Comparison of the Performance of the C.45 Algorithm with Naive Bayes in Analyzing Book Borrowing at the Library Pringsewu Muhammadiyah University

Dani Wilian[1]*, Sriyanto[2]
Departmen of Computer Science[1], [2]
Institut Informatika dan Bisnis Darmajaya
Lampung, Indonesia
wilian.2321210009p@mail.darmajaya.ac.id[1], sriyanto@darmajaya.ac.id [2]

***Abstract* — This study examines the effectiveness of the Naïve Bayes and C4.5 algorithms in analyzing book borrowing patterns at the Pringsewu Muhammadiyah University Library. As libraries increasingly serve as vital educational hubs, understanding user borrowing behavior is essential for effective collection management and service enhancement. The research follows the Cross-Industry Standard Process for Data Mining (CRISP-DM), which includes stages of business understanding, data understanding, preparation, modeling, evaluation, and implementation. A dataset consisting of 5,586 records and ten attributes related to book lending was utilized, with comprehensive data cleaning and preprocessing conducted. The performance of both algorithms was assessed using K-fold cross-validation, yielding an accuracy of 96.26% for C4.5, compared to 91.44% for Naïve Bayes. These results demonstrate that C4.5 is more adept at capturing complex relationships within the data, providing deeper insights into user preferences and enhancing library services. This research underscores the potential of data mining techniques to optimize library management and proposes avenues for future investigation, such as exploring advanced machine learning algorithms and expanding datasets for use in broader library contexts.**

Keywords*— *Book Borrowing Patterns, C4.5, Naive Bayes, Datamining

I. INTRODUCTION

In today's digital era, the library does not function as a place to store my books, but also as an important source of information and education for students, lecturers and researchers. With the increasing number of books and other materials available, and the diversity of users' information needs, managing library collections effectively has become a significant challenge[1]. One important aspect in library management is understanding book borrowing patterns by users.

Book borrowing patterns can provide valuable insight into user preferences, collection usage trends, and information needs that may be unmet[2]. However, with large data volumes and complexity, manual analysis becomes less efficient and prone to errors[3]. Muhammadiyah Pringsewu University as a higher education institution also faces challenges in managing its library so that it can meet the information needs of the entire academic community. This is where data mining plays an important role. By utilizing data mining to analyze book borrowing patterns, libraries can be more effective in developing collection management strategies and improving service quality. In recent years, data mining has become a powerful tool for enhancing library management by providing detailed analyses of user behavior and collection utilization. Several studies have examined various data mining techniques, such as association rule mining, clustering, and classification algorithms, to identify hidden patterns in library usage data. Researchers have found that algorithms like Naïve Bayes, Decision and K-means clustering can offer valuable insights into user preferences and improve services by enabling targeted recommendations and collection development. Despite these advancements, gaps remain in exploring the performance differences among these algorithms specifically in university libraries, as well as in developing strategies for real-time, automated recommendations. This research aims to bridge these gaps by focusing on the comparative effectiveness of Naïve Bayes and C4.5 in the context of a university library.

Data mining is a technique that can be used to extract valuable information from large and complex data[4][5]. In a library context, data mining can help identify hidden patterns in book lending data, which can then be used to improve library services, such as collection management, personalization of book recommendations, and budget planning for procuring new books[6].

To analyze book borrowing patterns, various data mining algorithms can be used, including the Naïve Bayes and C4.5 algorithms[7]. These two algorithms have different approaches to processing data and making predictions, so it is important to compare to determine which algorithm is more effective in a particular context. Naïve Bayes is an algorithm based on Bayes' theorem, which uses probability to make predictions[8]. This algorithm is simple, fast, and frequently used in text classification and pattern analysis. C4.5 is a decision tree-based machine learning algorithm that generates decision trees from training data[9]. These algorithms are known for their ability to handle varying data and provide easy-to-understand interpretations in the form of decision trees.

The application of data mining in library systems can

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i1.2300, Copyright ©2025
Submitted : October 14, 2024, Revised : November 12, 2024, Accepted : November 14, 2024, Published : January 31, 2025

help identify hidden book borrowing patterns and provide more personalized recommendations to library users[10]. Analyzing user behavior through data mining techniques can improve library collection management by understanding user preferences based on lending data[11]. The Naïve Bayes algorithm has advantages in text classification, especially because of its simplicity and speed in processing data, which makes it suitable for applications with large data volumes[12]. Naïve Bayes is effective in library book recommendation systems because this algorithm can accurately predict book categories of interest based on borrowing history[13]. The C4.5 algorithm is very effective in producing decision trees that are easy to interpret, especially in complex data analysis such as book borrowing patterns in libraries[14]. The C4.5 algorithm is able to handle varied data well and provides more interpretive results than other algorithms in analyzing library user behavior.

Therefore, this research aims to compare the performance of the Naïve Bayes and C4.5 algorithms in analyzing book borrowing patterns at the Pringsewu Muhammadiyah University Library. It is hoped that the results of this research can provide recommendations regarding the most appropriate methods to be implemented in library information systems in the future.

II. METHODOLOGY

A. Research Stage

The method used in this research follows the stages of the Cross-Industry Standard Process for Data Mining (CRISP-DM) model[15]. The research stages can be seen in Figure 1 below:

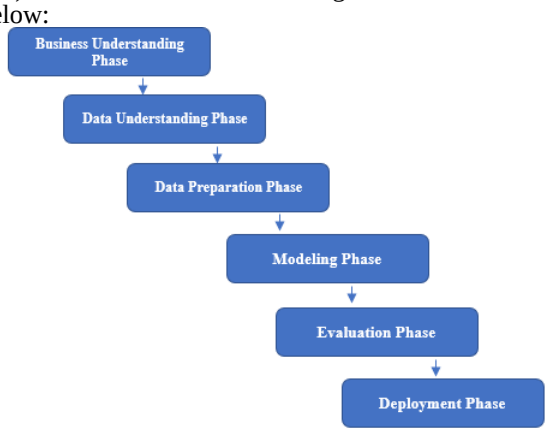


Figure 1 Research Flow

- *Business Understanding Phase*

At this stage the focus is on the research objective, namely to find out the best algorithm for analyzing book borrowing patterns in libraries by looking at what books are the most popular among borrowers, which students from study programs borrow books most often, then what are the book borrowing patterns over a certain period of time , so that the best model is obtained to fulfill the research objectives.

- *Data Understanding Phase*

The data that will be used in the research is Pringsewu Muhammadiyah University library information system data. There are 10 attributes that will be used in this research

Namely Borrower ID, Gender, Membership Type, Copy Code, Book Category, Length of Membership, Department/Faculty, Borrow Date, Return Date, Status.

- *Data Preparation*

Data preparation is one of the important stages in the CRISP-DM process which ensures that the data to be used for analysis is in optimal condition. Prepare data for analysis. This includes data cleaning, feature selection, and data transformation. After the data preparation stage is complete, the data will be ready to be used in the modeling stage. A careful data preparation process is essential to ensure that the analysis and models built are accurate and relevant.

- *Modeling (Modeling Phase)*

The algorithms used in this research are the C4.5 and Naive Bayes algorithms to classify book borrowing patterns at Pringsewu Muhammadiyah University and to obtain a model or function to describe graduation predictions by comparing the C4.5 and Naive Bayes algorithms.

- *Evaluation Phase (Evaluation Phase)*

At this stage, the performance evaluation of the two algorithms, namely the C4.5 and Naive Bayes Algorithms, is carried out by comparing the results of the average values of accuracy, recall and error rate contained in the confusion matrix table.

- *Deployment Phase (Deployment Phase)*

After the evaluation stage where the results of a model are assessed in detail, the model performance is monitored periodically and adjusted if necessary. Apart from that, adjustments were also made to the model so that it could produce results that were in line with the initial target of this CRISP-DM stage.

B. Data Collection Stage

Data collection methods are an important thing in research and are strategies or methods used by researchers to collect the data needed in their research. The data collection methods used in this research are:

- *Literature Review (Research Library)*
  The literature review is carried out by reading, quoting and making notes sourced from library materials that support and are related to research, in this case regarding C4.5 and Naive Bayes Algorithm data mining.
- *Field Studies (Field Research)*
  In this research, data collection through documents was carried out by studying the facts or data in the Pringsewu Muhammadiyah University Library information system.

C. Experiment Stages

This research will be carried out by applying two methods, namely the C4.5 Algorithm and Naive Bayes to classify book borrowing patterns.

- Decision tree (C4.5)

C4.5 is a collection of algorithms for classification techniques in machine learning and data mining. The goal is supervised learning, where each tuple in the data set can be described by a set of attribute values, and each tuple belongs to one of many different and incompatible classes[16]. The goal of C4.5 is to learn mappings from attribute values to categories that can be used to categorize unknown items into new categories. J. Rossi Quinlan suggests C4.5 based on ID3. A decision tree is constructed using the ID3 algorithm. A decision tree is a tree structure that is like a flowchart, with each internal node (nonleaf node) representing a test on an attribute, each branch representing a test result, and each leaf node holding a class label. After building a decision tree

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i1.2300, Copyright ©2025
Submitted : October 14, 2024, Revised : November 12, 2024, Accepted : November 14, 2024, Published : January 31, 2025

for tuples that do not provide classification labels, we select a path from the root node to the leaf node, and the path stores the prediction information of the tuple. Decision trees have the advantage of not requiring domain information or parameter configuration, making them ideal for exploratory information mining[17].

The C4.5 algorithm is based on ID3 added to continuous attributes, attribute values, and information processing, by generating a tree to build a pruning decision tree in two stages. For each attribute, with the C4.5 algorithm information calculation, we can find out the Gain Ratio, the rate of information acquisition. Finally, it is selected with the highest level of information gain from the given test set attributes to organize the branch. According to the test attribute values using a recursive algorithm, obtain an initial decision tree. The computational formula related to the C4.5 algorithm is as follows[18]. First, the expectation value required for sample classification is given as follows: Determine the root of the tree by calculating the highest gain value of each attribute or the lowest entropy index value. Previously, the entropy index value was calculated using the formula:

$$Entropy\ (i) = \sum_{j=1}^{m} \quad f(i,j).\ 2f[(i,j)] \tag{1}$$

Gain value using the formula:

$$gain = -\sum_{i=1}^{p} \quad .\ IE(i) \tag{2}$$

To calculate the gain ratio, you need to know a new term called Split Information with the formula:

$$SplitInformation = -\sum_{t=1}^{c} \frac{S1}{S} log2\ \frac{S1}{S} \tag{3}$$

Next, calculate the gain ratio

$$Gainratio(S,A) = \frac{\text{Gain(S.A)}}{\text{SplitInformation (S,A)}} \tag{4}$$

Repeat step 2 until all records have been split. The decision tree splitting process ends when:

1. *All tuples in node record m are of the same class.*
2. *The attributes in the dataset are not further divided.*
3. *An empty branch has no records*

- **Naïve Bayes**

Naive Bayes is a probabilistic classification algorithm that utilizes Bayes' theorem to classify data. This algorithm is called "naive" because it assumes that all features in the dataset are independent of each other when assigned a particular class. Although this assumption is often unrealistic in practice, Naive Bayes remains popular and effective for many classification tasks[19].

Bayes' theorem is the basic principle of this algorithm. This theorem can be used to update the probability of a hypothesis based on new evidence. In the context of classification, the hypothesis is the class we are trying to predict, and new evidence is the observed features of the data.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{5}$$

Where:

- *P(C/X): Posterior probability of the class C is given a feature X*
- *P(X/C): Feature probability X is given a class C (likelihood).*
- *P(C): Class prior probability C*
- *P(X): Marginal probability of a feature X*

D. Performance Evaluation

- K-fold Cross Validation

k-fold cross-validation is a technique for validating the accuracy of a model built on a certain data set, which divides the data set into two parts, namely training data and testing data. For prediction problems, the model is usually given a dataset of known data to train on (training dataset) and unknown data (or first-time data) to test the model (called validation). or test data)[20]. The goal of cross-validation is to test a model's ability to predict new data that was not used in its evaluation, to flag problems such as overfitting or selection bias, and to provide insight into how the model generalizes to independent data. set (i.e. unknown dataset, e.g. problem).

- Confusion matrices

Confusion matrix is a very popular measure used when solving classification problems. It can be applied to binary classification as well as to multiclass classification problems. This matrix is used to evaluate the performance of the method used after classification. The confusion matrix represents TP values that are correctly classified, FP values in the relevant class when they should be in other classes, and FN values in other classes when they should be in the relevant class and TN values that are correctly classified in other classes[21]. The most frequently used performance metrics for classification according to these values are accuracy (ACC), precision (P), sensitivity (Sn), specificity (Sp), and F-score values. The calculation of these performance metrics according to the values in the confusion matrix is made according to Eq.

$$ACC = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{6}$$

$$P = \frac{\text{TP}}{\text{TP+FP}} \tag{7}$$

$$Sn = \frac{\text{TP}}{\text{TP+FN}} \tag{8}$$

$$Sp = \frac{\text{TN}}{\text{TN+FP}} \tag{9}$$

$$F - score = 2\text{x}\frac{\text{P x Sn}}{\text{P+Sn}} \tag{10}$$

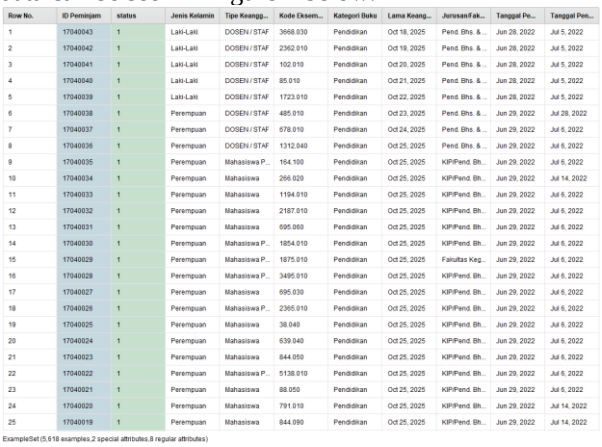
III. RESULTS AND DISCUSSION

A. Analysis and Preprocessing Data

Before applying the algorithm, data collected from the Pringsewu Muhammadiyah University Library was analyzed and processed first. The dataset consists of 5618 data with ten attributes including Borrower ID, Gender, Membership Type, Copy Code, Book Category,

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i1.2300, Copyright ©2025
Submitted : October 14, 2024, Revised : November 12, 2024, Accepted : November 14, 2024, Published : January 31, 2025

Membership Length, Department/Faculty, Borrowing Date, Return Date, and Status. A total of 170 records were removed due to missing values and duplicates identified. For example, all records with an incomplete Borrower ID or missing Return Date will be excluded from the data set. This data can be seen in Figure 2 below.



Row No.	ID Peminjam	status	Jenis Kelamin	Tipe Keangg...	Kode Eksem...	Kategori Buku	Lama Keang...	Jurusan/Fak...	Tanggal Pe...	Tanggal Pen...
1	17040043	1	Laki-Laki	DOSEN / STAF	3668.030	Pendidikan	Oct 18, 2025	Pend. Bhs. & ...	Jun 28, 2022	Jul 5, 2022
2	17040042	1	Laki-Laki	DOSEN / STAF	2362.010	Pendidikan	Oct 19, 2025	Pend. Bhs. & ...	Jun 28, 2022	Jul 5, 2022
3	17040041	1	Laki-Laki	DOSEN / STAF	102.010	Pendidikan	Oct 20, 2025	Pend. Bhs. & ...	Jun 28, 2022	Jul 5, 2022
4	17040040	1	Laki-Laki	DOSEN / STAF	85.010	Pendidikan	Oct 21, 2025	Pend. Bhs. & ...	Jun 28, 2022	Jul 5, 2022
5	17040039	1	Laki-Laki	DOSEN / STAF	1723.010	Pendidikan	Oct 22, 2025	Pend. Bhs. & ...	Jun 28, 2022	Jul 5, 2022
6	17040038	1	Perempuan	DOSEN / STAF	485.010	Pendidikan	Oct 23, 2025	Pend. Bhs. & ...	Jun 29, 2022	Jul 28, 2022
7	17040037	1	Perempuan	DOSEN / STAF	678.010	Pendidikan	Oct 24, 2025	Pend. Bhs. & ...	Jun 29, 2022	Jul 6, 2022
8	17040036	1	Perempuan	DOSEN / STAF	1312.040	Pendidikan	Oct 25, 2025	Pend. Bhs. & ...	Jun 29, 2022	Jul 6, 2022
9	17040035	1	Perempuan	Mahasiswa P...	164.100	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
10	17040034	1	Perempuan	Mahasiswa	266.020	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 14, 2022
11	17040033	1	Perempuan	Mahasiswa	1194.010	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
12	17040032	1	Perempuan	Mahasiswa	2187.010	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
13	17040031	1	Perempuan	Mahasiswa	695.060	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
14	17040030	1	Perempuan	Mahasiswa P...	1854.010	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
15	17040029	1	Perempuan	Mahasiswa P...	1875.010	Pendidikan	Oct 25, 2025	Fakultas Keg...	Jun 29, 2022	Jul 6, 2022
16	17040028	1	Perempuan	Mahasiswa P...	3495.010	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
17	17040027	1	Perempuan	Mahasiswa	695.030	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
18	17040026	1	Perempuan	Mahasiswa P...	2365.010	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
19	17040025	1	Perempuan	Mahasiswa	38.040	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
20	17040024	1	Perempuan	Mahasiswa	639.040	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
21	17040023	1	Perempuan	Mahasiswa	844.050	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
22	17040022	1	Perempuan	Mahasiswa P...	5138.010	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
23	17040021	1	Perempuan	Mahasiswa	88.050	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 6, 2022
24	17040020	1	Perempuan	Mahasiswa	791.010	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 14, 2022
25	17040019	1	Perempuan	Mahasiswa	844.090	Pendidikan	Oct 25, 2025	KIP/Pend. Bh...	Jun 29, 2022	Jul 14, 2022

ExampleSet (5,618 examples,2 special attributes,8 regular attributes)

Figure 2 Dataset

B. Modeling and Evaluation

Both algorithms (C4.5 and Naïve Bayes) were implemented using the preprocessed dataset. The performance of each algorithm was evaluated using K-fold cross-validation (with k=10) to ensure reliable accuracy metrics.

Algorithms C4.5

The application of data in Rapidminer for analyzing book borrowing using the C4.5 algorithm is shown in Figure 3 below:

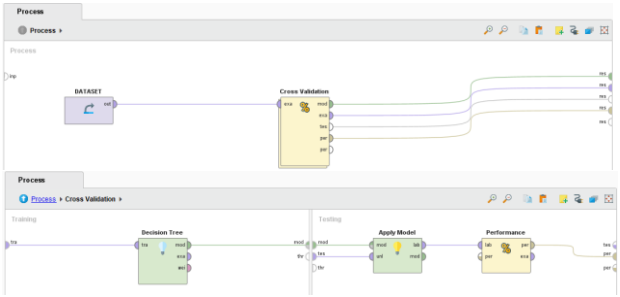


Figure 3 Testing scheme with C4.5

In Figure 3, the prepared dataset is applied to the Rapidminer application by conducting experiments using cross validation which can directly divide the data into training data and testing data because the data used is supervised and the algorithm used is C4.5. We can see the experimental results in Figure 4 below.

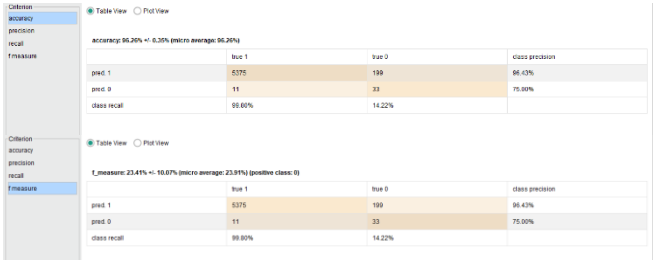


Figure 4 Accuracy Results with the C4.5 Algorithm

In figure 4 The model achieved an overall accuracy of 96.26%. This high accuracy indicates that the model correctly classifies approximately 96 out of every 100 instances, suggesting effective performance in distinguishing between the classes. Decision tree generated by the C4.5 algorithm for classifying book borrowing patterns we can see the experimental results in Figure 5 below.

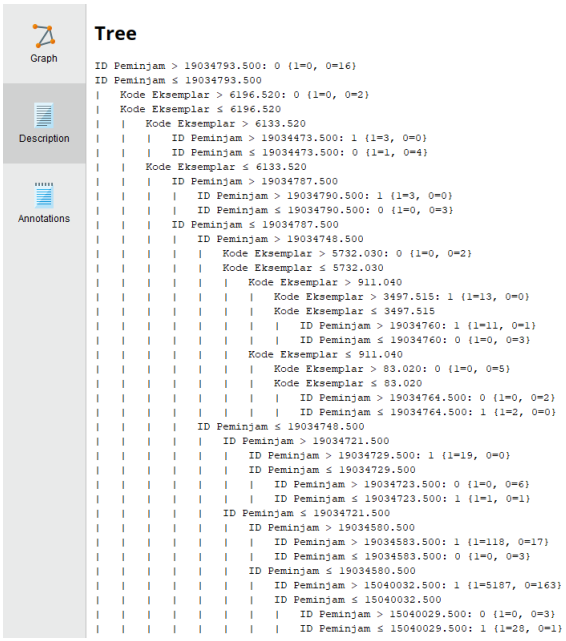


Figure 5 Decision tree

Figure 5 above depicts the decision tree structure used to classify Borrower IDs and Exemplar Codes based on various conditions. Each branch of the tree represents a decision or comparison made against the Borrower ID and Exemplar Code. For example, first a separation is carried out based on Borrower ID > 19034793.500, which then leads to further grouping based on Exemplar Code and Borrower ID values. Each branch contains information regarding the number of borrowers who meet or do not meet the given conditions, with the numbers in brackets indicating the number of borrowers in each category.

This decision tree structure helps in understanding various factors such as "Exemplar Code", "Borrower ID", etc. that influence borrowing patterns and helps librarians in making data-based decisions regarding book management. This can guide collection development, help identify popular books, and understand borrowing behavior at the

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i1.2300, Copyright ©2025
Submitted : October 14, 2024, Revised : November 12, 2024, Accepted : November 14, 2024, Published : January 31, 2025

Pringsewu Muhammadiyah University Library.

Algorithms Naïve Bayes

The application of data in Rapidminer for analyzing book borrowing using the Naïve Bayes algorithm is shown in Figure 6 below.

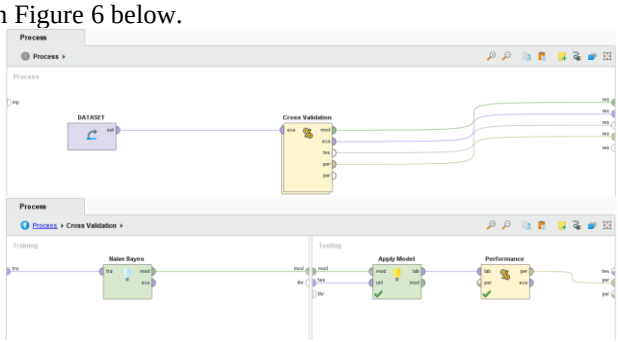


Figure 6 Testing scheme with Naïve Bayes

The experiment in Figure 3.5 is a validation technique for dividing training data and testing data using cross validation techniques. From this experiment, we got the results that we can see in Figure 7 below

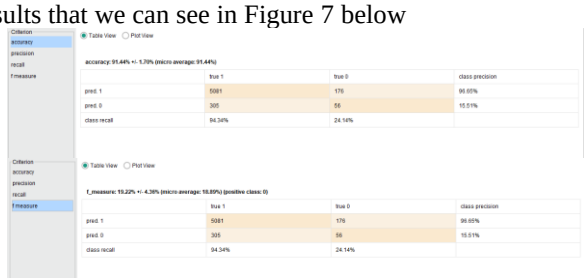


Figure 7 Accuracy Results with the Naïve Bayes Algorithm

In figure 7 The model achieved an overall accuracy of 91.44%. Apart from accuracy in the Naive Bayes algorithm, there is SimpleDistribution where The focus of this model is on the label attribute Status Pengembalian. This attribute likely classifies whether a book has been returned on time or not, reflected by binary classes (1 for returned on time, 0 for not returned on time). we got the results that we can see in Figure 8 below.

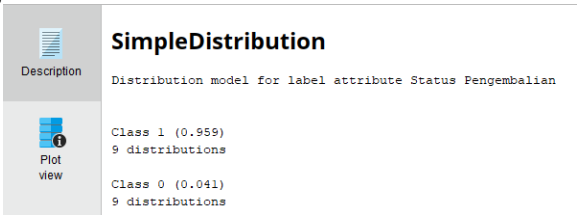


Figure 8 Sample Distribution

In figure 8 The model indicates that Class 1 has a probability of 0.959, suggesting that 95.9% of the instances in the dataset are classified as this class. This is a strong indication that most books are returned on time. In contrast, Class 0 has a probability of 0.041, indicating that only 4.1% of the instances are classified as not returned on time. This suggests that overdue returns are significantly less frequent compared to timely returns. Both classes have 9 distributions listed, which means the model used nine different feature distributions (attributes) to help classify the return status of the books.

C. Results

The results of the model performance are summarized in the following

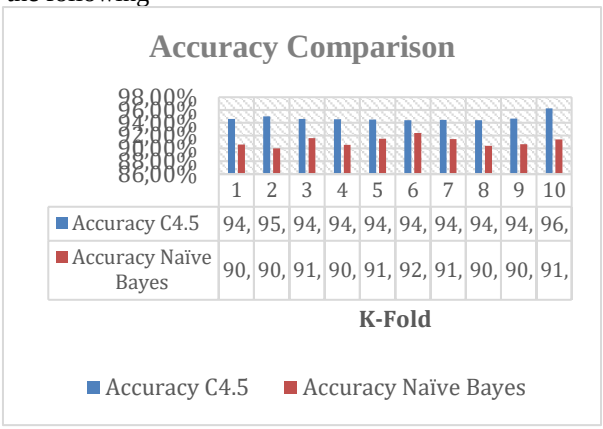


Figure 10 algorithm performance comparison results

Based on the table and diagram image above, it shows that changing the number of K-Folds in Cross Validation when carrying out classification will produce different accuracy so that we can produce the best accuracy. We can see that with K-Fold 10, the accuracy and precision results are the highest, in contrast to the recall results which are not higher than other uses of K-Fold. C4.5 achieved an accuracy of 96.26%, indicating it correctly classified a significant majority of the borrowing patterns. In contrast, Naïve Bayes, with an accuracy of 91.44%, showed a slightly lower capability in predicting user behavior.

D. Discussion

The C4.5 algorithm outperformed the Naïve Bayes algorithm across all evaluated metrics. This can be attributed to C4.5's ability to model complex relationships in the data through its decision tree structure. The tree structure allows for capturing interactions between features, which is particularly useful when analyzing book borrowing patterns that may be influenced by multiple factors (e.g., book categories, user demographics). C4.5 achieved an accuracy of 96.26%, indicating it correctly classified a significant majority of the borrowing patterns. In contrast, Naïve Bayes, with an accuracy of 91.44%, showed a slightly lower capability in predicting user behavior.

The findings emphasize the importance of utilizing data mining techniques for effective library management. By implementing the C4.5 algorithm, the Pringsewu Muhammadiyah University Library can enhance its collection management strategies, personalize recommendations for users, and allocate resources more efficiently based on borrowing trends.

IV. CONCLUSION

The study successfully demonstrated the application of the C4.5 and Naïve Bayes algorithms in analyzing book borrowing patterns at the Pringsewu Muhammadiyah University Library. The results showed that the C4.5 algorithm outperformed Naïve Bayes across all evaluation

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i1.2300, Copyright ©2025
Submitted : October 14, 2024, Revised : November 12, 2024, Accepted : November 14, 2024, Published : January 31, 2025

metrics, including accuracy, precision, recall, and F-score. Specifically, C4.5 achieved an accuracy of 96.26%, while Naïve Bayes reached 91.44%.

This indicates that C4.5 is more effective in capturing complex relationships in the data and predicting user borrowing behavior. The findings of this research underscore the importance of employing data mining methodologies in library management. By implementing the C4.5 algorithm, the Pringsewu Muhammadiyah University Library can improve its services, such as personalized book recommendations, targeted marketing efforts, and more efficient budgeting for new acquisitions. Understanding borrowing patterns also aids in developing strategies to meet the evolving information needs of the academic community.

While this study provides valuable insights, it also highlights several areas for future research. Subsequent studies could explore the application of other advanced machine learning algorithms, such as Random Forest or Support Vector Machines, to compare their performance against C4.5 and Naïve Bayes. Additionally, expanding the dataset to include a more extended time frame and more diverse attributes could enhance the robustness of the analysis and its applicability across different library settings.

ACKNOWLEDGMENT

The author would like to thank all parties who have provided support and assistance in preparing this journal, especially the supervisors and colleagues who have provided valuable input as well as their beloved family who always provide support.

REFERENCES

[1] L. Yulita, A. S. Sunge, and N. Nurhidayanti, "Optimasi Algoritma Genetika Dalam Memprediksi Minat Baca Siswa Pada Perpustakaan SMK Negeri 1 Gantar Dengan Metode Decision Tree."

[2] E. Irfiani, Y. Kusnadi, S. Sunarti, and F. Handayanna, "Implementasi Data Mining dalam Mengklasifikasi Minat Baca Pada Perpustakaan Daerah Menggunakan Algoritma C4.5," *JOINS (Journal of Information System)*, vol. 8, no. 2, pp. 106–114, Nov. 2023, doi: 10.33633/joins.v8i2.8004.

[3] T. T. Sang Nguyen, "Model-based book recommender systems using Naïve Bayes enhanced with optimal feature selection," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, 2019, pp. 217–222. doi: 10.1145/3316615.3316727.

[4] L. Feng, "Research on Higher Education Evaluation and Decision-Making Based on Data Mining," *Sci Program*, vol. 2021, 2021, doi: 10.1155/2021/6195067.

[5] D. Berrar, "Bayes' Theorem and Naive Bayes Classifier," in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Eds., Oxford: Academic Press, 2019, pp. 403–412. doi: https://doi.org/10.1016/B978-0-12-809633-8.20473-1.

[6] J. Moolayil, *Learn Keras for Deep Neural Networks*. Apress, 2019. doi: 10.1007/978-1-4842-4240-7.

[7] S. Xuanyuan, S. Xuanyuan, and Y. Yue, "Application of C4.5 Algorithm in Insurance and Financial Services Using Data Mining Methods," *Mobile Information Systems*, vol. 2022, 2022, doi: 10.1155/2022/5670784.

[8] X. Xiongjun and D. Lv, "The Evaluation of Music Teaching in Colleges and Universities Based on Machine Learning," *Journal of Mathematics*, vol. 2022, 2022, doi: 10.1155/2022/2678303.

[9] S. and Communication Networks, "Retracted: Analysis and Application of Data Mining Technology for College English Education Integration," *Security and Communication Networks*, vol. 2024, pp. 1–1, Jan. 2024, doi: 10.1155/2024/9836129.

[10] W. Liu *et al.*, "Using a classification model for determining the value of liver radiological reports of patients with colorectal cancer," *Front Oncol*, vol. 12, Nov. 2022, doi: 10.3389/fonc.2022.913806.

[11] P. V. Ngoc, C. V. T. Ngoc, T. V. T. Ngoc, and D. N. Duy, "A C4.5 algorithm for english emotional classification," *Evolving Systems*, vol. 10, no. 3, pp. 425–451, Sep. 2019, doi: 10.1007/s12530-017-9180-1.

[12] J. Wang, "Application of C4.5 Decision Tree Algorithm for Evaluating the College Music Education," *Mobile Information Systems*, vol. 2022, 2022, doi: 10.1155/2022/7442352.

[13] Y. A. Alsariera, Y. Baashar, G. Alkawsi, A. Mustafa, A. A. Alkahtani, and N. Ali, "Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance," 2022, *Hindawi Limited*. doi: 10.1155/2022/4151487.

[14] X. Yang and J. Ge, "Predicting Student Learning Effectiveness in Higher Education Based on Big Data Analysis," *Mobile Information Systems*, vol. 2022, 2022, doi: 10.1155/2022/8409780.

[15] Parteek Bhatia, "Data Mining and Data Warehousing," 2019.

[16] B. Mahesh, "Machine Learning Algorithms - A Review," *International Journal of Science and Research (IJSR)*, vol. 9, no. 1, pp. 381–386, Jan. 2020, doi: 10.21275/art20203995.

[17] T. Sinta Peringkat *et al.*, "Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu," 2020, [Online]. Available: www.bri-institute.ac.id

[18] S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes Orange Data Mining Implementation For Student Graduation Classification Using K-Nearest Neighbor, Decision Tree And Naive Bayes Models," 2021.

[19] D. Lianda and N. Surya Atmaja, "Prediksi Data Buku Favorit Menggunakan Metode Naïve Bayes (Studi Kasus: Universitas Dehasen Bengkulu)," 2021. [Online]. Available: www.ejournal.unib.ac.id/index.php/pseudocode

[20] Nurmalitasari, Z. Awang Long, and M. Faizuddin Mohd Noor, "Factors Influencing Dropout Students in Higher Education," *Educ Res Int*, vol. 2023, 2023, doi: 10.1155/2023/7704142.

[21] S. A. Alwarthan, N. Aslam, and I. U. Khan, "Predicting Student Academic Performance at Higher Education Using Data Mining: A Systematic Review," 2022, *Hindawi Limited*. doi: 10.1155/2022/8924028.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i1.2300, Copyright ©2025
Submitted : October 14, 2024, Revised : November 12, 2024, Accepted : November 14, 2024, Published : January 31, 2025