrelation Function (PACF) plot is utilized to identify the appropriate Autoregressive (AR) model. In Figure 10, the PACF reveals significant associations at lags 1 and 2 (-0.9 and 0.75, respectively), while all other lags fall within the confidence interval. This suggests that the time series primarily exhibits short-term dependence up to lag 2, after accounting for the effects of earlier lags. Given the steep decline observed after lag 2, these data may be best modeled by an Autoregressive model of order 2 (AR(2)). Higher-order AR terms may not be necessary, as the absence of substantial correlations beyond lag 2 indicates a lack of long-term dependencies.

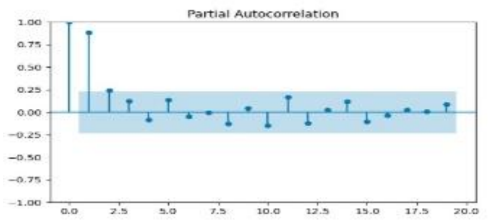
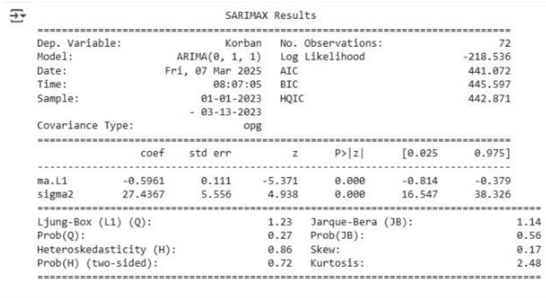


Fig. 9. Results of PACF Plots

D. Model Parameter Estimation

Model components encompassing periodic effects, regular patterns, error smoothing (MA), and lagged dependencies (AR) are considered all estimated at this stage, and their respective significance is evaluated. A model is deemed to fail the test if its parameters are not statistically significant. To assess and compare statistical models, researchers often rely on the Akaike and Bayesian Information Criteria (AIC and BIC) selecting ARIMA models and determining model parameters for both ARIMA and SARIMA. Below are the parameter estimations for the SARIMA model.

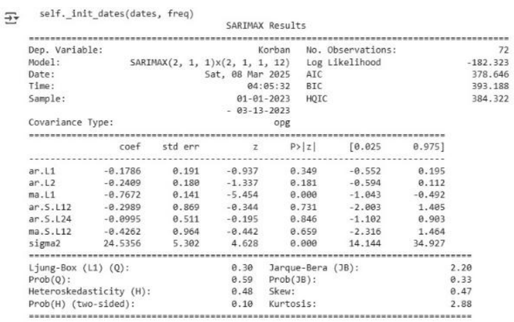
Figure 11 below presents the results of the ARIMA model significance test. An ARIMA (0,1,1) model was applied to the dataset, which consists of 72 observations. The results indicate a p-value of 0.000 and a moving average (MA) coefficient of -0.5961 at lag 1, demonstrating statistical significance. The variance of the residuals is represented by the sigma² value, which is 27.4367. Model fit is assessed using the model selection criteria: AIC (445.597), BIC (445.597), and HQIC (442.871). Diagnostic tests reveal that with a Jarque-Bera p-value of 0.72, the residuals exhibit characteristics consistent with a normal distribution. These findings suggest that while the model effectively captures short-term interdependence, it lacks an autoregressive component and is well-fitted.



```
                               SARIMAX Results
==============================================================================
Dep. Variable:                 Korban   No. Observations:                   72
Model:                 ARIMA(0, 1, 1)   Log Likelihood                -218.536
Date:                Fri, 07 Mar 2025   AIC                            441.072
Time:                        08:07:05   BIC                            445.597
Sample:                    01-01-2023   HQIC                           442.871
                         - 03-13-2023
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.5961      0.111     -5.371      0.000      -0.814      -0.379
sigma2        27.4367      5.556      4.938      0.000      16.547      38.326
===================================================================================
Ljung-Box (L1) (Q):                   1.23   Jarque-Bera (JB):                 1.14
Prob(Q):                              0.27   Prob(JB):                         0.56
Heteroskedasticity (H):               0.86   Skew:                             0.17
Prob(H) (two-sided):                  0.72   Kurtosis:                         2.48
===================================================================================
```

Fig. 10. Estimated Parameters of the ARIMA Model

The significant test results for the Sarima model are displayed in Figure 20 below. A Seasonal ARIMA (SARIMA) framework, specifically the SARIMAX (2,1,1)x(2,1,1,12), was applied to a dataset comprising 72 observations. The autoregressive (AR) and seasonal autoregressive (AR.S) terms were found to be insignificant (p-values < 0.05); however, the moving average (MA) and seasonal moving average (MA.S) terms were significant, indicating that the MA components are more effective at capturing short-term dependencies. Model diagnostics reveal that the residuals are normally distributed (Jarque-Bera p-value = 0.33) and show no meaningful evidence of correlation across lags, as indicated by a Ljung-Box p-value of 0.59. Overall, the model fits the data well, although there may be opportunities for improvement by reassessing the AR factors.



```
self._init_dates(dates, freq)
                                   SARIMAX Results
==========================================================================================
Dep. Variable:                             Korban   No. Observations:                   72
Model:             SARIMAX(2, 1, 1)x(2, 1, 1, 12)   Log Likelihood                -182.323
Date:                            Sat, 08 Mar 2025   AIC                            378.646
Time:                                    04:05:32   BIC                            393.188
Sample:                                01-01-2023   HQIC                           384.322
                                     - 03-13-2023
Covariance Type:                              opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.1786      0.191     -0.937      0.349      -0.552       0.195
ar.L2         -0.2409      0.180     -1.337      0.181      -0.594       0.112
ma.L1         -0.7672      0.141     -5.454      0.000      -1.043      -0.492
ar.S.L12      -0.2989      0.869     -0.344      0.731      -2.003       1.405
ar.S.L24      -0.0995      0.511     -0.195      0.846      -1.102       0.903
ma.S.L12      -0.4262      0.964     -0.442      0.659      -2.316       1.464
sigma2        24.5356      5.302      4.628      0.000      14.144      34.927
===================================================================================
Ljung-Box (L1) (Q):                   0.30   Jarque-Bera (JB):                 2.20
Prob(Q):                              0.59   Prob(JB):                         0.33
Heteroskedasticity (H):               0.48   Skew:                             0.47
Prob(H) (two-sided):                  0.10   Kurtosis:                         2.88
===================================================================================
```

Fig. 11. Estimated Parameters of the SARIMA Model

E. Examining The Diagnosis

Diagnostic verification follows the estimation of parameters for the ARIMA and SARIMA models to determine their appropriateness and effectiveness in forecasting. The results of the diagnostic testing on the ARIMA model, which includes an examination of density plots and residuals, are presented in Fig. 13. The histogram of the residuals reveals a right-skewed distribution, indicating that, rather than being perfectly normally distributed, the residuals are primarily centered around zero, with a few significant outliers. Since all lags fall within the confidence intervals, The residual ACF plot reveals no significant correlation patterns over time. This observation is reinforced by the results of the Ljung-Box test, which produces high p-values of 0.998861 and 0.999984, indicating a lack of serial dependence, suggesting that the residuals are uncorrelated and that the model effectively captures the time series patterns. The Ljung-Box statistical check was conducted to evaluate whether residuals show correlation across lags, and the findings suggest the model's adequacy in capturing time-dependent structure is well-specified, as the residuals behave like white noise.

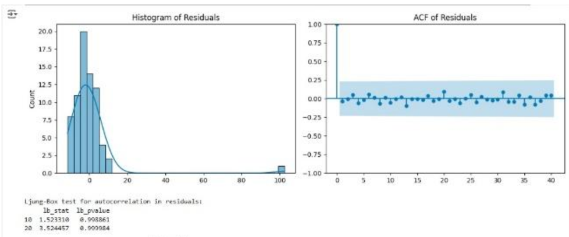


Fig. 12. Diagnostic Evaluation Outcomes for the ARIMA Model

Furthermore, diagnostic testing has been conducted on the SARIMA model. Potential outliers may be present, as the histogram of the residuals exhibits a right-skewed distribution, with most residuals concentrated near zero and a few extreme values. Since the majority of lag values fall remains inside the margin of estimation, the ACF plot of the residuals shows no significant autocorrelation.

According to the results of the Ljung-Box test, the lb_stat values at lags 10 and 20 are 2.283699 and 17.713411, respectively, with p-values of 0.993667 and 0.606281. There is no significant autocorrelation in the residuals, since both p-values exceed the threshold, the null hypothesis remains valid. This indicates that our SARIMA model is suitable for forecasting, as it accurately represents the time series structure and the residuals behave like white noise .

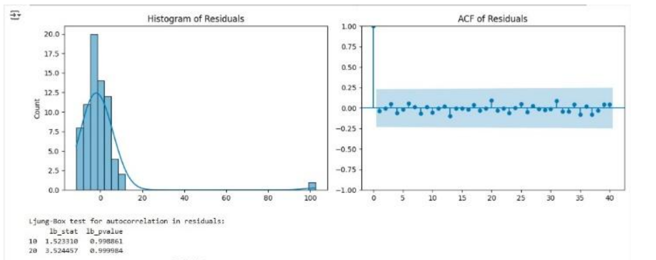


Fig. 13. SARIMA Model Diagnostics Checking Results

F. Forecasting

Following diagnostic testing, the forecast for each model is produced by contrasting the expected and actual values from the time series data. Forecasts become more accurate as a result of this process. Prediction results are obtained by applying ARIMA and SARIMA models to a reserved portion of the dataset designated for validation. To evaluate the correspondence between the actual and expected data, predictions are formulated. Figures 15 and 16, illustrating the forecasts generated by the ARIMA and SARIMA models respectively, demonstrate that the fluctuations in the predicted values closely resemble those observed in the recorded dataset.

The study on ARIMA-based predictions for child sexual abuse cases is illustrated in the graph, analyzing predicted values relative to empirical records, predicted cases, and the confidence interval. Initially, the training data and actual cases closely align in early 2025, with cases starting at nearly zero and progressively increasing. However, the actual data exhibits a more pronounced upward trend and begins to diverge from the forecast around March 2025. The ARIMA model predicts approximately 60 instances by June 2025, which underestimates the true trend, as actual cases exceed 80. As the confidence interval widens, it reflects increasing uncertainty, indicating that future projections may lose accuracy over time.

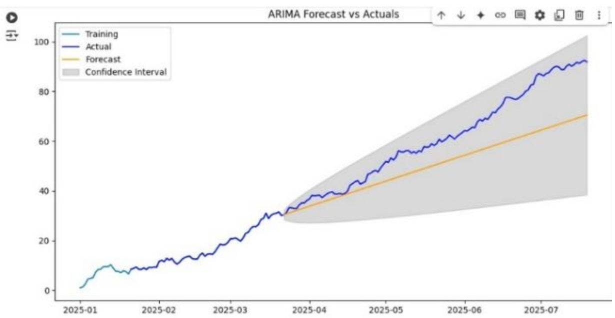


Fig. 14. ARIMA Model Training Data Forecasting Results

The chart provides a comparative analysis of training data,

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

actual data, and forecasted values to evaluate the predictive capability of the SARIMA (2,1,1) (1,1,1,12) model in estimating cases of child sexual violence. The actual data represents the real occurrences of such cases, while the training data comprises historical events used to develop the SARIMA model. Initially, from January to March 2025, the training data and actual instances align closely, showing a steady increase from 0 to approximately 30 cases. Beginning in April 2025, when the model commences its predictions, the forecasted values closely approximate the actual data, suggesting that the SARIMA model with parameters (2,1,1)(1,1,1,12) successfully models the underlying trend along with the recurring seasonal patterns in the data.

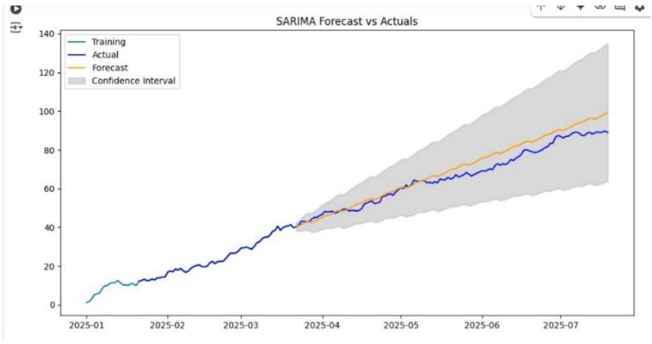


Fig. 15. SARIMA Model Training Data Forecasting Results

The actual instances show slight variations but usually fall within the predicted confidence interval when examined more closely. The prognosis somewhat overestimated at about 100 instances, but actual cases increase significantly from May to July 2025, reaching about 90 by that time. This implies that while the SARIMA model includes the general increasing trend, it significantly exaggerates the anticipated rise. In comparison to normal ARIMA, the model's order parameters (2,1,1,12) better account for seasonal impacts and short-term shocks, guaranteeing greater adaptability to periodic oscillations.

G. Prediction Model Evaluation

Following the prediction, the final step is to assess the outcomes of the forecasting process. Root Mean Square Error and Mean Squared Error calculations are used to test the model. The Root Mean Square Error (RMSE) serves as a metric to quantify the discrepancy between predicted outcomes and actual observations. When the RMSE score is minimal, it suggests strong agreement between the projected figures and the real-world measurements. Mean Square Error is a forecasting metric used to assess how accurate the forecasting results are. The accuracy of the forecasting findings increases with a smaller MSE value.

The prediction model's performance across several error measures is shown by the evaluation results. The model's predictions often deviate from the actual values by about 32.23 units, according to the Root Mean Square Error (RMSE) of 32.23; bigger errors are penalized more severely. The average absolute difference between the predicted and actual values is around 32.09 units, as indicated by the Mean Absolute Error (MAE) of 32.09. Mean Absolute Percentage (MAPE) of 5.19%, on the other hand, suggests that the model's predictions perform quite well, with an average error of 5.19% about the actual values. Table 2, which follows, displays the specifics of the model evaluation results.

TABLE II. EVALUATION OF THE PREDICTION MODELS OF SARIMA MODELS, AND THE ARIMA

Method	ARIMA	SARIMA
RMSE	32.22	80.26
MAE	32.09	66.21
MAPE	5.19	-

V. CONCLUSIONS

According to modeling conducted to forecast the incidence of violence against women and children in Jakarta. The models that performed the best were ARIMA (0,1,1) and SARIMA (0,1,1) (2,2,1) 12. The evaluation's findings indicate that the ARIMA model predicts violence against women and children in Jakarta more accurately than the SARIMA model. The ARIMA model demonstrates greater prediction accuracy and consistency, as evidenced by its significantly lower RMSE (32.22) and MAE (32.09) compared to SARIMA's RMSE of 80.26 and MAE of 66.21. Furthermore, the ARIMA model's MAPE of 5.19% reflects a comparatively low percentage error, while the SARIMA model's MAPE could not be calculated, possibly due to the presence of zero or negative actual values in the data. According to these results, a simpler ARIMA model performs better on this forecasting task than its complex seasonal equivalent. Hybrid modeling approaches, including merging ARIMA with machine learning techniques or adding external variables, like socioeconomic indicators or law enforcement data, are suggested for further research, as they may increase prediction power

Furthermore, a more thorough seasonality analysis might help in improving SARIMA combinations' effectiveness. Effectiveness can also be enhanced by incorporating advanced techniques such as time series decomposition or dynamic regression models, which allow for a more nuanced insight into the hidden structures within the dataset. By exploring these methods, researchers may uncover additional insights that can lead to more accurate and reliable forecasts.

REFERENCES

[1] M. Myall, S. Morgan, and S. Scott, "Editorial: Domestic violence and abuse: increasing global and intersectional understanding," *Frontiers in Health Services*, vol. 4, Sep. 2024, doi: 10.3389/frhs.2024.1465688.

[2] A. Bluschke, N. Faedda, J. Friedrich, and E. J. Dommett, "Editorial: Women in psychiatry 2023: ADHD," *Front Psychiatry*, vol. 15, Aug. 2024, doi: 10.3389/fpsyt.2024.1447958.

[3] Y. F. Wismayanti, P. O'Leary, C. Tilbury, and Y. Tjoe, "Child sexual abuse in Indonesia: A systematic review of literature, law and policy,"

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

Child Abuse Negl, vol. 95, p. 104034, Sep. 2019, doi: 10.1016/j.chiabu.2019.104034.

[4] A. L. Wirtz *et al.*, "Development of a screening tool to identify female survivors of gender-based violence in a humanitarian setting: qualitative evidence from research among refugees in Ethiopia," *Confl Health*, vol. 7, no. 1, p. 13, Dec. 2013, doi: 10.1186/1752-1505-7-13.

[5] C. Stoicescu, B. Medley, E. Wu, N. El-Bassel, P. Tanjung, and L. Gilbert, "Synergistic effects of exposure to multiple types of violence on non-fatal drug overdose among women who inject drugs in Indonesia," *International Journal of Drug Policy*, vol. 129, p. 104486, Jul. 2024, doi: 10.1016/j.drugpo.2024.104486.

[6] N. Lepcha and S. Paul, "Exploring Violence Against Children Under Sustainable Development Goals," 2021, pp. 286–296. doi: 10.1007/978-3-319-95687-9_72.

[7] Ore-ofe Loveth Oluwajobi, Chidinma Favour Udechukwu, and Toluwanimi Oreoluwa Arogundade, "Understanding the impact of domestic violence on children's mental health and exploring effective intervention strategies," *World Journal of Advanced Research and Reviews*, vol. 23, no. 3, pp. 1405–1418, Sep. 2024, doi: 10.30574/wjarr.2024.23.3.2812.

[8] Riswanda, J. McIntyre-Mills, and Y. Corcoran-Nantes, "Prostitution and Human Rights in Indonesia: A Critical Systemic Review of Policy Discourses and Scenarios," *Syst Pract Action Res*, vol. 30, no. 3, pp. 213–237, Jun. 2017, doi: 10.1007/s11213-016-9393-4.

[9] G. Dhamija, P. Roychowdhury, and B. Shankar, "Does urbanization empower women? Evidence from India," *J Popul Econ*, vol. 38, no. 1, p. 27, Mar. 2025, doi: 10.1007/s00148-025-01085-4.

[10] A. Al Mutair *et al.*, "Domestic violence and childhood trauma among married women using machine learning approach: a cross-sectional study," *BMC Public Health*, vol. 25, no. 1, p. 1340, Apr. 2025, doi: 10.1186/s12889-025-22537-2.

[11] D. Rukmana and D. Ramadhani, "Income Inequality and Socioeconomic Segregation in Jakarta," 2021, pp. 135–152. doi: 10.1007/978-3-030-64569-4_7.

[12] O. O. Okedare, M. M. Salawu, and O. I. Fawole, "Intimate partner violence and quality of life of young women in urban slum and non-slum communities, Ibadan, Nigeria," *BMC Public Health*, vol. 25, no. 1, p. 1199, Mar. 2025, doi: 10.1186/s12889-025-22385-0.

[13] Md. R. Kabir, S. Ghosh, and A. Shawly, "Causes of Early Marriage and Its Effect on Reproductive Health of Young Mothers in Bangladesh," *Am J Appl Sci*, vol. 16, no. 9, pp. 289–297, Sep. 2019, doi: 10.3844/ajassp.2019.289.297.

[14] D. Tunas and A. Peresthu, "The self-help housing in Indonesia: The only option for the poor?," *Habitat Int*, vol. 34, no. 3, pp. 315–322, Jul. 2010, doi: 10.1016/j.habitatint.2009.11.007.

[15] Mr. B. R. A. -, "Harnessing Technology and Data Analytics to Advance Prevention and Treatment in the Opioid Crisis," *International Journal For Multidisciplinary Research*, vol. 5, no. 5, Oct. 2023, doi: 10.36948/ijfmr.2023.v05i05.31737.

[16] Y. Cao *et al.*, "Recognize Human Activities from Partially Observed Videos," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2013, pp. 2658–2665. doi: 10.1109/CVPR.2013.343.

[17] J. W. Petty, "Research in Small-Firm Entrepreneurial Finance: A Note on Developing a Paradigm," *The Journal of Entrepreneurial Finance*, vol. 1, no. 1, pp. 88–90, Dec. 1991, doi: 10.57229/2373-1761.1114.

[18] A. A. Fagan *et al.*, "Scaling up Evidence-Based Interventions in US Public Systems to Prevent Behavioral Health Problems: Challenges and Opportunities," *Prevention Science*, vol. 20, no. 8, pp. 1147–1168, Nov. 2019, doi: 10.1007/s11121-019-01048-8.

[19] G. M. Campedelli, A. Aziani, and S. Favarin, "Exploring the Immediate Effects of COVID-19 Containment Policies on Crime: an Empirical Analysis of the Short-Term Aftermath in Los Angeles," *American Journal of Criminal Justice*, vol. 46, no. 5, pp. 704–727, Oct. 2021, 10.1007/s12103-020-09578-6.

[20] H. Seyidoglu, G. Farrell, A. Dixon, J. Pina-Sánchez, and N. Malleson, "Post-pandemic crime trends in England and Wales," *Crime Sci*, vol. 13, no. 1, p. 6, Mar. 2024, doi: 10.1186/s40163-024-00201-1.

[21] S. Yao *et al.*, "Predicting Land Use Changes under Shared Socioeconomic Pathway–Representative Concentration Pathway Scenarios to Support Sustainable Planning in High-Density Urban Areas: A Case Study of Hangzhou, Southeastern China," *Buildings*, vol. 14, no. 7, p. 2165, Jul. 2024, doi: 10.3390/buildings14072165.

[22] C. V. Redoblo, J. L. G. Redoblo, R. A. Salmingo, C. M. Padilla, and J. C. T. Arroyo, "Forecasting the influx of crime cases using seasonal autoregressive integrated moving average model," *International Journal of ADVANCED AND APPLIED SCIENCES*, vol. 10, no. 8, pp. 158–165, Aug. 2023, doi: 10.21833/ijaas.2023.08.018.

[23] S. Siamba, A. Otieno, and J. Koech, "Application of ARIMA, and hybrid ARIMA Models in predicting and forecasting tuberculosis incidences among children in Homa Bay and Turkana Counties, Kenya," *PLOS Digital Health*, vol. 2, no. 2, p. e0000084, Feb. 2023, doi: 10.1371/journal.pdig.0000084.

[24] M. K. Lubeya *et al.*, "Using the ARIMA Model to forecast sexual and gender-based violence cases reported to a tertiary hospital in Lusaka, Zambia," *PAMJ - One Health*, vol. 5, 2021, doi: 10.11604/pamj-oh.2021.5.4.27590.

[25] A. Anavatan and E. Y. Kayacan, "Investigation of femicide in Turkey: modeling time series of counts," *Qual Quant*, vol. 58, no. 3, pp. 2013–2028, Jun. 2024, doi: 10.1007/s11135-023-01619-6.

[26] S. Rashid, "Impact of COVID-19 on Selected Criminal Activities in Dhaka, Bangladesh," *Asian J Criminol*, vol. 16, no. 1, pp. 5–17, Mar. 2021, doi: 10.1007/s11417-020-09341-0.

[27] A. Luong *et al.*, "Comparison of Machine Learning Models to a Novel Score in the Identification of Patients at Low Risk for Diabetic Retinopathy," *Ophthalmology Science*, vol. 5, no. 1, p. 100592, Jan. 2025, doi: 10.1016/j.xops.2024.100592.

[28] F. Bolikulov, R. Nasimov, A. Rashidov, F. Akhmedov, and Y.-I. Cho, "Effective Methods of Categorical Data Encoding for Artificial Intelligence Algorithms," *Mathematics*, vol. 12, no. 16, p. 2553, Aug. 2024, doi: 10.3390/math12162553.

[29] Z. Liang and M. T. Ismail, "Advanced CEEMD hybrid model for VIX forecasting: optimized decision trees and ARIMA integration," *Evol Intell*, vol. 18, no. 1, p. 12, Feb. 2025, doi: 10.1007/s12065-024-00984-x.

[30] Y. Shen *et al.*, "Near real-time corn and soybean mapping at field-scale by blending crop phenometrics with growth magnitude from multiple temporal and spatial satellite observations," *Remote Sens Environ*, vol. 318, p. 114605, Mar. 2025, doi: 10.1016/j.rse.2025.114605.

[31] R. Refinetti, "Non-stationary time series and the robustness of circadian rhythms," *J Theor Biol*, vol. 227, no. 4, pp. 571–581, Apr. 2004, doi: 10.1016/j.jtbi.2003.11.032.

[32] K. M. Wantzen, K.-O. Rothhaupt, M. Mörtl, M. Cantonati, L. G.-Tóth, and P. Fischer, "Ecological effects of water-level fluctuations in lakes: an urgent issue," *Hydrobiologia*, vol. 613, no. 1, pp. 1–4, Nov. 2008, doi: 10.1007/s10750-008-9466-1.

[33] S. Putri and A. Sofro, "Peramalan Jumlah Keberangkatan Penumpang Pelayaran Dalam Negeri di Pelabuhan Tanjung Perak Menggunakan Metode ARIMA dan SARIMA," *MATHunesa: Jurnal Ilmiah Matematika*, vol. 10, no. 1, pp. 61–67, Apr. 2022, doi: 10.26740/mathunesa.v10n1.p61-67.

[34] V. P. Ariyanti and Tristyanti Yusnitasari, "Comparison of ARIMA and SARIMA for Forecasting Crude Oil Prices," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, pp. 405–413, Mar. 2023, doi: 10.29207/resti.v7i2.4895.

[35] Elvina Catria, A. A. Putra, D. Permana, and D. Fitria, "Adding Exogenous Variable in Forming ARIMAX Model to Predict Export Load Goods in Tanjung Priok Port," *UNP Journal of Statistics and Data Science*, vol. 1, no. 1, pp. 31–38, Feb. 2023, doi: 10.24036/ujsds/vol1-iss1/10.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2349, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 16, 2025, Published : May 26, 2025

Implementation of Triple Exponential Smoothing Method To Predict Palm Oil Production of PT.Lonsum Web-Based

Syafhira Ananda Galasca[1]*, Aninda Muliani Harahap[2]
Department of Information Systems[1],[2]
University of Islamic Negeri Sumatera Utara
Medan, Indonesia
syafhiragalasca@gmail.com[1]*, anindamh@gmail.com[2]

***Abstract*— This research aims to develop a web-based palm oil (CPO) production forecasting system by applying the Triple Exponential Smoothing (TES) method to the production data of PT Lonsum Turangi. The data used includes 60 monthly data from 2020 to 2024. The first 36 data were used for model training, while the remaining 24 data were used for validation. Research instruments included semi-structured interviews and participatory observations to understand the operational patterns and needs of the system in the field. Triple Exponential Smoothing method was chosen for its ability to handle level, trend and seasonal components simultaneously, making it superior to other time series forecasting methods that require large volumes of data. The system was developed using the Rapid Application Development (RAD) method, PHP programming language, and MySQL database. The test results show a good level of prediction accuracy with a Mean Absolute Percentage Error (MAPE) value of 17.34% at an alpha value of 0.1. This system not only improves prediction accuracy, but also provides practical benefits in production planning, meeting market demand, and reducing potential losses due to production imbalances. The novelty of this research lies in the integration of the TES method into a web-based decision support system specific to the CPO industry.**

Keywords*— *Crude Palm Oill (CPO), Prediction, Production, Triple Exponential Smoothing, Website

I. INTRODUCTION

Information technology continues to develop rapidly where this is also influenced by the use of various types of technological media, one of which is in the industrial sector[1]. The palm oil industry is one of the sectors that continues to innovate with technology. This industry is one of the economic sectors that plays an important role as a foreign exchange earner as well as a significant provider of employment [2]. The majority of foreign exchange from the palm oil industry is generated by crude palm oil (CPO) and palm kernel oil (PKO). CPO is the oil derived from the flesh of the oil palm fruit while PKO is the oil derived from the seeds of the oil palm fruit [3].

PT. PP London Sumatera Utara Tbk (Lonsum) is a leading agro-industrial company engaged in the management of agricultural commodities, one of which is oil palm. Overall Crude Palm Oil (CPO) production at PT.Lonsum fluctuates from year to year with total production reaching 194 thousand tons in 2024. In the previous year the amount of production had increased even though in 2024 it decreased. The amount decreased by 10% from the previous year. One of PT.Lonsum palm oil mills is located in the village of Turangie, Langkat Regency focusing on the management of Fresh Fruit Bunches (FFB) into palm oil derivative products, especially Crude Palm Oil (CPO). PT.Lonsum palm oil mill faces a problem of inaccurate prediction of the amount of palm oil (CPO) production. PT.Lonsum palm oil mill still predicts the amount of CPO production manually, without utilizing historical data that is systematically integrated. Estimated production results are still measured based on the number of incoming FFB, machine capacity and weather conditions. This has an impact on not fulfilling the demand for CPO, less than optimal factory production and company losses. The gap between market demand and production capacity is increasingly visible due to the inability to utilize historical production data for more accurate forecasting results.

Forecasting is a process carried out to estimate future situations based on historical past data [4]. Forecasting is an effort to determine the quantity of products in the future with certain constraints and conditions [5]. The use of forecasting plays a role in careful planning, determining the resources needed as well as the basis for making the right decisions so that the results needed in the future can meet the target [6]. Forecasting production quantities is one of the crucial aspect of operations and supply chain management. In forecasting production quantities, one of the processes involves estimating the quantity of products that will be produced in the future. This is measured based on historical data, market trends and market demand. Forecasting is widely used in various companies to predict future conditions by testing past conditions [7]. However, previous studies have shown that problems with the accuracy and effectiveness of palm oil production forecasting have serious implications for the gap between demand and production capacity [8].

Triple Exponential Smoothing is one of the improved algorithms from the Single and Double Exponential Smoothing algorithms. This algorithm is an appropriate prediction algorithm by considering level, trend and seasonal factors [9].

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2358, Copyright ©2025
Submitted : May 5, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 28, 2025

This algorithm is considered very suitable for Crude Palm Oil (CPO) production patterns which are also influenced by seasonal factors. Unlike other forecasting algorithms such as ARIMA that require stationary data or machine learning models that require large amounts of data [10], Triple Exponential Smoothing can work effectively even with more limited stationary data. It is also able to capture regularly recurring seasonal patterns. Forecasting algorithms that are applied to an integrated website can make their effectiveness and efficiency better. A website-based prediction system can also make it easier for companies to predict future production results.

In research conducted by Nelfi Yolanda [11] using the Triple Exponential Smoothing method to predict the amount of pineapple production in Riau in the first quarter to the third quarter of 2025 resulted in a value of MAPE = 3.7% MAD = 1.93% MSD = 7.05% which can be categorized as accurate and very good. This is characterized by previous data and forecast data that has been made. Prediction calculations in this study were made using Minitab software but were not developed in the form of a website. The next research by Dewi et al., [12] Using the Triple Exponential Smoothing Method to predict motor vehicle tax revenue in North Sumatra. The results showed the smallest error value with a MAPE value below 10% which proves the prediction accuracy is very good.

Research by Sandika et al., [13] combine Triple Exponential Smoothing and Double Moving Average method to predict palm kernel production. This study produces an optimum alpha value with the MAPE level of the Triple Exponential Smoothing method of 9.48% where this is categorized as very good and the Double Moving Average MAPE value of 11.2% with a good category. The results of the study are expected to anticipate palm kernel production more optimally. Saputro et al., [14] implemented Triple Exponential Smoothing method for predicting helmet sales with the results of Cargloss helmet sales data obtained resulted in forecasting for 2022. The results of the accuracy calculation with Mean Absolute Percentage Error with Alpha, Beta and Gamma values of 0.3 are 44.4%. Based on the Mean Absolute Percentage Error value, helmet sales forecasting with the Triple Exponential Smoothing method is feasible to use. Research conducted by R.Puspita [15] use Triple Exponential Smoothing method to predict Banten Province's Unemployement rate for the next 7 periods. The resultsd from the forecasting in a value of MAPE = 8.858859% which that shows very accurated prediction.

The new contribution of this research is the integration of Triple Exponential Smoothing method into a web-based decision support system specifically for the palm oil industry. This research will allow the process of forecasting the amount of CPO production to be carried out automatically, real-time, and integrated, improving the accuracy of predictions and the efficiency of operational decision making. With the ability of Triple Exponential Smoothing to handle trends and seasonality simultaneously, this system is very relevant to be applied to volatile production data such as in PT.Lonsum Turangi palm oil mill. This research is a review of several previous studies, namely the use of proven predictions to streamline planning, preparation and optimization of operations for research sites, both in the form of companies and businesses of various levels.

This research uses the Triple Exponential Smoothing method and develops it in the form of a website to predict future Crude Palm Oil (CPO) production results so that the resulting production is maximized, market demand can be met and minimize company losses. By implementing the Triple Exponential Smoothing method into an integrated website, this research can be a reference and technology development in the industry, especially the palm oil industry in the future..

II. METHODOLOGY

This research uses quantitative methods where this method emphasizes objective measurement, measurement and generalization of research results. The main purpose of using this method is to produce findings that can later be measured and tested statistically to support or reject the research hypothesis[16]. This research uses several stages in implementing the Triple Exponential Smoothing method into a web-based system. The research stage include observation or problem identification, data collection, system planning and development, algorithm implementation and testing and evaluation. The making of research stages aims to make the research carried out have a clear development structure and can implement the algorithm properly in accordance with the problems and observations that have been made.

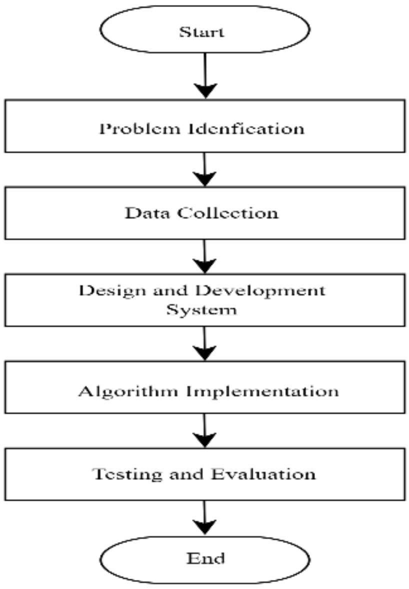


Fig. 1. Research Stage

A. Research Data and Research Instruments

This research use Palm Oil (CPO) production data of PT.Lonsum Turangie for the last 5 year period (2020-2024) collected from the production department of PT.Lonsum. The data was collected in kilograms (Kg) per month, with total of 60 data points. Then, the data was divided into two parts: The first 36 data points (2020-2022) were used as training data to build the forecasting model, while the next 24 data points (2023-2024) were used for model validation.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2358, Copyright ©2025
Submitted : May 5, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 28, 2025

TABLE I. SAMPLE PRODUCTION DATA

CPO Production Data (Kg)			
Month	**2020**	**2021**	**2022**
Jan	2921590	2546240	1912060
Feb	2961510	2584770	2309680
Mar	3151180	3816720	3135080
Apr	4030690	3154050	2767410
Mei	2902720	2784580	2476290
June	3586180	3195620	3060090
July	3567940	2972110	2792920
Aug	2925480	2672330	2337800
Sept	2523180	3068400	2545780
Okt	2994030	2260500	2093150
Nov	2202440	2297550	2449280
Des	2638310	2622660	3280290
Total	36405250	33975530	31159830

In this research, the instruments used for data collection were semi-structured interviews and participative observation. Interview were conducted with stakeholders who have a direct role in the production and planning process at PT.Lonsum Turangi palm oil mill. The interview aimed to collect depth information about production patterns, operational challenges and forecasting methods that have been used by the company. Direct observations were also made at PT.Lonsum Turangi palm oil mill to learn about the flow of CPO production process, daily data recording and the interaction between the working unit in making production decisions. The combination of interviews and observations provides a comprehensive overview of the needs of an accurate and real-time forecasting system, and becomes the basis for designing a system that is in accordance with field conditions.

B. *System Development Method*

This research uses one of the models from the System Development Life Cycle (SLDC), namely the Rapid Application Development (RAD) or Rapid Prototyping development method. RAD is an incremental software development method. The Rapid Application Development (RAD) method emphasizes a short, short and fast software development cycle [17]. Rapid Application Development (RAD) develops systems through a working model that is contributed at the beginning of development. Where this aims to determine the final user needs (requirements). In this system development model, the working model is only used once as an initial design as well as the final system implementation. The Rapid Application Development (RAD) method is also a linear sequential development method that emphasizes a short development cycle [18].

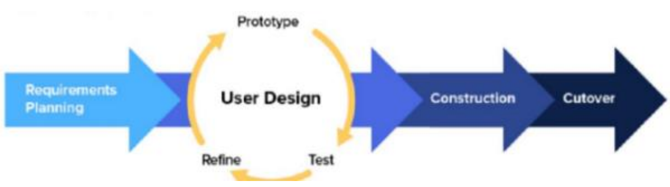


Fig. 2. Rapid Application Development (RAD) Stage Method

The stages of development using the Rapid Application Development (RAD) method that the author does:

1) *Requirements planning*

Requirements planning is based on observations and identification of problems that have been carried out previously. This stage aims to make the system that will be made in accordance with user needs and can be a solution to the problems faced by users [19]. The requirement panning stage is carried out in the following way

- *Observation*

  Author makes direct observations at the Palm Oil Mill PT.Lonsum Turangi operational area where production data management is done using Microsoft Excel, however there is no system that can predict the amount of CPO production in the future.

- *Interview*

  Author conducted an interview with the Head of Administration in PT Lonsum Palm Oil Mill Turangi operational area regarding the production process, production raw materials, and production data management. The interview process is also dug up information and data information and data available at the PT Lonsum Palm Oil Mill in the Turangi operational area.

- *Literature study*

  Author conducts a literature study by studying journals, books and literature related to the research, including forecasting, Exponential smoothing method, Triple exponential smoothing method and website implementation.

2) *User design*

The user design stage is carried out to design the system process and interface that will be proposed to users [20]. User design design is carried out using UML (Unified Modeling Language) as system modelling. This stage aims to make sure that the design made is in accordance with user needs.

3) *System development (construction)*

After the design has been made and confirmed according to user needs, the design is developed into the form of a system that has been planned previously [21]. In this research, system development is carried out starting from coding to adjusting the design that has been made into the system using PHP and MySQL as a database.

4) *Testing (cutover)*

After the development is carried out, the system will be developed. Testing (cutover) After system development is carried out, testing is carried out on the system that has been developed[22]. Testing is carried out on the program module then continued with black-box testing, verification of system features, forecasting accuracy testing and comparison of system forecasting results with the manual forecasting method used in PT Lonsum Turangi palm oil mill.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2358, Copyright ©2025
Submitted : May 5, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 28, 2025

C. Triple Exponential Smoothing Method

Triple Exponential Smoothing is one of the forecasting methods developed by Brown. The Triple Exponential Smoothing method performs forecasting with three times smoothing analysis using a predetermined constant value with the aim that the forecasting results obtained are effective and appropriate so that prediction errors can be minimized [23]. This method is used on stationary and non-stationary data that has a regularly recurring seasonal pattern [24].

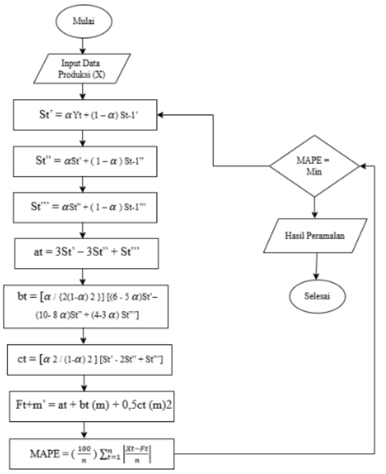


Fig. 3. Triple Exponential Smoothing Algorithm Flowchart

In the Triple Exponential Smoothing method, the first step is to determine the value based on trial and error by finding the value closest to the actual value. This method is applied to CPO production data of PT.Lonsum Turangi palm oil mill with the parameter $\alpha = 0.1$ which is choosen based on comparative trials. Then the Triple Exponential Smoothing is implemented in the system as follows :

1) Parameter initialization

At this stage, the value of parameter $\alpha = 0.1$ is determined based on the trial results. Then initialize the first values of S'1, S''1 and S'''1 using the actual data of the first period.

2) Calculating of Smoothing Value

After obtaining the parameter value, three smoothing calculations are carried out including : (1) calculating the Single Exponential Smoothing (S't) value for level data, (2) calculating Double Exponential Smooting (S''t) value for trend data and (3) calculating Triple Exponential Smooting (S'''t) value for seasonal data.

3) Forecasting parameter calculation

After calculating the smoothing value, the next step is calculate the constant (at) as the base of forecasting value, the slope (bt) for the trend component and ct value for seasonal component.

4) Forecasting value calculation

At this stage, the calculation of the forecasting value for the upcoming period is carried out with the following equation:

$$Ft + m = at + bt\,(m) + ½\, ct\,(m^2) \qquad (1)$$

5) Accuracy measurement

After obtaining the forecasting value is obtained, the next test is carried out on the forecasting results using the Mean Absolute Percentage Error (MAPE) equation. Mean Absolute Percentage Error (MAPE) is amethod of measuring the absolute error rate in forecasting period by finding the percentage error value of the difference between actual data and forecasting data. The smaller MAPE value, more accurate the forecasting [25]. This following table is the criteria for determining MAPE.

TABLE II. INTERPRETATION OF MAPE VALUE

MAPE	Interpretation
<10%	Highly accurate forecasting
10% -20%	Good forecasting
20% - 50%	Reasonable forecasting
>50%	Innacurate forecasting

To calculate the MAPE value using the following equation:

$$MAPE = \frac{1}{n} \sum^{n} i{=}0 \mid Ft{-}yi\ yi \mid n\ i{=}0 \times 100\% \qquad (2)$$

Where :
Ft = forecasting value
yi = actual value
n = data

III. Result And Discussion

A Data Analysis

Based on the sample data, there is a pattern of decreasing CPO production over a 3-year period, with total production decreasing from 36,405,250 kg in 2020 to 31,159,830 kg in 2022 (a decrease of approximately 14.4%). This pattern indicates a downward trend that needs to be considered in the forecasting analysis. In addition, the data shows significant monthly fluctuations, with the highest production pattern generally occurring in March-April and the lowest in November, indicating seasonal factors affecting CPO production.

Some of the factors that can affect this production pattern include:

- *The oil palm harvest cycle which is influenced by the rainy and dry seasons.*
- *Productive age of oil palm plants in PT.Lonsum plantations*
- *Fluctuating weather and climate conditions during the period*
- *Technical factors such as machine capacity and mill operational efficiency*

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2358, Copyright ©2025
Submitted : May 5, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 28, 2025

TABLE III. SAMPLE PRODUCTION DATA

CPO Production Data (Kg)			
Month	**2020**	**2021**	**2022**
Jan	2921590	2546240	1912060
Feb	2961510	2584770	2309680
Mar	3151180	3816720	3135080
Apr	4030690	3154050	2767410
Mei	2902720	2784580	2476290
June	3586180	3195620	3060090
July	3567940	2972110	2792920
Aug	2925480	2672330	2337800
Sept	2523180	3068400	2545780
Okt	2994030	2260500	2093150
Nov	2202440	2297550	2449280
Des	2638310	2622660	3280290
Total	36405250	33975530	31159830

B User Requirement Analysis

Identification as well as data analysis, then an analysis of user needs for the system is carried out. The analysis process includes information collection, modeling and system development according to user needs [26]. By analyzing user needs, it will be easier for researchers to build a system that can be a solution that is functional as well as intuitive and user-friendly [27]. User needs analysis was carried out on three users, namely the Administrator, Production Manager and Production Supervisor.

TABLE IV. USER REQUIREMENT ANALYSIS

User Requirements for the System			
No	**Actor**	**Feature**	**Description**
1	Administrator	- Login and access - Prediction chart dashboard - User management - Historical data management - Monitoring sistem - Export reports	- Administrator can manage other users account - Upload, Update and export production data
2	Production Manager	- Login - Prediction graphic dashboard - View historical and prediction reports - Export reports	- Production managers can access historical production data and view prediction results
3	Production Supervisor	- Login - View historical data - Perform prediction - View prediction results - Export reports	- Production supervisor can access historical data and prediction data - Using the *Triple Exponential Smoothing* Algorithm

C Unified Modelling Language (UML)

As system modeling, researchers use UML (Unified Modeling Language) use case diagrams so that the system workflow can be described properly. The following is a use case diagram used in the palm oil (CPO) production forecasting system at PT.Lonsum with three users. The use case diagram created has been adjusted to the user needs analysis that has done previously.

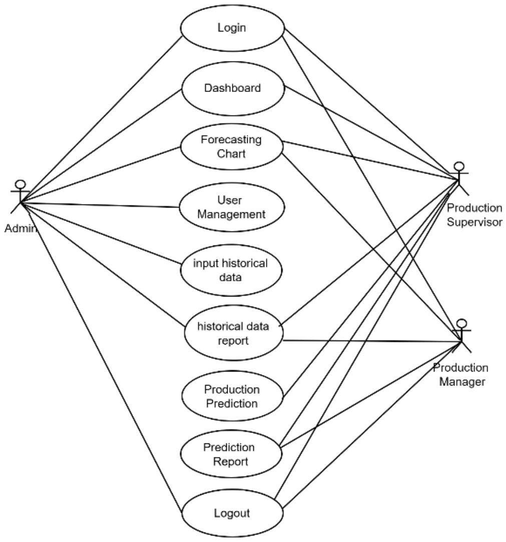


Fig. 4. Unified Modelling Language (UML)

D Application of the Triple Exponential Smoothing Method

In applying the triple exponential smoothing method, the first thing to do is to determine the parameter values first trough trial and error. The parameter value used is 0.1. where is the value choosen as a value that is close to the actual value. Next,the 12th period Single Exponential Smoothing (S′t) is performed as follows :

$$S't = a\,Xt + (1 - a)\,S't - 1 \quad (1)$$

$S'1 = 1775040$

$S'2 = (0.1)\ 1959700 + (1 – 0.1)\ 1775040$
$\quad = 1793506.00$

$S'3 = (0.1)2899260 + (1 – 0.1)\ 1793506.00$
$\quad = 1904081.40$………….etc

After the Single Exponential Smooting is done, the second smoothing or Double Exponential Smooting is done as follows:

$$S''t = a\,S't + (1 - a)\,S''t - 1 \quad (2)$$

$S''1 = 1775040$

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2358, Copyright ©2025
Submitted : May 5, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 28, 2025

S″2 = (0.1) 1793506,00 + (1 – 0.1)1775040
= 1776886.60
S″3 = (0.1) 1904081,40 + (1 – 0.1)1776886.60

= 1789606.60…………..etc

Next, the third smoothing or Triple Exponential Smoothing is carried out as follows:

S‴t = a S″t+ (1 − a) S‴t - 1 (3)
S‴1 = 1775040
S‴2 = (0.1) 1775040 + (1 – 0.1) 1776886,60
= 1775224.66
S‴3 = (0.1) 1789606,08 + (1 – 0.1) 1776886,60

= 1778158.55..………..etc

After obtaining the results of the three smoothing, the equation is then used to find the constant value (at) as follows:

at = 3S′t - 3S″t + S‴t (4)
at1 = 3(1775040) - 3(1775040) + 1775040
= 1775040
at2 = 3(1793506,00) - 3(1776886,60) + 1775224,66
= 1825082,86
at3 = 3(1904081,40) - 3(1789606,08) + 1778158,55
= 2121584,51..………………etc

After getting the constant value (at) then do the equation to find the slope value (bt) and the ct value as follows:

bt = a / 2(1- a)(6 - 5. a)S' - (10-8. a) S" + (4- 3. a) S‴t (5)
bt1 = (0,1/2) x (1 - 0.1)² x ((6 - (5 x 0,1)) x 1775040 - ((10 – (8 x 0,1)) x 1775040+ ((4 – (3 x 0,1) x 1775040))
= 1775040
bt2 = (0,1/2) x (1 - 0.1)² x ((6 - (5 x 0,1)) x 1793506,00 - ((10 – (8 x 0,1)) x 1776886,60 + ((4 – (3 x 0,1) x 1775224,66))
= -267772.96
bt3 = (0,1/2) x (1 - 0.1)² x ((6 - (5 x 0,1)) x 1904081,40 - ((10 – (8 x 0,1)) x 1789606,08 + ((4 – (3 x 0,1) x 1778158,55)) = -245631.81..………………etc

ct = a 2 (1- a) 2 (S′t - 2S″t + S‴t) (6)

ct1 = 0,1² / (1 – 0,1)² x (1775040 – (2 x 775040) + 1775040)
= 0
ct2 = 0,1² / (1 – 0,1)² x (1793506,00 – (2 x 1776886,60) + 1775224,66)
= 184.66
ct3 = 0,1² / (1 – 0,1)² x (1904081,40 – (2 x 1789606,08) + 1778158,55)
= 1271.95..……….etc

The next step is to determine the production forecasting value for the next period followed by determining the MAPE value to determine the accuracy of the prediction results.

Ft + m = at + bt (1) + ½ ct (1) (7)
Ft + m = 2999947.48 + -248536.92 (1) + (1,89 x 3389.71 (1)²)
= 2753105.41

The results of palm oil (CPO) forecasting using the PT.Lonsum Triple Exponential Smoothing method for the January 2025 period produced 2753105 Kg of Crude Palm Oil (CPO) with the MAPE values :

$MAPE = \frac{1}{n} \sum^{n} i{=}0 \mid Ft{-}yi\ yi \mid n\ i{=}0 \times 100\%$ (8)
MAPE = 17.34 (Good forecasting)

E Analysis of Forecasting Results

The results of forecasting with the Triple Exponential Smoothing (TES) method show a MAPE value of 17.34%, which according to the MAPE interpretation category is included in Good Forecasting category (10%-20%). However, it is necessary to analyze more deeply the factors that affect the level of accuracy

1) Alpha (α) parameter selection analysis

The value of alpha (α) = 0.1 selected through trial and error is a relatively small value, which indicates that the model gives greater weight to historical data compared to recent data. This small alpha value selection is suitable for data that has high random fluctuations, as seen in PT Lonsum's CPO production data which has significant monthly variations. To validate this parameter selection, a comparison was made with several other alpha values

TABLE V. COMPARISON OF OTHER ALPHA VALUES

Alpha	MAPE	Interpretation
0.1	17.34%	Good forecasting
0.2	19.26%	Good forecasting
0.3	21.58%	Reasonable forecasting
0.4	23.75%	Reasonable forecasting

From this comparison, it can be seen that the alpha = 0.1 value does provide the best MAPE results, validates the accuracy of the parameter selection.

2) Model and Data limitation

Although the forecasting results show a good level of accuracy, there are some limitations in the model and data that need to be considered. The model only uses 3 years of data (2020-2022) which may not adequately represent the long-term pattern of CPO production, especially given the consistent downward trend in production over the period.

Triple Exponential Smoothing model does not explicitly take into account external factors such as extreme weather changes, company policy changes, or market conditions that may affect CPO production. Historical data shows a downward trend in production from year to year, which may affect the forecasting results if this trend does not continue in the future. The use of a

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2358, Copyright ©2025
Submitted : May 5, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 28, 2025

single alpha value (α = 0.1) for all model components (level, trend, and seasonality) may not be optimal, as each component can have different dynamic characteristics.

F System Implementation

System implementation using the Rapid Application Development (RAD) approach, which focuses on accelerating the development process through continuous iterations. The implementation stages include historical data analysis, user requirements analysis and system modelling using UML and application of Triple Exponential Smoothing (TES) method to generate the accurate forecasting results. After carrying out these stages, it is necessary to implementation and testing of the system where this test aims to check that the system running process is in accordance with user needs and to check whether the system running is in accordance with the planning in the research method used.

- *Login View*

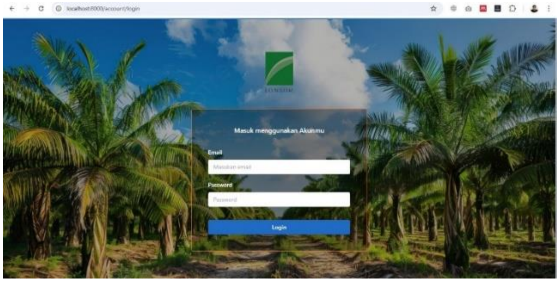


Fig. 5. Login view

On this view users must enter their username and password in order to have access to the system. There are 3 users with each username and password that have been registered : Admin, Supervisor and Manager. Users must ensure that the username and password entered are correct. The login page also limits each user to different access rights that have been determined by the system.

- *Dashboard View*

Fig. 6. Dashboard View

On the dashboard view there is a comparison graph of actual data and forecasting results that have been carried out. there is also a resource graph : CPO and FFB. each user can access the dashboard view.

- *Page as Admin*

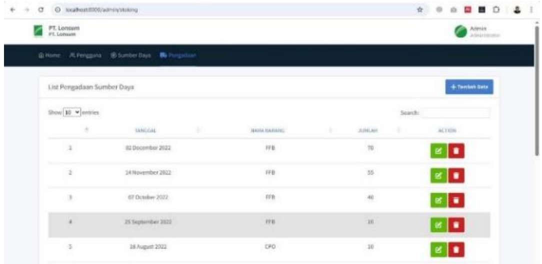


Fig. 7. Resource procurement view

After logging in as Administrator, user is directed to several features : dashboard, user management, resources and resource procurement. User can management system user : edit, update and delete user data that has access to the system. On the resource view and resource procurement view, users can add resources and input historical production data. User can also view historical data report.

- *Page as Supervisor*

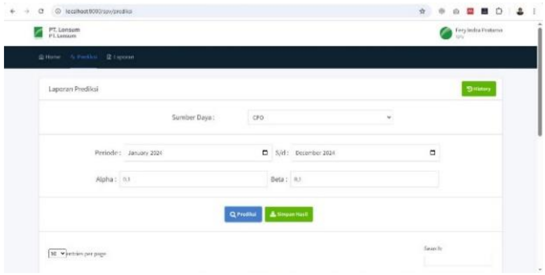


Fig. 8. Forecasting process view

After logging in as a production supervisor, user is directed to the dashboard display in the form of a graph. users can access the forecasting view, historical data report and forecasting data report.

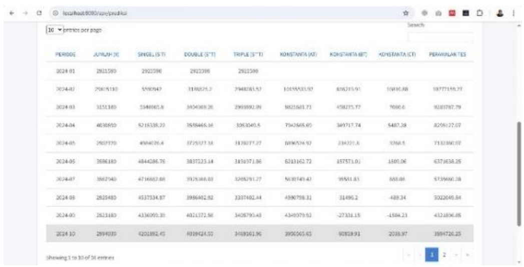


Fig. 9. Forecasting results view

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2358, Copyright ©2025
Submitted : May 5, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 28, 2025

- *Page as Manager*

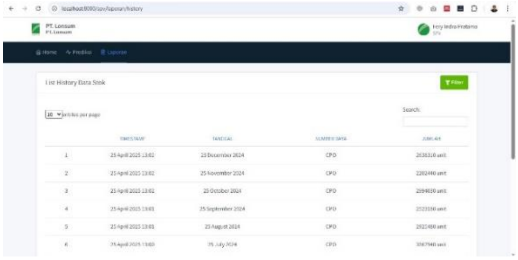


Fig. 10. Historical data report

After logging in as a production manager, user is directed to dashboard display in the form of graph. Users can access Forecasting data report and Historical data report.

IV. CLOSING

The implementation of a production prediction system based on the Triple Exponential Smoothing method has been successfully implemented at PT Lonsum Turangi Palm Oil Mill with good results. By using alpha value = 0.1, the system is able to predict palm oil (CPO) production with a good level of accuracy, as evidenced by the MAPE value of 17.34%. The predicted amount of CPO production for the January 2025 period is 2,753,105 Kg. This research contributes to the integration of the Triple Exponential Smoothing method into a web-based decision support system specifically designed for the palm oil industry. The system enables better data-driven production planning by visualizing historical data and prediction results in an user-friendly interface, which helps management in optimizing mill operations and meeting market demand.

However, this research has several limitations as follows:

1) The prediction is only based on historical production data without considering external variables such as weather conditions, market prices, and other environmental factors that may affect production.
2) The current system only focuses on prediction for one factory location, so it cannot provide a comprehensive picture of the company's overall production.

For future development, some potential areas that can be explored are:

1) Expansion of the prediction system to multi-production sites in one integrated platform.
2) Integration of adaptive smoothing techniques that can automatically adjust parameters based on recent data patterns
3) Integration of external variables such as weather data, soil conditions, and market factors to improve prediction accuracy
4) Development of a "what-if" analysis module to simulate various production scenarios
5) Implementation of machine learning and artificial intelligence to improve the system's predictive capabilities

With this web-based production prediction system, PT. Lonsum and the palm oil industry are expected can be more responsive to market fluctuations, optimize the use of resources, and ultimately improve profitability and operational sustainability.

REFERENCES

[1] M. Y. R. Siahaan and D. Darianto, "Karakteristik Koefisien Serap Suara Material Concrete Foam Dicampur Serat Tandan Kosong Kelapa Sawit (TKKS) dengan Metode Impedance Tube," *J. Mech. Eng. Manuf. Mater. Energy*, vol. 4, no. 1, pp. 85–93, 2020, doi: 10.31289/jmemme.v4i1.3823.

[2] A. E. Batubara, M. F. Yahya, R. Nasyaa, and P. R. Silalahi, "Analisis Ekspor Impor Kelapa Sawit Indonesia Dalam Meningkatkan Pertumbuhan Ekonomi," Jurnal Manajemen, 2023.

[3] E. Lette, M. Zunaidi, and W. R. Maya, "Prediksi Penjualan Crude Palm Oil (CPO) Menggunakan Metode Regresi Linear Berganda," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 3, p. 128, 2022, doi: 10.53513/jursi.v1i3.5106.

[4] M. I. Wiladibrata, N. Azizah, and K. Rifai, "Smoothing dengan Algoritma Golden Section," pp. 507–511, 2022.

[5] P. A. Qori, D. S. Oktafani, and I. Kharisudin, "Analisis Peramalan dengan Long Short Term Memory pada Data Kasus Covid-19 di Provinsi Jawa Tengah," *Prism. Pros. Semin. Nas. Mat.*, vol. 5, pp. 752–758, 2022, [Online]. Available: https://journal.unnes.ac.id/sju/prisma/article/view/54319

[6] M. A. Siregar and N. B. Puspitasari, "Peramalan Hasil Produksi Minyak Kelapa Sawit PT . Bakrie Pasaman Plantations Dengan Metode Holt-Winter ' S Exponential Smoothing," *Ind. Eng. Online J.*, vol. 12, no. 2, p. 10, 2023.

[7] Anisah Anisah and Hadita Hadita, "Penerapan Metode Forecasting Dalam Menentukan Persediaan Kopi Susu Pada Usaha Mikro Kecil Menengah Dalam Hal Ini Sir Coffeehouse Bekasi," *J. Manag. Creat. Bus.*, vol. 2, no. 1, pp. 97–107, Jan. 2024, doi: 10.30640/jmcbus.v2i1.2070.

[8] M. N. Akhtar, E. Ansari, S. S. N. Alhady, and E. Abu Bakar, "Leveraging on Advanced Remote Sensing- and Artificial Intelligence-Based Technologies to Manage Palm Oil Plantation for Current Global Scenario: A Review," Feb. 01, 2023, *MDPI*. doi: 10.3390/agriculture13020504.

[9] S. Madianto, E. Utami, and A. D. Hartanto, "Algoritma Triple Exponential Smoothing Untuk Prediksi Trend Turis Pariwisata Jatim Park Batu saat Pandemi Covid-19," 2021. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[10] J. Ryan and H. Wijaya, "Implementasi Data Mining untuk Sales Forecasting Berbasis Website dengan Metode ARIMA," *bit-Tech*, vol. 7, no. 1, pp. 19–27, 2024, doi: 10.32877/bt.v7i1.1332.

[11] R. Nelfi Yolanda, D. Rahmi, A. Kurniati, S. Yuniati, J. H. Pendidikan Matematika Fakultas Tarbiyah dan Keguruan Universitas Islam Negeri Sultan Syarif Kasim Riau Jl Soebrantas NoKm, and T. Karya Kec Tampan Riau, "Penerapan Metode Triple Exponential Smoothing dalam Peramalan Produksi Buah Nenas di Provinsi Riau," *J. Teknol. dan Manaj. Ind. Terap. (JTMIT*, vol. 3, no. 1, pp. 1–10, 2024.

[12] R. S. Dewi, I. Jaya, and I. Husein, "Peramalan Penerimaan Pajak Kendaraan Bermotor Menggunakan Metode Triple Exponential Smoothing di Sumatera Utara," *Prox. J. Penelit. Mat. dan Pendidik. Mat.*, vol. 7, no. 2, pp. 572–583, 2024, doi: 10.30605/proximal.v7i2.3724.

[13] R. A. Sandika, S. K. Gusti, L. Handayani, and S. Ramadhani, "Implementasi Triple Exponential Smoothing dan Double Moving Average Untuk Peramalan Produksi Kernel Kelapa Sawit," *J. Inf. Syst.*

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2358, Copyright ©2025
Submitted : May 5, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 28, 2025

Res., vol. 4, no. 3, pp. 883–893, 2023, doi: 10.47065/josh.v4i3.3359.

[14] R. B. Saputro, K. P. Kartika, and W. D. Puspitasari, "Implementation of the Triple Exponential Smoothing Method for Predicting Helmet Sales," *JOINCS (Journal Informatics, Network, Comput. Sci.*, vol. 5, no. 2, pp. 30–34, 2022, doi: 10.21070/joincs.v5i2.1607.

[15] R. N. Puspita, "Peramalan Tingkat Pengangguran Terbuka Provinsi Banten Dengan Metode Triple Exponential Smoothing," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 3, no. 2, pp. 358–366, 2022, doi: 10.46306/lb.v3i2.138.

[16] F. Wajdi *et al.*, *Metode Penelitian Kuantitatif*, vol. 7, no. 2. 2024.

[17] D. Murdiani and M. Sobirin, "Perbandingan Metodologi Waterfall Dan RAD Dalam Pengembangan Sistem Informasi," *JINTEKS (Jurnal Inform. Teknol. dan Sains)*, vol. 4, no. 4, pp. 302–306, 2022, [Online]. Available: http://www.jurnal.uts.ac.id/index.php/JINTEKS/article/view/2008

[18] Nurman Hidayat and Kusuma Hati, "Penerapan Metode Rapid Application Development (RAD) dalam Rancang Bangun Sistem Informasi Rapor Online (SIRALINE)," *J. Sist. Inf.*, vol. 10, no. 1, pp. 8–17, 2021, doi: 10.51998/jsi.v10i1.352.

[19] A. K. Nalendra, "Rapid Application Development (RAD) model method for creating an agricultural irrigation system based on internet of things," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1098, no. 2, p. 022103, 2021, doi: 10.1088/1757-899x/1098/2/022103.

[20] N. Purwati, O. R. Fadhlurrahman, D. Iswahyuni, S. Kiswati, and H. Faqih, "Sistem Informasi Cuti Karyawan Menggunakan Berbasis Web dengan Metode Rapid Application Development (RAD)," *Infomatek*, vol. 25, no. 1, pp. 61–68, 2023, doi: 10.23969/infomatek.v25i1.7822.

[21] S. Suendri, T. Triase, and S. Afzalena, "Implementasi Metode Job Order Costing Pada Sistem Informasi Produksi Berbasis Web," *Js (Jurnal Sekolah)*, vol. 4, no. 2, p. 97, 2021, doi: 10.24114/js.v4i2.17954.

[22] A. Azzahra, W. Ramdhan, and W. M. Kifti, "Single Exponential Smoothing: Metode Peramalan Kebutuhan Vaksin Campak," *Edumatic J. Pendidik. Inform.*, vol. 6, no. 2, pp. 215–223, 2022, doi: 10.29408/edumatic.v6i2.6299.

[23] A. T. Hidayat, D. P. Sari, and P. Andriani, "Forecasting Penjualan Produk Sembako Menggunakan Metode Triple Exponential Smoothing," *RESOLUSI Rekayasa Tek. Inform. dan Inf.* , vol. 4, no. 4, pp. 436–445, 2024, [Online]. Available: https://djournals.com/resolusi

[24] R. F. Putri and E. Ekadiansyah, "Metode Triple Exponential Smoothing Dalam Prediksi Persediaan Bahan Baku Pada PT. Bumi Menara Internusa Berbasis Web," *UNES J. Sci. Res.*, vol. 3, no. 1, pp. 81–87, 2022.

[25] I. Yulian, D. S. Anggraeni, and Q. Aini, "Penerapan Metode Trend Moment Dalam Forecasting Penjualan Produk CV. Rabbani Asyisa," *J. Teknol. dan Sist. Inf.*, vol. 6, no. 2, pp. 193–200, 2020.

[26] L. P. Nugraha, R. S. Sianturi, and L. Fanani, "Perancangan Pengalaman Pengguna Aplikasi Knowledge Management System UMKM menggunakan Metode Human Centered Design (Studi Kasus: UMKM Bogor)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 10, pp. 4829–4838, 2022, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11699

[27] F. Oliyan, R. Heriyanto, Y. Septriani, and K. Tania, "Analisis Kebutuhan Pengguna Untuk Perancangan Aplikasi Database Laporan Keuangan Dengan Menggunakan Microsoft Access untuk UMKM," vol. 3, no. 2, pp. 1–10, 2024.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2358, Copyright ©2025
Submitted : May 5, 2025, Revised : May 16, 2025, Accepted : May 18, 2025, Published : May 28, 2025

Comparative Analysis of Random Forest and Support Vector Machine for Sundanese Dialect Classification Using Speech Recognition Features

Abdull Halim Anshor[1]*, Tri Ngudi Wiyatno[2]
Department of Informatics Engineering[1], Department of Industrial Engineering[2]
Universitas Pelita Bangsa
Bekasi, Indonesia
abdulhalimanshor@pelitabangsa.ac.id[1], tringudi@pelitabangsa.ac.id[2]

Abstract— This study investigates the classification of West and South Sundanese dialects using Random Forest (RF) and Support Vector Machine (SVM). Using a dataset of 100 recordings with features extracted via Mel Frequency Cepstral Coefficient (MFCC), models were evaluated by accuracy, precision, recall, and F1-score. Results show RF achieved an accuracy of 93.33%, outperforming SVM's 73.33%. The analysis demonstrates that RF is more reliable in distinguishing dialectal features. This research contributes to regional speech recognition, supporting language preservation and improved dialectal analysis.

Keywords—Classification of Sundanese Dialects, Machine Learning, Random Forest, Support Vector Machine, Mel-Frequency Cepstral Coefficient

I. INTRODUCTION

Machine learning (ML) and speech recognition technologies have advanced significantly in recent years, impacting numerous sectors such as linguistics, healthcare, and information systems [1], [2], [3]. The ability to recognize dialects accurately plays an important role in preserving regional languages and improving personalized speech applications [4], [5]. However, the focus of most speech recognition systems remains on major languages, often neglecting the acoustic complexity of regional dialects like West and South Sundanese [6], [7].

The limitation of existing models to differentiate subtle variations among dialects leads to decreased inclusivity and accuracy in voice-driven systems [8], [9]. In previous studies, Random Forest (RF) and Support Vector Machine (SVM) algorithms have shown effectiveness in various classification tasks. RF has been successfully used for sentiment analysis [10], land use classification [7], speech emotion recognition , and English dialect identification [1]. RF models demonstrate advantages in handling high-dimensional and nonlinear datasets [11], [12], which are common in speech and audio analysis [13], [14].

Support Vector Machine (SVM), on the other hand, has been applied for tasks such as bird species audio classification , gender recognition from voice [14], and emotion expression detection in speech [15]. Studies have indicated that SVM can achieve high accuracy in well-structured datasets, although it may struggle with highly complex patterns compared to ensemble methods like RF [8], [16].

Hybrid and optimized models have also emerged to enhance classification accuracy. For example, CNN-Attention-Optimized RF models have been developed for detecting abusive tweets [5], and multimodal machine learning has been applied to detect markers of mental health through speech [9]. Feature extraction methods, particularly Mel-Frequency Cepstral Coefficients (MFCC), are commonly employed in speech analysis for their ability to capture critical acoustic features [17], [18], [19].

The availability of Sundanese speech datasets [13], opens opportunities for studying regional dialects. However, existing research has focused more on disease detection [11], [16], urban land use mapping [17], action recognition [15], and sentiment analysis [20], leaving a gap in the classification of Sundanese dialects using RF and SVM approaches.

Furthermore, studies have highlighted the importance of adapting classification systems to specific acoustic patterns in regional languages [21], [22], [23]. The combination of feature selection and ensemble models has shown potential for improving classification performance in complex datasets, including speech impairments and voice disorders [24], [22].

Although comparative studies between RF and SVM have been conducted in other domains such as medical diagnostics [24], tree species classification [25], risk analysis in peer-to-peer lending [26], and judicial decision predictions [27], research directly comparing the two algorithms for Sundanese dialect classification remains scarce.

This research contributes to filling this gap by directly comparing RF and SVM in Sundanese dialect classification, which has not been extensively explored in past studies. The novelty lies in the application of MFCC features on local dialect audio combined with classical ML classifiers. Therefore, this study aims to fill this research gap by conducting a comparative analysis of the Random Forest and Support Vector Machine algorithms, using MFCC feature extraction, for the

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

classification of West and South Sundanese dialects.

II. RESEARCH METHODOLOGY

This study implements the Random Forest (RF) and Support Vector Machine (SVM) algorithms for the classification of limited sources of discussion of West and South Sundanese dialects. For this reason, this study is divided into a number of stages consisting of data collection, sound feature actracy, data sharing, model training, and model evaluation using relevant metrics [1] [2]. The research stages are shown in Figure 1.

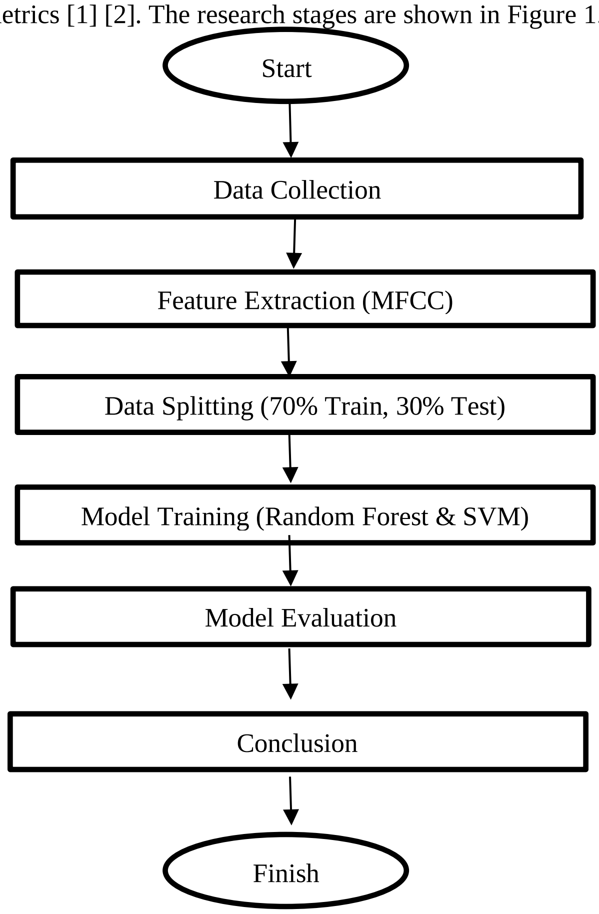


Fig. 1. The research stages

The studied literature has been categorized into three subject categories for clarity: (1) RF/SVM in audio classification, (2) MFCC in dialect detection, and (3) speech recognition in low-resource languages.

A. RF and SVM in Audio Classification

Random Forest (RF) and Support Vector Machine (SVM) are extensively employed in audio classification endeavors. Random Forest (RF) has demonstrated considerable accuracy and resilience in tasks such as dialect identification and emotion detection, attributed to its ensemble characteristics and capacity to handle nonlinear data [1], [18]. SVM demonstrates efficacy with smaller, well-organized datasets and has been employed in applications such as avian species identification, and gender categorization from audio . Both models function as robust baselines for dialect classification [2], [14].

B. MFCC in Dialect Identification

The Mel-Frequency Cepstral Coefficients (MFCC) methodology is a preeminent feature extraction method in speech analysis, esteemed for its capacity to emulate the human auditory system. It catches acoustic subtleties, rendering it especially proficient in dialect and accent categorization [11], [17]. MFCC has been extensively utilized in applications like emotion recognition, speaker verification, and dialect profiling.

C. Speech Recognition in Resource-Scarce Languages

Sundanese is classified as a low-resource language, and current studies suggest that conventional machine learning models such as Random Forest (RF) and Support Vector Machines (SVM) may surpass deep learning models in scenarios with minimal data [9], [21]. Previous research underscores the necessity of employing efficient algorithms and strong feature extraction methods in environments characterized by limited annotated data and significant dialectal complexity.

D. Data Collection

The Sundanese ASR dataset, available at Kaggle, contains a collection of Sundanese-language audio recordings accompanied by their corresponding text transcriptions. This dataset serves as a reliable resource for speech recognition tasks, particularly in dialect classification.

In this study, the data was divided into 70% for training and 30% for testing. This split was chosen to ensure balanced model evaluation, helping to maintain validity while minimizing the risk of overfitting [6].

For the training process, Random Forest (RF) and Support Vector Machine (SVM) algorithms were implemented using features extracted through Mel-Frequency Cepstral Coefficients (MFCC) [7]. The performance of both models was assessed using standard evaluation metrics, including accuracy, precision, recall, and F1-score. Additionally, ROC curves and confusion matrices were employed to provide deeper insights into classification effectiveness and error distribution [1].

E. Feature Extraction with Mel-Frequency Cepstral Coefficients (MFCC)

MFCC is a crucial technique in speech signal processing used to extract the key features of spoken audio. The process begins with pre-emphasis, which amplifies high frequency components, followed by segmenting the signal into frames and applying a Hamming window to ensure smooth transitions between segments. Next, the Fast Fourier Transform (FFT) converts the signal from the time domain into the frequency domain. The Mel scale is then applied to adjust the frequency representation in a way that aligns with human auditory perception. Finally, the Discrete Cosine Transform (DCT) reduces the spectral information into a set of essential coefficients that are efficient for the classification process. [2], [14].

.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

F. *Model Training*

a. *Random Forest (RF)*

Random Forest (RF) is an ensemble-based machine learning algorithm that builds multiple decision trees to improve classification accuracy. Each tree is trained using a random subset of the data and features, thus minimizing the risk of overfitting. Two important parameters in RF are the number of trees (n_trees) and the maximum depth of the tree (max_depth), which are usually adjusted through an optimization process [1], [14].

The RF model combines the results of a number of trees to produce a final prediction. The final prediction is calculated by the decision of combining each tree with an average or majority vote, which can be expressed by equation 1.

$$\hat{y} = \frac{1}{n\_trees} \sum_{i+1}^{n\_trees} Ti(x) \qquad (1)$$

where Ti(x) is the prediction from the i-th tree to the input x. The main parameters of RF are the number of trees (n_trees) and the maximum depth (max_depth), which are set to achieve the best performance [1]. RF has advantages in handling complex data and often shows better performance compared to other algorithms on non-linear data, as proven in a number of previous studies [9].

b. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) is a classification algorithm that works by determining the best hyperplane to separate two groups of data in a high-dimensional space. This algorithm is known to be reliable in handling linear and non-linear data, thanks to its ability to utilize kernel functions. In this context, linear kernels and Radial Basis Function (RBF) are used to test the performance of SVM on voice data in various dialects. [21]. The decision function in the SVM method can be explained through Equation 2

$$f(x) = sign\left(\sum_{i-1}^{N} \alpha iyiK(xi,x) + b\right) \qquad (2)$$

Where:
α i is the coefficient obtained during the training process and is only evaluated as zero For the support vector,
yi is the class label of the training data,
K(xi,x) is the kernel function that measures the similarity between the training data xi and the input bias x,
B For the Radial Basis Function (RBF) kernel, The kernel function K(xi,x) is defined by equation 3.

$$K(xi,x) = \exp(-\gamma \parallel xi - x \parallel 2) \qquad (3)$$

In the SVM algorithm, there are two main parameters that need to be set, namely the C value and gamma (γ). The C parameter plays a role in controlling the balance between the class separation margin and tolerance for misclassification, while γ regulates the extent to which a data point influences the shape of the decision boundary. In this study, both parameters were optimized with the aim of obtaining the best classification accuracy in recognizing Sundanese dialects [6]

G. *Model Evaluation*

In the classification process, evaluating model performance is not just a formality, but an important part so that we know how well the model reads patterns from complex data, especially if the data distribution is unbalanced. In this study, several indicators were used, ranging from accuracy to ROC-AUC. Indeed, accuracy is often the initial benchmark, but to be honest, this metric can be misleading if one class is much larger in number. Therefore, precision and recall are more reliable, especially to see how often the model actually recognizes important classes. F1-score is also a complement, because it balances the two metrics in one number. Finally, we also need a confusion matrix and ROC curve so that we can see the error map and how well the model distinguishes between West and South Sundanese dialects, especially when comparing the Random Forest and SVM algorithms.

III. RESULT AND DISCUSSION

This section explains the results of the voice data processing process from the West and South Sundanese dialects. The voice data was first converted into numbers using the MFCC method, so that it could be read by a computer. Now, after becoming numbers, the data was processed using two models Random Forest and SVM to help recognize the characteristics of each dialect. The goal is for the system to be able to know the difference between the voices from the two regions more accurately.

A. *Feature Extraction Results Using MFCC*

Using the MFCC feature extraction process, the Sundanese dialect voice signal is converted into numeric data form through several stages, namely pre-emphasis, framing, windowing, fast Fourier transform (FFT), Mel filter, and discrete cosine transform (DCT). From a total of 100 voice recordings analyzed, each produced 13 MFCC coefficient values, which were then used as input for the Random Forest and Support Vector Machine algorithms. Both models are used to distinguish the characteristics between the West and South Sundanese dialects. Details of the results are presented in Table 1.

TABLE I. TRANSFORMATION RESULTS

Reckoning Voice	MFCC1	MFCC2	MFCC3	MFCC4	MFCC5	…….	MFCC13	Dialect
1	-3.245.688	-3.186.747	-3.120.912	-32.146.894	-3.329.107	…….	4.187.491	West
2	-32.848.688	-32.132.829	-320.362.86	-3.292.678	-3.415.152	…….	4.022.270	West
3	-	-	-	-	-	…	-	We

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

	32.676.094	320.413.05	3.182.289	32.322.820	33.684.574	….	6.420.176	st
4	-32.186.078	-3.149.075	-3.134.935	-31.918.075	-3.329.881	…….	-4.886.838	West
5	-3.364.437	-3.273.691	-3.203.315	-32.598.596	-3.320.739	…….	4.307.632	West
6	-31.687.822	-31.824.039	-3.173.723	-32.729.593	-3.377.658	…….	-9.547.366	West
7	-3.305.284	-3.285.394	-3.272.659	-3.361.831	-35.028.644	…….	4.800.604	West
8	-3.085.055	-2.969.825	-29.699.222	-31.208.631	-3.244.687	…….	5.396.771	West
9	-3.262.164	-32.435.082	-3.191.192	-3.271.938	-33.982.604	…….	-2.950.396	West
10	-3.131.350	-2.987.862	-2.969.718	-299.796.63	-30.636.218	…….	7.292.064	West
…	…….	…….	…….	…….	…….	………	…….	…….
100	-3.255.414	-31.221.118	-32.074.761	-3.199.827	-33.019.434	…….	-32.013.347	Shouth

a. *Pre-emphasis*

The purpose of pre-emphasis is to amplify the signal at high frequencies and reduce noise. With a constant filter. The calculation at the pre-emphasis stage is done as follows:

y[0] = 5.5348501
y[1] = y[1]-0.97(y[0]) = (-2.7390197)-0.97(5.5348501) = -8.1078242

The pre-emphasis process as shown in Figure 2 aims to enhance high-frequency components, which are vital for phoneme recognition and improve the performance of Random Forest and SVM in classifying Sundanese dialects. This process adds validity to the extraction and training of the model.

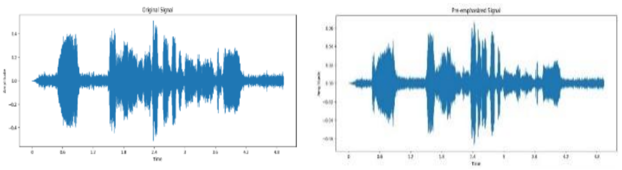


Fig. 2. Comparison Signal Before and After Pre-emphasis

b. Windowing

If the study of the function window is a hamming window, then the n-th function signal data window with the hamming window can be calculated using the n-th function signal data window calculation formula as follows:

w(0) = 0,54 – 0,46 (2π(0) 520)
= 0,54 – 0,46 cos(0) =0.08

After marking the nth window function obtained from the signal data, the windowing formula is used to calculate the windowing result x(n). The windowing result x(n) is obtained by the nth double sign signal frame signal (y(n)) with the window function w(n). The windowing calculation is done as follows:

x(0) =y(0) × w(0)
= (7,13790068) × 0,08 = 5.71032054

In Figure 3, it is shown that the use of Hamming window reduces the amplitude at the edges of the frame, while retaining it in the center, and helps reduce distortion and improve frequency analysis

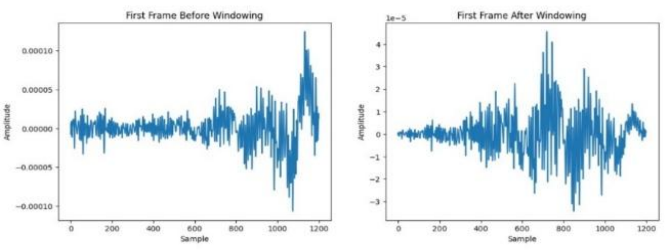


Fig. 3. Effect of windowing process on the first frame

c. Fast Fourier Transform (FFT)

The FFT process converts signals from the time domain to the frequency domain. With the number of samples N = 512, the calculation process in the Fast Fourier Transform process is as follows:

s(0) = ∑ 512−1 n = 0 x (n) e − j 2 πn (0) /512
s(0) = (5.71032054)e− j 2 π (0)(0) /512 + (8.19981570)e− j 2 π (1)(0) /512
+ (6.11131231)e− j 2 π (2)(0) /512 + (1.22343630)e− j 2 π (3)(0) /512
+ (7.83836882)e− j 2 π (4)(0) / 512 + (7.12290406)e− j 2 π (511)(0)/512
s(0) = 7.08910173

Figure 4 shows the amplitude spectrum of the first frame of the audio signal after going through the FFT process. There, the frequency (in Hertz) is displayed on the horizontal axis, while the vertical axis shows how much signal energy appears at each frequency.

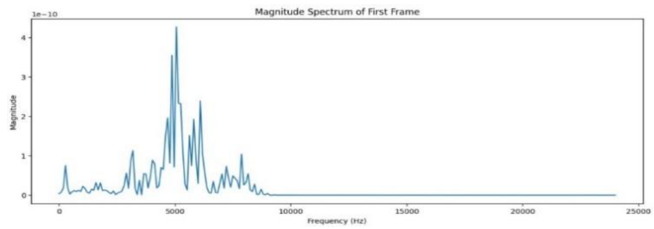


Fig. 4. First frame magnitude spectrum after FFT

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

d. *Mel Filter Bank*

The FFT energy spectrum is filtered using the Mel scale to approximate human hearing perception. The Mel Scale calculation is done as follows:
When zero frequency is entered into the Mel scale formula, the result is indeed zero too this is logical because there is no frequency that can be measured yet. But when we try to enter the number 5000 Hz into the same formula, the result is around 236.4. From there, we can see that in the context of this measurement, the lowest Mel value is 0, and the highest is around 236. So, this scale difference is what will later be used to adjust the perception of sound based on frequency.

e. *Discrete Cosine Transform (DCT)*

DCT can be used to compress frequency information and generate MFCC coefficients. For the calculation of the first coefficient, N is set to 40

$$c(o) = 2\sum_{n=0}^{40-1} \text{xncos}\frac{(2n+1)(0)}{2N}$$
$$=2[236.40458953\cos(0) + (-234.26287304\cos(0)) + (-208.0816647\cos(0))$$
$$+\ldots+(164.66021165)\cos(0))] = -41.2797366$$
$$c(0) = (-41.2797366) \times 1/\sqrt{4(40)} = -3.24568825$$

The value of the first DCT coefficient is -3.24568825. The distribution of MFCC coefficients, which reflects the acoustic characteristics, is illustrated in Figure 5

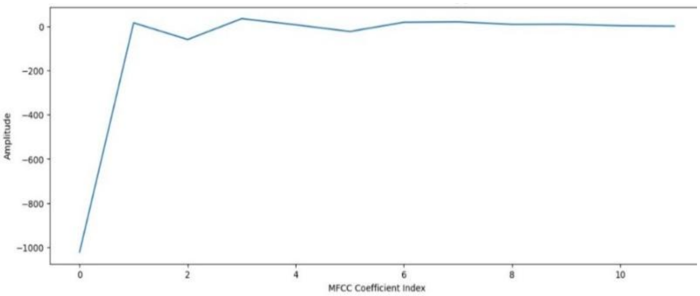


Fig. 5. Distribution of MFCC coefisients after discrete cosine transform

B. Classification Results Using Random Forest (RF) and Support Vector Machine (SVM) Algorithms

The performance of both models was evaluated using accuracy, precision, recall, and F1-score metrics. Additionally, the balance between true positive detection and false positive errors was analyzed using the ROC curve, while the overall quality of the models was assessed by measuring the Area Under the Curve (AUC).

a. *Analysis of Random Forest (RF) Algorithm Results*

The Random Forest model recorded an accuracy of 93.33%, with 28 out of 30 predictions correctly classified. For the West Sundanese dialect, the precision reached 0.89, recall 1.00, and F1-score 0.94. Meanwhile, for the Southern dialect, the precision was perfect at 1.00, recall 0.85, and F1-score 0.92. The ROC curve shows that this model is able to distinguish the two dialects well, as seen from the high true positive rate and low false positive rate at various decision thresholds.

Based on results displayed In Figure 6 Confusion Matrix for RF, we can understand how is this model succeed classifying voice data dialect in a way accurate . The RF model shows very high accuracy , namely 93.33%, which means that Of the 30 predictions made , 28 were correct . classified with Correct

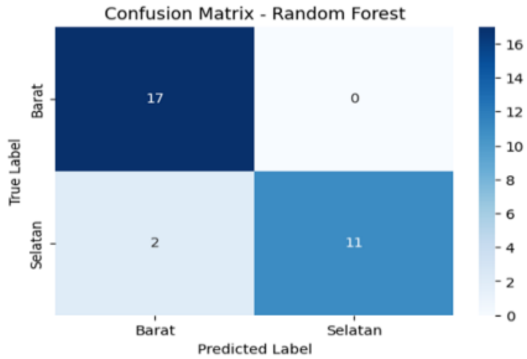


Fig. 6. Confusion matrix Random Forest

The following data is the result of the Random Forest method classification, shown in Figure 7.

Classification Report - Random Forest

	precision	recall	f1-score	support
Barat	0.89	1.00	0.94	17
Selatan	1.00	0.85	0.92	13
accuracy			0.93	30
macro avg	0.95	0.92	0.93	30
weighted avg	0.94	0.93	0.93	30

Fig. 7. Classification results with Random Forest

Figure 8 shows that Random Forest effectively differentiates West and South Sundanese dialects with high precision, recall, and F1-score. The model achieves 93% accuracy, with a precision of 0.89 and recall of 1.00 for the Western dialect, and precision of 1.00 with recall of 0.85 for the Southern dialect. The F1-scores (0.94 for West, 0.92 for South) indicate balanced performance. The ROC curve further confirms its strong classification ability..

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

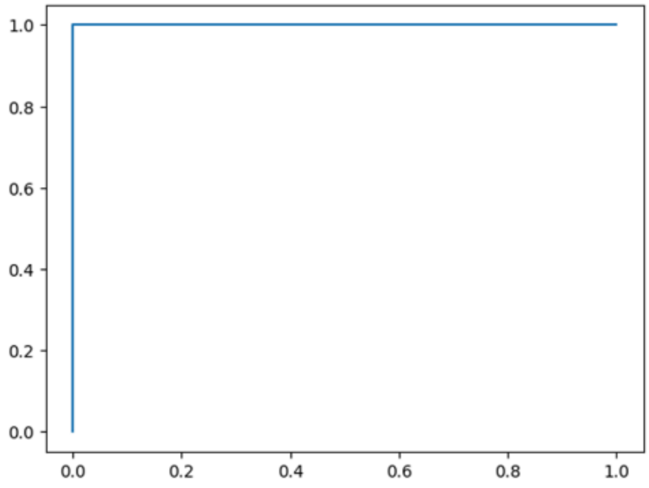


Fig. 8. ROC Curve for classification model with Random Forest

From the ROC curve above , it can be seen that the model has excellent performance , marked with almost curve approach corner left on graph . This is show that the model can differentiate second class with great accuracy on a wide range of decision threshold . The ROC line that reaches TPR value 1.0 without a significant increase in FPR indicates optimal performance , which means this model own level error very low prediction .

b. Analysis of Support Vector Machine (SVM) Algorithm Results

SVM managed to correctly recognize 14 samples from the Western dialect and 8 from the Southern. However, it still made some mistakes three samples that should have been negative were marked as positive, and five were missed entirely. This suggests that the model isn't very sensitive to the sound patterns of the Southern dialect. On the other hand, when applied to the Western dialect, its performance was fairly steady, with a precision of 0.74 and recall of 0.82. In contrast, the recall for the Southern dialect dropped to 0.62, which also caused its F1-score to fall to 0.67. The AUC came out to 0.81 decent, but not outstanding. Compared to Random Forest, SVM clearly fell short. RF delivered a higher accuracy of 93.33%, though it did require slightly more time and Recall to complete the process.

Different with RF, SVM shows more performance low in classification voice dialect , with accuracy overall 73.33%. Difference This possibility big due to the limitations of SVM in handle complex data distribution , especially when there is variation frequency non - linear sound . Figure 9 shows the Confusion Matrix for the applied Support Vector Machine (SVM) model. in classification West and South Sundanese dialects . Matrix This illustrate number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) in classification performed by the model.

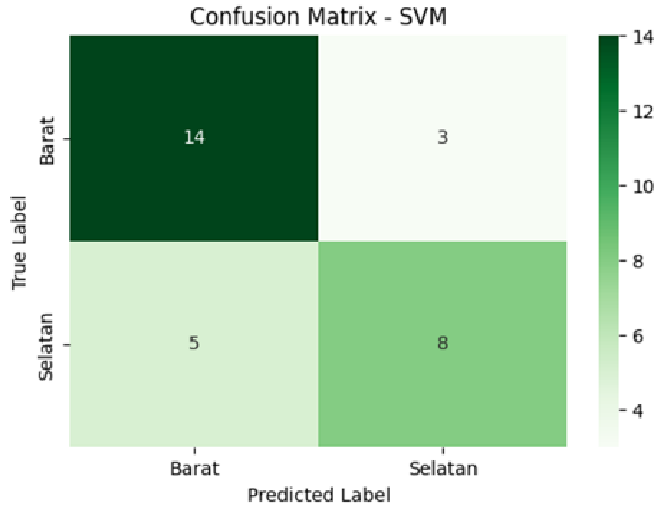


Fig. 9. Confusion matrix SVM

From the Confusion Matrix above , it can be seen that the SVM model has 14 True Positives (TP) for Western dialect and 8 True Positives (TN) for Southern dialect , showing correct prediction on both class . This model also recorded 3 False Positives (FP) for the West and 5 False Negatives (FN) for the South, indicating a number of error classification , especially in recognize Southern dialect . Although the SVM model has good performance in a way overall , the number of FNs is more high in the South class shows that the model is a bit difficulty in detect all example Southern dialect with appropriate .

The classification results of the Support Vector Machine (SVM) method are shown in Figure 10.

Classification Report - SVM

	precision	recall	f1-score	support
Barat	0.74	0.82	0.78	17
Selatan	0.73	0.62	0.67	13
accuracy			0.73	30
macro avg	0.73	0.72	0.72	30
weighted avg	0.73	0.73	0.73	30

Fig. 10. Classification results with Support Vector Machine

The SVM model performs better on the Western dialect (precision 0.74, recall 0.82, F1-score 0.78) than the Southern dialect (recall 0.62, F1-score 0.67), with an overall accuracy of 73%. Figure 11 presents the ROC Curve, illustrating the model's ability to distinguish both dialects based on TPR and FPR.

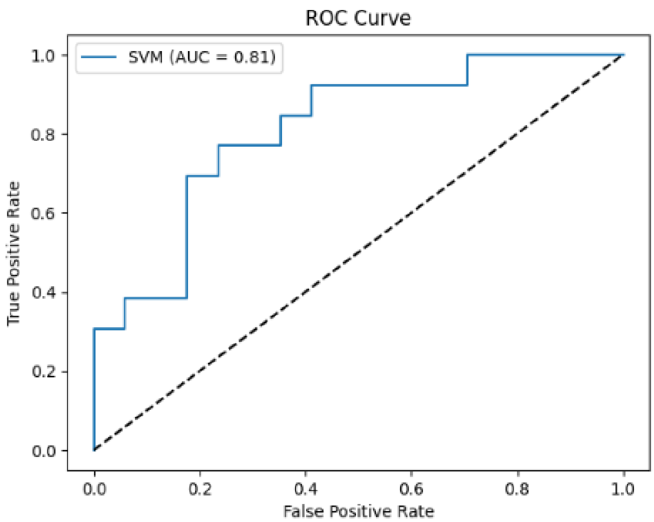


Fig. 11. ROC Curve for classification model with SVM

The ROC Curve in Figure 13 shows that the SVM model has an AUC of 0.81, indicating good classification performance but with room for improvement in distinguishing West and South Sundanese dialects. Table 3 compares RF and SVM based on processing time, accuracy, and memory usage.

TABLE II. COMPARISON RANDOM FOREST AND SVM ALGORITHMS

Algorithm	Processing Time	Accuracy	Memory
Random Forest (RF)	2.3 seconds	93.33%	520 KB
SVM	1.8 seconds	73.33%	480 KB

Table 2 shows the comparison items consists of from Process Time namely the duration required for each algorithm For carry out the classification process , accuracy is level accuracy prediction from each algorithm against test data, memory : Memory is required by every algorithm during the classification process . From the table this , looks that Random Forest has more precision tall compared to SVM, although need little processing time and memory more big.

The performance gap between RF and SVM shows how ensemble techniques can better generalize on constrained, high-dimensional voice data.

Misclassifications identified in the confusion matrix indicate that specific dialectal characteristics may intersect, presenting a problem for boundary-based classifiers such as SVM. These overlaps illustrate the practical challenge of differentiating regional dialects that possess common phonetic origins. Additional enhancements could be realized by integrating a broader range of audio properties or employing more sophisticated models with temporal sensitivity.

IV. CONCLUSION

The results of this study indicate that both Random Forest (RF) and Support Vector Machine (SVM) are effective in classifying West and South Sundanese dialects using MFCC features since RF consistently outperforms SVM across accuracy, precision, recall, and F1-score. RF was particularly consistent in identifying the Southern dialect, indicating its superiority in relation to possible feature change.

Nevertheless, this study has certain restrictions. The small size of the dataset can restrict the model's generalizability. Moreover, the absence of data augmentation or noise-handling methods could affect performance in real-world scenarios where audio quality varies.

Future research is fascinating to explore using deep learning models as convolutional neural networks (CNNs) or recurrent neural networks (RNNs). These models offer more feature learning capability and better scalability for larger and more complex datasets. They can also adapt to minor acoustic patterns that standard machine learning models overlook, hence perhaps enhancing accurate and generalizable dialect detection techniques.

REFERENCES

[1] M. Azhar and H. F. Pardede, “Klasifikasi Dialek Pengujar Bahasa Inggris Menggunakan Random Forest,” J. MEDIA Inform. BUDIDARMA, vol. 5, no. 2, p. 439, Apr. 2021, doi: 10.30865/mib.v5i2.2754.

[2] A. M. Afida, “KLASIFIKASI JENIS BURUNG BERDASARKAN SUARA MENGGUNAKAN ALGORITME SUPPORT VECTOR MACHINE,” MATHunesa J. Ilm. Mat., vol. 8, no. 1, pp. 1–6, Jan. 2020, doi: 10.26740/mathunesa.v8n1.p1-6.

[3] Akhiril Anwar Harahap, R. Novita, T. K. Ahsyar, and Z. Zarnelly, “Classification of Beef and Pork with Deep Learning Approach,” J. Sist. Cerdas, vol. 7, no. 1, pp. 55–65, Apr. 2024, doi: 10.37396/jsc.v7i1.393.

[4] A. Alam, S. Urooj, and A. Q. Ansari, “Design and Development of a Non-Contact ECG-Based Human Emotion Recognition System Using SVM and RF Classifiers,” Diagnostics, vol. 13, no. 12, p. 2097, Jun. 2023, doi: 10.3390/diagnostics13122097.

[5] A. Aljohani, N. Alharbe, R. E. Al Mamlook, and M. M. Khayyat, “A hybrid combination of CNN Attention with optimized random forest with grey wolf optimizer to discriminate between Arabic hateful, abusive tweets,” J. King Saud Univ. - Comput. Inf. Sci., vol. 36, no. 2, p. 101961, Feb. 2024, doi: 10.1016/j.jksuci.2024.101961.

[6] M. A. As Sarofi, I. Irhamah, and A. Mukarromah, “Identifikasi Genre Musik dengan Menggunakan Metode Random Forest,” J. Sains Dan Seni ITS, vol. 9, no. 1, pp. D79–D86, Jun. 2020, doi: 10.12962/j23373520.v9i1.51311.

[7] C. Avci, M. Budak, N. Yağmur, and F. Balçik, “Comparison between random forest and support vector machine algorithms for LULC classification,” Int. J. Eng. Geosci., vol. 8, no. 1, pp. 1–10, Feb. 2023, doi: 10.26833/ijeg.987605.

[8] C. Doğdu, T. Kessler, D. Schneider, M. Shadaydeh, and S. R. Schweinberger, “A Comparison of Machine Learning Algorithms and Feature Sets for Automatic Vocal Emotion Recognition in Speech,” Sensors, vol. 22, no. 19, p. 7561, Oct. 2022, doi: 10.3390/s22197561.

[9] G. Drougkas, E. Bakker, and M. Spruit, “Multimodal Machine Learning for Language and Speech Markers Identification in Mental Health,” Sep. 27, 2024, In Review. doi: 10.21203/rs.3.rs-4925232/v1.

[10] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, “Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB,” J. Inform. Upgris, vol. 7, no. 1, Jun. 2021, doi: 10.26877/jiu.v7i1.7099.

[11] Q. Du, J. Shen, P. Wen, and X. Chen, “Parkinson’s Disease Detection by Using Machine Learning Method based on Local Classification on Class Boundary,” Discov. Appl. Sci., vol. 6, no. 11, p. 576, Oct. 2024, doi: 10.1007/s42452-024-06295-1.

[12] U. Farooq, K. K. S. Reddy, K. S. Shishira, M. G. Jayanthi, and P. Kannadaguli, “Comparing Hindustani Music Raga Prediction Systems using DL and ML Models,” in 2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Applications (ICETCS), Bengaluru, India: IEEE, Apr. 2024, pp. 1–6. doi: 10.1109/ICETCS61022.2024.10543647.

[13] S. Garg and B. Raghavan, “Comparison of machine learning algorithms for the classification of spinal cord tumor,” Ir. J. Med. Sci. 1971 -, vol. 193, no. 2, pp. 571–575, Apr. 2024, doi: 10.1007/s11845-023-03487-3.

[14] R. B. Handoko and S. Suyanto, “Klasifikasi Gender Berdasarkan Suara Menggunakan Support Vector Machine,” Indones. J. Comput. Indo-JC, vol. 4, no. 1, p. 9, Mar. 2019, doi: 10.21108/INDOJC.2019.4.1.244.

[15] O. Peña-Cáceres, H. Silva-Marchan, M. Albert, and M. Gil, “Recognition of Human Actions through Speech or Voice Using Machine Learning Techniques,” Comput. Mater. Contin., vol. 77, no. 2, pp. 1873–1891, 2023, doi: 10.32604/cmc.2023.043176.

[16] S.-M. Jeong, Y.-D. Song, C.-L. Seok, J.-Y. Lee, E. C. Lee, and H.-J. Kim, “Machine learning-based classification of Parkinson’s disease using acoustic features: Insights from multilingual speech tasks,” Comput. Biol. Med., vol. 182, p. 109078, Nov. 2024, doi: 10.1016/j.compbiomed.2024.109078.

[17] M. Kasahun and A. Legesse, “Machine learning for urban land use/ cover mapping: Comparison of artificial neural network, random forest and support vector machine, a case study of Dilla town,” Heliyon, vol. 10, no. 20, p. e39146, Oct. 2024, doi: 10.1016/j.heliyon.2024.e39146.

[18] S. Madanian et al., “Speech emotion recognition using machine learning — A systematic review,” Intell. Syst. Appl., vol. 20, p. 200266, Nov. 2023, doi: 10.1016/j.iswa.2023.200266.

[19] G. H. Mohmad Dar and R. Delhibabu, “Speech Databases, Speech Features, and Classifiers in Speech Emotion Recognition: A Review,” IEEE Access, vol. 12, pp. 151122–151152, 2024, doi: 10.1109/ACCESS.2024.3476960.

[20] A. Omar and T. Abd El-Hafeez, “Quantum computing and machine learning for Arabic language sentiment classification in social media,” Sci. Rep., vol. 13, no. 1, p. 17305, Oct. 2023, doi: 10.1038/s41598-023-44113-7.

[21] D. Subhash, J. L. G., P. B., and V. Ravi, “A robust accent classification system based on variational mode decomposition,” Eng. Appl. Artif. Intell., vol. 139, p. 109512, Jan. 2025, doi: 10.1016/j.engappai.2024.109512.

[22] M. Ur Rehman, A. Shafique, Q.-U.-A. Azhar, S. S. Jamal, Y. Gheraibia, and A. B. Usman, “Voice disorder detection using machine learning algorithms: An application in speech and language pathology,” Eng. Appl. Artif. Intell., vol. 133, p. 108047, Jul. 2024, doi: 10.1016/j.engappai.2024.108047.

[23] N. Widjiyati, “Implementasi Algoritme Random Forest Pada Klasifikasi Dataset Credit Approval,” J. Janitra Inform. Dan Sist. Inf., vol. 1, no. 1, pp. 1–7, Apr. 2021, doi: 10.25008/janitra.v1i1.118.

[24] N. Zhantileuov and S. Ospanov, “A Comparative Study of Supervised Machine Learning and Deep Learning Techniques with Feature Selection Methods for Classifying Parkinson’s Disease Based on Speech Impairments,” in 2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan: IEEE, May 2024, pp. 124–129. doi: 10.1109/SIST61555.2024.10629274.

[25] E. Raczko and B. Zagajewski, “Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images,” Eur. J. Remote Sens., vol. 50, no. 1, pp. 144–154, Jan. 2017, doi: 10.1080/22797254.2017.1299557.

[26] E. Renata and M. Ayub, “Penerapan Metode Random forest untuk Analisis Risiko pada dataset Peer to peer lending,” J. Tek. Inform. Dan Sist. Inf., vol. 6, no. 3, Dec. 2020, doi: 10.28932/jutisi.v6i3.2890.

[27] A. B. Dina, R. Sarno, R. N. E. Anggraini, A. T. Haryono, and A. F. Septiyanto, “Comparison of Oversampling Techniques in Prediction Judicial Decisions of Divorce Trials in Family Courts,” in 2024 International Conference on Information Technology Research and Innovation (ICITRI), Jakarta, Indonesia: IEEE, Sep. 2024, pp. 13–18. doi: 10.1109/ICITRI62858.2024.10699016.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2347, Copyright ©2025
Submitted : April 30, 2025, Revised : May 9, 2025, Accepted : May 14, 2025, Published : May 26, 2025