Major Recommendation System for New Students at SMK Muhammadiyah 1 Lamongan with Naive Bayes Algorithm

Wildan Irsyad Muzaqi[1]*, M. Ghofar Rohman[2], Danang Bagus Reknadi[3]
Faculty of Science and Technology [1], [2], [3]
Lamongan Islamic University
Lamongan, Indonesia
alzaqiahmad@gmail.com[1], m.ghofarrohman@unisla.ac.id[2], danz.0907@unisla.ac.id[3]

***Abstract*— Students' majors in Vocational High Schools (SMK) are very important in determining the direction of their education and career, but the process carried out so far is often subjective and does not consider academic grades and interests objectively. To overcome this, this study develops a website-based major recommendation system at SMK Muhammadiyah 1 Lamongan using the Naive Bayes algorithm that is able to provide accurate major recommendations based on student data. This system is designed using a structured Waterfall Model software development method, starting from needs analysis, design, implementation, to testing. The Naive Bayes algorithm was chosen because of its simplicity and ability to work with relatively small datasets, such as new student data at the school. Of the total 675 student data collected, 60% or 405 data were used as training data to train the Naive Bayes algorithm, while the remaining 40% or 270 data were used as test data to measure the accuracy level of the recommendation system. The test results show that the system achieves an average accuracy of 90.91%, with precision above 0.73 for each major, recall above 0.80 except for the Office Management major which reaches 0.75, and an average F1 score of 81.72%. These findings indicate that the website-based major recommendation system with the Naive Bayes algorithm is effective and can help students determine majors that suit their potential and interests objectively and accurately, thus supporting a more precise and targeted major selection process.**

Keywords*— Major Recommendation System; Naive Bayes; Vocational High School Majors; Waterfall Model; Web-Based;

I. Introduction

Student majors in Vocational High Schools (SMK) play a very important role in determining the direction of students' education and careers in the future[1]. The majoring process is usually carried out based on academic grades, interests, and recommendations from guidance and counseling teachers. However, often, this process is carried out subjectively or only relies on short interviews, so that the results are not entirely accurate and in accordance with the student's potential. This can cause students to feel less suited to the chosen major, which has the potential to reduce their learning motivation and academic performance[2].

To overcome this problem, a system is needed that can provide objective major recommendations based on data[3]. Website-Based Technology is very relevant to be applied in this context because of its easy access for students, teachers, and parents. Website-Based Systems also allow fast and accurate data processing, so that it can help the majoring process more effectively[4], [5].

One of the algorithms that can be used in developing a recommendation system is Naive Bayes. This algorithm works based on probability to determine the match between student profiles and available majors[6]. The advantages of Naive Bayes are its simplicity in implementation and its ability to work well with relatively small datasets, such as new student data at SMK Muhammadiyah 1 Lamongan[7]. By using historical data of academic grades, interests, and preferences of students, this algorithm can provide objective and data-based major recommendations.

This study aims to develop a Website-Based Major Recommendation System at SMK Muhammadiyah 1 Lamongan using the Naive Bayes algorithm which is extended by considering the dimensions of student personalization and integration of educational environment data. The main advantage of this system compared to previous studies lies in the holistic approach that simultaneously combines academic, psychological, student interest variables, and external environmental data as determinant factors, which were previously rarely integrated in major recommendation systems. In addition, this system implements an adaptive learning model that dynamically adjusts recommendations based on real-time feedback from users (students and guidance counselors), an innovation that has not been widely applied in similar studies. With a user-friendly interface feature and transparent monitoring process, this system provides added value in the form of easier, more accurate and reliable decision-making for all stakeholders.

Previous research related to the student major recommendation system using the Naive Bayes algorithm has been conducted in several educational institutions[8]. One relevant study developed a final project recommendation system for students, using Naive Bayes to suggest appropriate topics based on academic data and student interests[9]. Other

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2390, Copyright ©2025
Submitted : Mey 20, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

studies also implemented similar algorithms to help determine majors for students based on their grades and preferences, aiming to improve the match between student abilities and the chosen major. The results of these studies demonstrate the effectiveness of the Naive Bayes algorithm in producing precise and accurate recommendations.

Research related to recommendation systems based on the Naive Bayes algorithm has been widely conducted with various approaches and applications. One interesting approach was developed by researchers in the journal, which proposed a hybrid method by combining product attributes using a genetic algorithm and a Naive Bayes classifier. This method is designed to overcome the problems of sparsity and cold-start in recommendation systems, using a combination of explicit and implicit attributes to provide more personalized and relevant recommendations. This hybrid approach shows better performance than conventional methods due to the ability of genetic algorithms to optimize features used by Naive Bayes [10]. Meanwhile, another article examines the collaborative filtering approach that utilizes the Naive Bayes classifier as the main classification mechanism to improve the accuracy of recommendation predictions. The focus on managing user data and their preferences in the system allows this model to better adjust recommendations. This study also discusses how combining collaborative filtering with the Naive Bayes statistical method can help overcome the problem of data sparseness and improve user satisfaction in the context of recommendation systems[11]. In addition, research in other journals focuses on the application of trust in recommendation systems using the Naive Bayes classifier to improve the reliability and accuracy of predictions, especially in social networks. The methodology used involves three stages, namely user profile creation, dataset preparation, and classification with Naive Bayes. This study highlights the importance of the trust aspect in improving the quality of recommendations that are not only accurate but also trustworthy by users, thus improving the user experience in using the recommendation system [12].

Various previous studies have applied the Naive Bayes algorithm in the development of recommendation systems in the field of education with promising results, namely by developing a final project title selection recommendation system using the Naive Bayes method and development model Waterfall [13]. This system is able to provide recommendations for final project topics based on course grades and student interests objectively. The main advantage of this study is its ability to adjust recommendations to student academic data, but the limitation found is the lack of broader historical data, which has the potential to limit the accuracy of the system's predictions.

Subsequent studies also applied the Naive Bayes Classifier algorithm to recommend thesis supervisors based on competence, functional positions, and lecturer homebases. This system helps students choose supervisors more objectively and in a structured manner. However, this study also revealed shortcomings in terms of the lack of consideration of student personalization factors, which can affect the compatibility between students and supervisors[14].

Another example of developing a web-based High School selection recommendation system using Naive Bayes and data mining. This system considers criteria such as distance, cost, accreditation, and school graduation rates. The results show high accuracy in the classification of recommended schools. However, this study has not included student psychological factors, which can be an important aspect in making major decisions[15].

Another study combines text mining with the Naive Bayes algorithm to recommend thesis supervisors based on the suitability of the thesis topic with the lecturer's knowledge. This system speeds up the process of determining the title and supervisor, thus providing efficiency in the academic process. However, this study does not explicitly mention the shortcomings found, so there is still room for further evaluation[16].

Another literature study also developed a study program selection recommendation system using the Naive Bayes algorithm with an accuracy of 73.4%. This system is able to work with relatively little training data, which is an advantage in the context of limited data. However, the accuracy of the system can still be improved by adding other relevant variables to enrich the model[17].

Overall, these studies show that the Naive Bayes algorithm is an effective and efficient method in developing recommendation systems in education. The main advantages of this algorithm lie in the simplicity of implementation and its ability to work with relatively small datasets. However, several studies also highlight limitations such as the need for more complete historical data and the need to consider personalization factors and additional variables to improve the accuracy and relevance of recommendations. Therefore, the development of a Naive Bayes-based recommendation system still has great potential to be further developed by considering these aspects.

II. METHODOLOGY

The research method is a systematic step used in developing this web-based major recommendation system. This study focuses on the application of the Naive Bayes algorithm in the major recommendation process for new students at SMK Muhammadiyah 1 Lamongan. In addition, this study also applies a software development method using the Waterfall Model, which allows the system to be developed in stages, starting from needs analysis, system design, implementation, to testing to measure the accuracy of the resulting recommendation system. This model was chosen because it has structured and systematic stages, so that each stage must be completed before proceeding to the next stage. With this approach, system development can be carried out in stages, starting from needs analysis to testing and maintenance.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2390, Copyright ©2025
Submitted : Mey 20, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

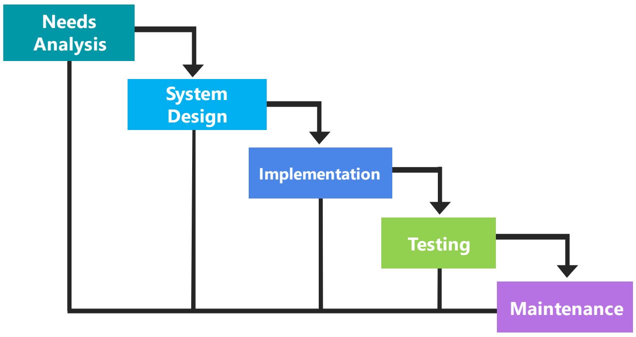

Fig. 1. Waterfall model in this research

The Waterfall model consists of several main stages, namely needs analysis, system design, implementation, testing, and maintenance.

A. Needs Analysis

The needs analysis aims to determine the important things that must be met so that the system can run well and in accordance with the needs of its users. This analysis is divided into two types, namely functional and non-functional needs. Functional needs include the system's ability to receive student data such as academic grades, interests, and preferences, and apply the Naive Bayes algorithm to analyze the data and provide recommendations for majors based on probability. Before the analysis is carried out, the received data is first processed to handle missing values using the average imputation method and data balancing is carried out if an imbalance in the major class is found using the oversampling technique to ensure that the model is not biased towards certain majors. In addition, the system must be able to store student data and recommendation results in a well-organized database, as well as provide an interface for school administrators to manage data and monitor recommendation results. While non-functional needs include the speed of the system in processing data and producing recommendations, compatibility with various devices that can be accessed via a web browser, a simple, easy-to-understand, and user-friendly interface design for both students and administrators, and system reliability with a very low error rate, especially in the analysis process and recommendation results.

B. System Design

The website-based major recommendation system for new students at SMK Muhammadiyah 1 Lamongan is designed to provide an effective solution in helping students determine majors that suit their abilities and interests.

This website-based major recommendation system begins with collecting student data, such as academic grades and interests. The features used in the model are selected based on their relevance to major decision-making, such as academic grades for the last semester's report card, major subject grades, and interest data filled in by students through questionnaires. The selection of these features is supported by a literature study that shows a strong relationship between these features and major needs. Once the data is ready, the Naive Bayes algorithm will be applied to perform probability and classification analysis to determine the major that best suits the characteristics of the students. Based on the results of the algorithm, the system will generate recommendations in the form of a list of majors that are suitable for each student.

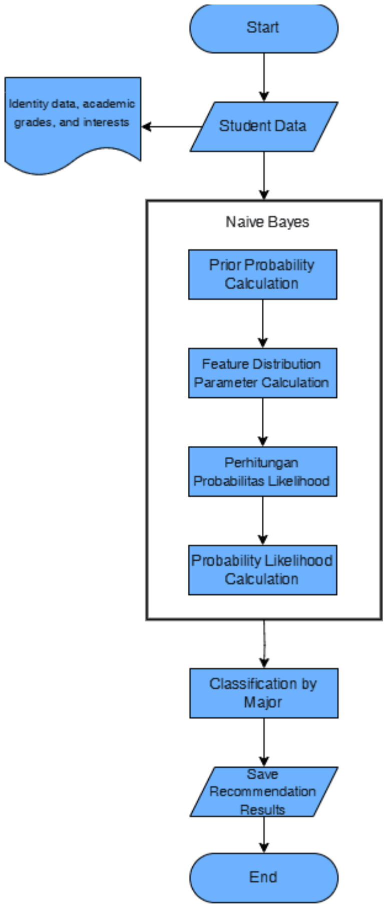

Fig. 2. Flowchart of Website-Based Major Recommendation System for New Students

As in the figure 2. This stage reflects the logical flow in processing student data to produce recommendations based on probabilistic analysis. The system flowchart provides an organized structure to support more accurate and efficient decision making.

C. Implementation

The implementation stage is the step where the web-based major recommendation system begins to be built based on the design that has been made previously. At this stage, coding and integration of the Naive Bayes algorithm into the system are carried out to process student data and produce major recommendations. The implementation of this system consists of several main components, namely:

- The user interface development is done by utilizing web technologies such as HTML, CSS, and JavaScript to build an interactive and easy-to-use display. To ensure a responsive display that can adjust to various screen sizes, the Bootstrap framework is used. The main page of the system includes a form for student data input, a display of department recommendation results, and a special page for administrators that

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2390, Copyright ©2025
Submitted : Mey 20, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

functions to manage data efficiently.

- The development of application logic or back-end is done using the Python programming language and the Flask framework. This system is tasked with processing student data that has been entered, then classifying it using the Naive Bayes algorithm. In this section, the probability calculation function is implemented based on student features such as academic grades and interests, so that the system can provide accurate recommendations for appropriate majors.
- The integration of the Naive Bayes algorithm into the system is done by calculating the probability for each major based on the data entered by the student. This process uses the Bayes Theorem formula to determine the major with the highest probability that best suits the student's profile. After the calculation is complete, the system saves the classification results and displays major recommendations to the user directly and easily understood.

After the system is implemented, a testing phase is carried out to ensure that the system functions properly and is able to provide accurate major recommendations. Testing is carried out using the Black Box Testing method and accuracy evaluation with Confusion Matrix.

D. Testing

The testing phase aims to ensure that the web-based major recommendation system functions properly and is able to provide results that are in accordance with the research objectives. Testing is carried out to detect errors in the system, evaluate the accuracy of the Naive Bayes algorithm, and ensure that the system can be used easily by users.

In this study, testing was conducted with two main approaches. First, functional testing or Black Box Testing, which aims to ensure that every feature in the system runs according to the design that has been made. This testing is done by providing certain inputs to the system and then evaluating whether the output produced is as expected.

In addition, cross-validation was performed using the *k*-fold method (*k*=5) to measure the stability and consistency of the Naive Bayes model performance on different data randomly. This helps reduce evaluation bias that can arise in the division of single training and testing data.

To measure how accurate the system is in providing major recommendations, the Confusion Matrix method is used. This method compares the recommendations given by the system with the major data that has been determined by the school. From this comparison, calculations of accuracy, precision, and recall are carried out using certain formulas to assess the overall performance of the system[18].

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{1}$$

$$Presisi = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 = 2 \times \frac{presisi \times recall}{presisi + recall} \tag{4}$$

The testing procedure begins by preparing test data derived from a new student dataset that has been classified by the school as validation data.

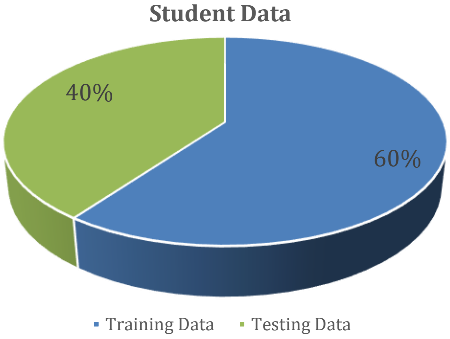

Fig. 3. Student Data Graph obtained from observation results

As in Fig. 3 of the total 675 student data collected, 60% or 405 data were used as training data to train the Naive Bayes algorithm, while the remaining 40%, or 270 data, were used as test data to measure the accuracy level of the recommendation system. This approach ensures that the system is tested with representative and valid data.

E. Maintenance

The maintenance stage is the final stage in the Waterfall model which aims to ensure that the web-based major recommendation system continues to function optimally after implementation.

The maintenance of this system is carried out in two main forms. First, corrective maintenance which aims to fix errors or bugs that appear after the system is used by students and school administrators, such as problems with data input, errors in the calculation of the Naive Bayes algorithm, or inconsistencies in the interface display. Second, adaptive maintenance which focuses on adjusting the system to changes in school needs, for example when there is the addition of a new major or a change in the major policy. In addition, adaptive maintenance also includes updating the Naive Bayes algorithm if necessary, such as adding new factors in the recommendation calculation process so that the system remains relevant and effective.

III. Result and Analysis

An in-depth description of the research results and system design that has been carried out, especially regarding the implementation of the Naive Bayes algorithm in the department recommendation system.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2390, Copyright ©2025
Submitted : Mey 20, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

A. Manual Calculation of Naive Bayes Algorithm

The Naive Bayes algorithm is a probabilistic classification method based on Bayes' Theorem with the assumption of independence between features. The basic principle of this algorithm is to calculate the posterior probability of a class based on the features owned by the data, then choose the class with the highest probability as the classification result. The basic Naive Bayes formula for calculating the probability of a class C_k against feature data $X = (x_1, x_2, \dots, x_n)$ is as follows:

$$aP(C_k \mid X) = \frac{P(C_k) \times \prod_{i=1}^{n} P(x_i \mid C_k)}{P(X)} \tag{5}$$

Since $P(X)$ constant for all classes, the calculation is usually focused on the numerator value only, so that it becomes:

$$(C_k \mid X) \propto P(C_k) \times \prod_{i=1}^{n} P(x_i \mid C_k) \tag{6}$$

Where:

- $P(C_k)$ is the prior probability of class C_k, which is the frequency of occurrence of that class in the training data.
- $P(x_i \mid C_k)$ is the likelihood probability of feature x_i given class C_k.
- C_k is the notation used to denote the k-th class in a classification problem. For example, if several classes or categories are to be predicted, such as:
  - C_1 = class “Teknik Kendaraan Ringan Otomotif”
  - C_2 = class “Farmasi Klinis Komunitas”
  - C_3 = class “Multimedia”
  - C_4 = class “Manajemen Perkantoran”
- X is a notation consisting of several attributes that describe the student profile, namely:
  - X_1 = value Matematika (Mtk)
  - X_2 = value Ilmu Pendidikan Alam (Ipa)
  - X_3 = value Ilmu Pendidikan Sosial (Ips)

    X_4 = Minat : category minat siswa, for example “Teknik”, “Kesehatan”, or “Administrasi”.

To provide a more concrete picture of how the Naive Bayes algorithm works, here is an example of manual calculation using the training data that has been collected. Suppose there is student data with the following feature values:

TABLE I. STUDENT SAMPLING DATA

Student	Mtk	IPA	IPS	Minat
A	80	85	70	Teknik
B	75	80	65	Teknik
C	60	70	80	Kesehatan
D	65	75	85	Kesehatan

Step 1: Calculating Prior Probability

Calculate the frequency of occurrence of each major:

- Teknik : 2 students
- Kesehatan : 2 students
- Administrasi : 1 student

Total data = 5 students, so the prior probability:

$$P(Teknik) = \frac{2}{5} = 0.4$$

$$P(Kesehatan) = \frac{2}{5} = 0.4$$

$$P(Administrasi) = \frac{2}{5} = 0.4$$

Step 2: Calculating the Mean and Variance of Features per Major

Calculate the mean (μ) and variance (σ^2) for each feature in each major.

TABLE II. CALCULATE THE MEAN AND VARIANCE

Minat	Academic	Mean (μ)	Varians (σ^2)
Teknik	Mtk	$\mu = \frac{80-75}{2} = 77.5$	$\sigma^2 = \frac{(80-77.5)^2 + (75-77.5)}{2} = 6.25$
Teknik	IPA	$\mu = \frac{85-80}{2} = 82.5$	$\sigma^2 = \frac{(85-82.5)^2 + (80-82.5)}{2} = 6.25$
Teknik	IPS	$\mu = \frac{70-65}{2} = 77.5$	$\sigma^2 = \frac{(70-67.5)^2 + (65-67.5)}{2} = 6.25$
Kesehatan	Mtk	$\mu = \frac{60-65}{2} = 62.5$	$\sigma^2 = \frac{(60-62.5)^2 + (65-62.5)}{2} = 6.25$
Kesehatan	IPA	$\mu = \frac{70-75}{2} = 72.5$	$\sigma^2 = \frac{(70-72.5)^2 + (75-72.5)}{2} = 6.25$
Kesehatan	IPS	$\mu = \frac{80-85}{2} = 82.5$	$\sigma^2 = \frac{(80-82.5)^2 + (85-82.5)}{2} = 6.25$
Administrasi	Mtk	70	0 (only one data)
Administrasi	IPA	60	0
Administrasi	IPS	90	0

Step 3: Calculating Likelihood Probability for New Data

Suppose there is a new student with the following scores:

- Mtk = 72
- Ipa = 78
- Ips = 68

Calculate the likelihood probability for each feature and class using the Gaussian distribution formula:

$$P(x_i \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_{C_k}^2}} exp\left(-\frac{\left(x_i - \mu_{C_k}\right)^2}{2\sigma_{C_k}^2}\right) \tag{7}$$

Example calculation for the academic feature of Mathematics (Mtk) in the Engineering class:

$$PP(72 \mid Teknik) = \frac{1}{\sqrt{2\pi \times 5^2}} exp\left(-\frac{(72-77.5)^2}{2 \times 6.25}\right) \tag{8}$$

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2390, Copyright ©2025
Submitted : Mey 20, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

Calculate this value using a calculator or statistical software.

Step 4: Calculating Posterior Probability
Calculate the prior probability with all feature likelihood probabilities for each class:

$$PP(C_k \mid X) \propto P(C_k) \times P(Mtk \mid C_k) \times P(Ipa \mid C_k) \times P(Ips \mid C_k) \quad (9)$$

Perform this calculation for all majors, then select the major with the highest posterior value as the prediction of the new student's major.

TABLE III. POSTERIOR PROBABILITY CALCULATION EXAMPLE

Jurusan	$P(C_k)$	$P(Mtk \mid ($	$P(Ipa$	$P(Ips \mid$	$P(C_k \mid X)$ (Posterior
Teknik	0.4	0.05	0.06	0.07	0.4 × 0.05 × 0.06 × 0.07 = 0.000084
Kesehatan	0.4	0.04	0.05	0.06	0.4 × 0.04× 0.05 × 0.06 = 0.000048
Administrasi	0.2	0.03	0.02	0.01	0.2 × 0.03 ×0.02 × 0.01 = 0.0000012

From Table 3, the Engineering class has the highest posterior value, so the student is predicted to choose the Engineering major.

B. Web-based recommendation system

In developing a department recommendation system using the Naïve Bayes algorithm, various programming functions and processes utilize certain Python libraries to facilitate implementation and increase efficiency.

The major recommendation system developed using the Naive Bayes algorithm consists of several main stages. In its implementation, the back-end part is programmed using Python by utilizing the Flask library to build a server and connect data processing logic. Meanwhile, the front-end part is designed using HTML and Bootstrap so that the interface display is responsive and easy to use.

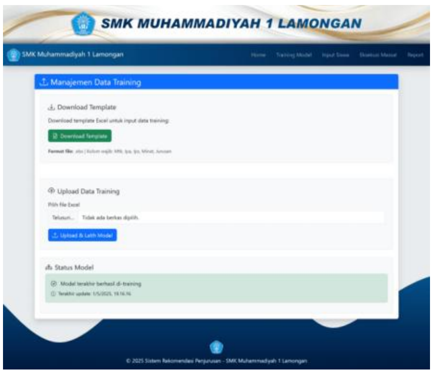

Fig. 4. Page training_model.html

Model Training Process and System Implementation that functions as a place to manage training data for the major recommendation system On this page, users can download the training data template in Excel file format (.xlsx) as shown in fig. 4 .as a guide to filling in the correct data. In addition, there is a training data upload feature that allows users to upload Excel files containing the latest data to update the model.

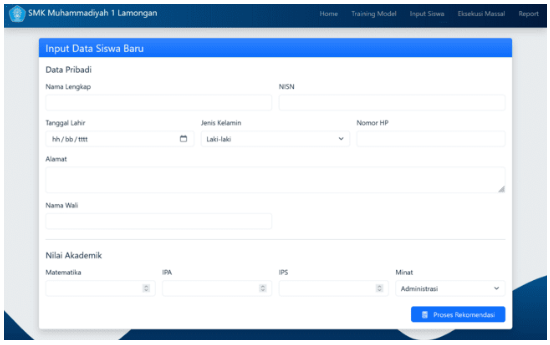

Fig. 5. Page input_siswa.html

to enter student data into the major recommendation system. Users can input students' academic scores, namely Mathematics, Science, Social Studies, and select student interests as supporting features, as in figure 5

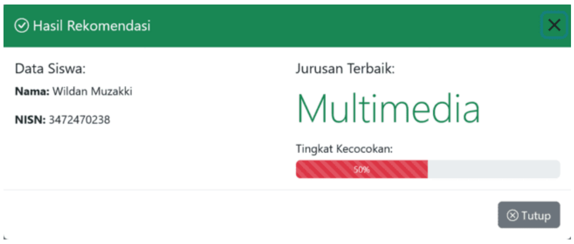

Fig. 6. Recommendation Results Notification View

After the data is entered, the system automatically processes and displays the most appropriate major prediction results using the Naive Bayes algorithm as shown in figure 6 . In addition, all student input data along with prediction results are automatically stored directly into the database, making it easier to manage and track data in a structured and secure manner. This page is designed to make the input and prediction process fast, easy, and efficient for users.

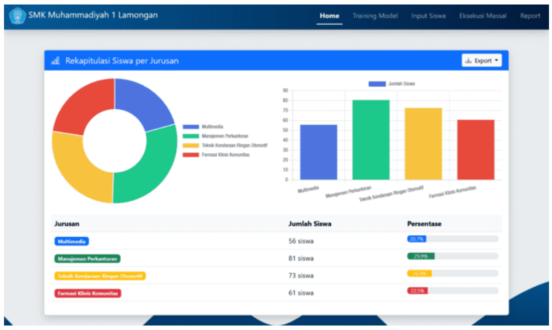

Fig. 7. Index.html page

Figure 7 displays a graph showing the distribution of data from four majors based on the results of the majoring process of 271 students. This graph helps users to see comparisons and trends in data visually more clearly. In addition, this page is also equipped with an export feature that allows users to save the graph in PNG image format or export the data recapitulation to an Excel file. This feature is very useful for practical and fast

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2390, Copyright ©2025
Submitted : Mey 20, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

documentation.

C. Model Evaluation

In evaluating the performance of a classification model, four important terms used to measure the prediction results are True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). These four values are obtained from the confusion matrix and are the basis for calculating evaluation metrics such as precision, recall, and F1-score.

For the case of multi-class classification as in this study with four majors, the calculation of TP, FP, TN, and FN is done in a one-vs-all manner for each major. This means that each major is considered a positive class, while the other majors are combined as negative classes. As in Table 4 are the TP, FP, TN, and FN values obtained for each major based on the model prediction results.

TABLE IV. CONFUSION MATRIX SEARCH RESULTS

No	Jurusan	TP	FP	TN	FN
1.	Teknik Kendaraan Ringan Otomotif	52	18	193	8
2.	Farmasi Klinis Komunitas	47	15	203	6
3.	Multimedia	48	11	201	11
4.	Manajemen Perkantora	75	5	167	24

At this stage, the confusion matrix is used as the main tool to evaluate the performance of the model. The confusion matrix allows this study to see in detail the number of correct and incorrect predictions for each major, so that it can measure how accurate the system is in classifying students into the right major.

From table 4, the next step is to calculate the evaluation metrics commonly used to assess the performance of the classification model, namely Accuracy, Precision, Recall, and F1-Score. For example, for the “Teknik Kendaraan Ringan Otomotif” major with a score of TP=52, FP=18, TN=193, and FN=8, the calculation is:

- Accuracy $= \frac{52+193}{52+18+193+8} = \frac{245}{271} = 0.904 = 90.4\%$
- Presisi $= \frac{52}{52+18} = \frac{52}{70} = 0.743$
- Recall $= \frac{52}{52+8} = \frac{52}{60} = 0.867$
- F1-Score $= 2 \times \frac{0.743 \times 0.867}{0.743 + 0.867} = 0.801 = 80.00\%$

Similar calculations were performed for other majors using the TP, FP, TN, and FN values that had been obtained. And the results can be seen in Table 5.

TABLE V. PRECISION & RECALL CALCULATION RESULTS FOR ALL MAJORS

No	Jurusan	Precision	Recall
1.	Teknik Kendaraan Ringan Otomotif (TKRO)	0.743	0.867
2.	Farmasi Klinis Komunitas (FKK)	0.758	0.887
3.	Multimedia (MM)	0.814	0.814
4.	Manajemen Perkantoran (MP)	0.938	0.758

To provide a more comprehensive picture of the model performance, the Accuracy and F1-score values for each department are visualized in the form of bar graphs as in Figure 8 and Figure 9.

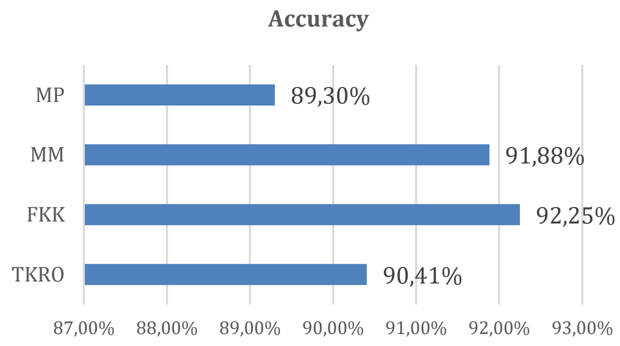

Fig. 8. Graphic Accuracy All Majors

Testing the Naive Bayes model on the department recommendation system produced adequate evaluation metrics, as seen in Figure 8. The model accuracy per department is as follows: MP (89.30%), MM (91.88%), FKK (92.25%), and TKRO (90.41%).

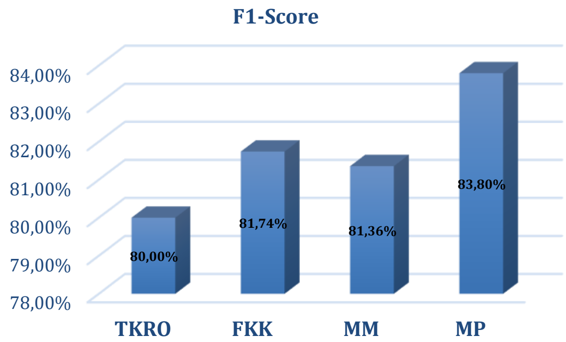

Fig. 9. Graphic F1-Score All Majors

Performance data analysis shows that the FKK class has the highest accuracy but not the highest F1-score, while MP shows the highest F1-score even though its accuracy is relatively lower than FKK. The recall of the TKRO class is relatively lower, which is an indication that the model has difficulty recognizing all positive TKRO data correctly.

One factor causing this difference is the similarity of features between certain classes, such as between TKRO and MM, which can cause the model to misclassify. In addition, the more limited data distribution in classes with low recall also contributes to less than optimal performance. This leads to the

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2390, Copyright ©2025
Submitted : Mey 20, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

need to increase the proportion and quality of training data so that the model can better learn the feature patterns of each class.

These factors are important to consider in the development of a recommendation system so that the prediction rate can be improved and the system can provide more accurate and relevant recommendations. To ensure the reliability of the evaluation results, the model was tested using the cross-validation method with 5-fold cross-validation. This method ensures that each subset of data is used alternately as training and testing data, thereby reducing bias due to certain data divisions. The average results of the evaluation metrics from cross-validation confirm that the Naive Bayes model provides consistent and reliable performance in the context of this department's recommendation system.

IV. Conclusion

After going through a series of research and development processes of the Naive Bayes algorithm-based major recommendation system, this section will present a conclusion that summarizes the important results obtained. The following are the main conclusions obtained from the results of this study.

- This study has succeeded in developing a website-based system that is able to provide major recommendations for new students at SMK Muhammadiyah 1 Lamongan. This system is designed with a user-friendly interface to facilitate students and schools in the process of selecting majors online and in a structured manner.
- The Naive Bayes algorithm is used as the main classification method in this system to provide major recommendations based on student academic grades, interests, and talents. This algorithm calculates the posterior probability of each major by considering these features, so that it can predict the major that best suits the student's characteristics effectively.
- System evaluation using testing data shows that the Naive Bayes algorithm is able to provide major recommendations with a fairly high level of accuracy. The test results show that the accuracy reaches a satisfactory value, where the average accuracy reaches 90.91%, while the precision in each major is above 0.73, the recall in each major is above 0.80 (except for the Office Management major which is only 0.75), and the average F1-score of all majors reaches 81.72% which supports the reliability of the system in classifying students into the right major.

V. Suggestions for Development

To improve the quality and effectiveness of the major recommendation system, several important suggestions can be applied. First, adding new features such as psychological test results, non-academic achievements, and environmental factors to make recommendations more complete and accurate. Next, trying other algorithms such as Random Forest or Support Vector Machine to improve prediction performance and overcome the limitations of Naive Bayes. The use of cross-validation techniques is also recommended to optimize the model and prevent prediction errors. In addition, handling incomplete data needs to be improved so that the results remain accurate. Developing a more user-friendly interface and integrating the system with the school database in real-time is also important so that users can easily access the latest information. With these steps, the recommendation system can become a more reliable and useful tool in helping students choose majors according to their potential and interests.

References

[1] A. Santika, E. R. Simanjuntak, R. Amalia, and S. R. Kurniasari, “Peran pendidikan sekolah menengah kejuruan dalam memposisikan lulusan siswanya mencari pekerjaan,” *Paedagoria J. Kajian, Penelit. dan Pengemb. Kependidikan*, vol. 14, no. 1, pp. 84–94, 2023.

[2] A. Mulyana *et al.*, *Metode penelitian kualitatif*. Penerbit Widina, 2024.

[3] A. Rizka, R. E. Putri, Y. Yusman, and M. Fajar, “Sistem Rekomendasi Jurusan Kuliah dalam Pengambilan Keputusan Menggunakan Metode MOORA,” *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 4, no. 2, pp. 364–373, 2023.

[4] D. B. Reknadi, M. Munif, and M. Mustain, “Optimasi Layanan KKN: Implementasi e-KKN Berbasis Web Pada Universitas Islam Lamongan,” *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 5, no. 1, pp. 152–161, 2024.

[5] A. Rohman, A. S. Hidaytullah, and Mg. Rohman, “Implementasi Metode Waterfall pada Rancang Bangun Sistem Pengarsipan Surat Berbasis Website,” *Gener. J.*, vol. 6, no. 2, pp. 134–143, 2022.

[6] A. F. Sitanggang, E. Rasyawir, and L. Aryani, “Klasifikasi Akreditasi Sekolah Di Kota Jambi Menggunakan Metode Na{\"\i}ve Bayes Di Bpmp Jambi,” *J. Inform. Dan Rekayasa Komput.*, vol. 4, no. 1, pp. 865–873, 2024.

[7] R. R. Sani, Y. A. Pratiwi, S. Winarno, E. D. Udayanti, and F. Alzami, “Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Berita Hoax pada Berita Online Indonesia,” *J. Masy. Inform.*, vol. 13, no. 2, pp. 85–98, 2022.

[8] A. Natuzzuhriyyah and N. Nafisah, “Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara Daring Menggunakan Algoritma Naive Bayes,” *Techno Xplore J. Ilmu Komput. Dan Teknol. Inf.*, vol. 6, no. 2, pp. 61–67, 2021.

[9] H. A. Pratama, “Implementasi algoritma c4. 5 dan naive bayes classification Untuk mengklasifikasi mahasiswa berpotensi drop out,” Fakultas Sains dan Teknologi UIN Syarif HIdayatullah Jakarta.

[10] M. Salehi and I. N. Kamalabadi, “A hybrid recommendation approach based on attributes of products using genetic algorithm and naive Bayes classifier,” *Int. J. Bus. Inf. Syst.*, vol. 13, no. 4, p. 381, 2022, doi: 10.1504/IJBIS.2013.055297.

[11] P. Valdiviezo-Diaz, F. Ortega, E. Cobos, and R. Lara-Cabrera, “A Collaborative Filtering Approach Based on Naïve Bayes Classifier,” *IEEE Access*, vol. 7, pp. 108581–108592, 2022, doi: 10.1109/ACCESS.2019.2933048.

[12] K. Rrmoku, B. Selimi, and L. Ahmedi, “Application of Trust in Recommender Systems—Utilizing Naive Bayes Classifier,” *Computation*, vol. 10, no. 1, p. 6, Jan. 2023, doi: 10.3390/computation10010006.

[13] Y. Yulindawati, S. Lailiyah, A. Yusnita, and A. Hafifah, “Rekomendasi Pemilihan Judul Tugas Akhir Menggunakan Metode Naive Bayes,” *J. Inf. Syst. Manag.*, vol. 5, no. 2, pp. 171–175, 2024.

[14] M. Asfi and N. Fitrianingsih, “Implementasi Algoritma Naive Bayes Classifier sebagai Sistem Rekomendasi Pembimbing Skripsi,” *InfoTekJar J. Nas. Inform. dan Teknol. Jar*, vol. 5, no. 1, p. 44, 2020.

[15] T. D. Salma and Y. S. Nugroho, “Sistem Rekomendasi Pemilihan Sekolah Menengah Tingkat Atas Menggunakan Metode Naive Bayes,” *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 2, no. 2, pp. 85–94, 2016.

[16] M. Ikhsan and others, “Penerapan Text Mining Pada Sistem Rekomendasi Pembimbing Skripsi Mahasiswa Menggunakan Algoritma Naive Bayes Classifier di Program Studi Ilmu Komputer UIN Sumatera Utara Medan,” *Indones. J. Comput. Sci.*, vol. 12, no. 6, 2023.

[17] H. N. F. Fikrillah and D. Kurniadi, “Rekomendasi Pemilihan Program Studi Menggunakan Algoritma Naive Bayes,” *J. Algoritm.*, vol. 20, no. 1, pp. 42–49, 2023.

[18] I. Markoulidakis, G. Kopsiaftis, I. Rallis, and I. Georgoulas, “Multi-class confusion matrix reduction method and its application on net promoter score classification problem,” in *Proceedings of the 14th PErvasive technologies related to assistive environments conference*, 2021, pp. 412–419.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2390, Copyright ©2025
Submitted : Mey 20, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025