

Research Article

Data Mining Classification in The New Student Admission Process Using The K-Nearest Neighbors Method (Case Study: Yapmi Boarding School)

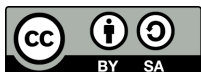
Veri Arinal ¹, Tri Wahyudi ², and Mesra Betty Yel ³, Nurul Khoiriyah ⁴

- 1 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (STIKOM CKI) Jakarta; Email: veriarinal@stikomcki.ac.id
 - 2 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (STIKOM CKI) Jakarta; Email: triwahyudi@stikomcki.ac.id
 - 3 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (STIKOM CKI) Jakarta; Email: mesrabettyyel@stikomcki.ac.id
 - 4 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (STIKOM CKI) Jakarta
- * Corresponding Author: veriarinal@stikomcki.ac.id

Abstract: The new student admission process is an important activity in educational institutions to ensure that prospective students meet the established admission criteria. However, the selection process is often conducted manually, making it less efficient and prone to subjective assessments. This study aims to implement a data mining classification approach using the K-Nearest Neighbors (K-NN) method to support decision-making in the new student admission process at Yapmi Boarding School. The research utilizes historical admission data consisting of academic scores, interview results, and other admission criteria as classification attributes. The K-NN algorithm was applied to classify prospective students into accepted and rejected categories based on the similarity of their characteristics to previously evaluated applicants. The research methodology includes data collection, preprocessing, classification modeling, and performance evaluation using accuracy metrics. The results demonstrate that the K-NN method is capable of classifying prospective students effectively and can assist admission committees in making more objective and accurate decisions. The implementation of this model contributes to improving the efficiency, consistency, and reliability of the student admission process at Yapmi Boarding School. Therefore, the K-NN algorithm can be considered a viable alternative for supporting educational admission decision systems.

Keywords: Classification; Data Mining; K-Nearest Neighbors; Student Admission; Student Selection.

Received: 12 Augst 2024
Revised: 17 September 2024
Accepted: 25 October 2024
Published: 30 October 2024
Curr. Ver.: 30 October 2024



Copyright: © 2025 by the authors.
Submitted for possible open
access publication under the
terms and conditions of the
Creative Commons Attribution
(CC BY SA) license
(<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Introduction

Student admission is a crucial process that influences the quality of education and the overall performance of educational institutions. The selection of prospective students requires an objective, efficient, and accurate assessment mechanism to ensure that admitted students meet the academic and non-academic criteria established by the institution. However, many educational institutions still rely on conventional and manual selection procedures, which often require significant time and resources and may lead to inconsistencies in decision-making.

YAPMI Boarding School faces similar challenges in the student admission process. The existing admission procedure relies heavily on manual assessment, making it difficult to evaluate large volumes of applicant data while simultaneously considering multiple selection criteria. Furthermore, manual evaluation may introduce subjectivity and increase the possibility of human error, potentially affecting the accuracy of admission decisions. These challenges indicate the need for a more systematic and data-driven approach to support the student selection process.

Recent advances in data mining have provided various techniques for extracting meaningful knowledge from large datasets and supporting decision-making processes in different domains, including education. Data mining classification techniques have been widely applied to identify patterns, predict outcomes, and improve decision accuracy. Among the available classification methods, the K-Nearest Neighbors (KNN) algorithm has gained considerable attention due to its simplicity, effectiveness, and ability to classify data based on similarity measures. The KNN algorithm determines the class of a new instance by analyzing the characteristics of the nearest neighboring data points within a predefined distance metric.

Several previous studies have demonstrated the applicability of data mining techniques in educational decision support systems. Research on student admission, scholarship recipient selection, student major recommendation, and educational promotion strategies has shown that classification algorithms can improve decision-making performance and reduce manual processing efforts. In particular, the KNN algorithm has been reported to achieve satisfactory classification performance in various educational datasets, making it a promising approach for student admission classification.

Despite the growing adoption of data mining methods in educational settings, the implementation of KNN for classifying prospective students in boarding school admission processes remains relatively limited. Therefore, this study focuses on the application of the K-Nearest Neighbors algorithm to support the admission process at YAPMI Boarding School. The proposed approach is expected to assist admission committees in evaluating prospective students more objectively and efficiently by utilizing historical admission data and similarity-based classification techniques.

The objectives of this study are to implement the K-Nearest Neighbors algorithm for student admission classification, identify the factors that influence admission decisions, evaluate the classification performance of the algorithm, and determine the most appropriate parameter configuration for the admission dataset used in this research. In addition, the study aims to provide a prototype decision-support model that can be integrated into the student admission process at YAPMI Boarding School.

This research is expected to contribute to the application of data mining techniques in educational management by providing an alternative approach for improving the effectiveness, efficiency, and objectivity of student admission processes. Furthermore, the findings may serve as a reference for other educational institutions seeking to implement data-driven admission systems based on classification methods.

2. Literature Review

Data Mining

Data mining is a process of extracting meaningful patterns, relationships, and knowledge from large datasets using statistical, computational, and machine learning techniques. The primary objective of data mining is to transform raw data into useful information that can support decision-making processes. In educational institutions, data mining has been widely utilized to analyze student performance, predict academic outcomes, support admission decisions, and improve institutional management.

Classification is one of the most commonly used techniques in data mining. Classification aims to assign data objects into predefined categories based on their attributes. In the context of student admission, classification techniques can assist decision-makers in

identifying prospective students who meet the admission criteria established by educational institutions. The implementation of classification methods can reduce subjectivity in the decision-making process and improve the consistency of admission outcomes.

Several studies have demonstrated the effectiveness of data mining techniques in educational environments. Data mining has been applied to student admission selection, scholarship recipient classification, academic performance prediction, and educational recommendation systems. These applications indicate that data-driven approaches can enhance the efficiency and accuracy of educational decision-making processes (Budiman, 2021; Saifudin, 2018).

Classification in Educational Data Mining

Classification is a supervised learning technique that utilizes historical data with known labels to build predictive models. The generated model can subsequently be used to classify new instances into specific categories. The success of classification depends on the quality of the training data, the selected features, and the algorithm employed.

Within educational data mining, classification methods have become increasingly important due to the growing volume of educational data. Educational institutions often possess extensive datasets related to student demographics, academic performance, admission records, and behavioral characteristics. Proper utilization of these datasets can support evidence-based decision-making and improve institutional effectiveness.

The application of classification techniques in student admission processes enables educational institutions to evaluate applicants based on multiple criteria simultaneously. Such an approach provides a more objective assessment compared to conventional manual selection methods, which may be susceptible to human bias and inconsistency.

K-Nearest Neighbors Algorithm

K-Nearest Neighbors (KNN) is a supervised machine learning algorithm that performs classification by identifying the nearest neighboring instances within a dataset. The algorithm classifies a new data object according to the majority class of its nearest neighbors. Due to its simplicity and effectiveness, KNN has become one of the most widely used classification algorithms in data mining applications.

The fundamental principle of KNN is based on similarity measurement. Data instances that exhibit similar characteristics are expected to belong to the same class. To determine similarity, KNN calculates the distance between a new data instance and all instances in the training dataset. The most commonly used distance metric is Euclidean Distance.

The performance of KNN is strongly influenced by the value of K, which represents the number of nearest neighbors considered during classification. A small value of K may increase sensitivity to noise, whereas a large value may reduce classification specificity. Therefore, selecting an appropriate K value is essential to achieve optimal classification performance (Shichao et al., 2018; Zhang, 2022).

Several studies have reported successful applications of KNN in educational domains. KNN has been utilized for scholarship recipient selection, student major recommendation, student admission classification, and educational decision support systems. These studies indicate that KNN can provide satisfactory classification performance while maintaining computational simplicity (Yanasari & Arifin, 2023; Kurniawan & Saputra, 2019).

Student Admission Selection

Student admission is a strategic process undertaken by educational institutions to identify and select applicants who satisfy predetermined academic and non-academic

requirements. The quality of admitted students significantly influences institutional performance, educational outcomes, and overall competitiveness.

Traditional admission procedures often rely on manual evaluation processes, which can be time-consuming and prone to inconsistencies. As the number of applicants increases, educational institutions face greater challenges in managing and evaluating admission data effectively. Consequently, the integration of data mining techniques into admission systems has emerged as a promising solution for improving decision quality and operational efficiency.

The use of classification algorithms in admission systems enables institutions to evaluate applicants systematically based on historical data patterns. Such approaches facilitate objective decision-making, minimize human error, and provide consistent admission recommendations.

K-Nearest Neighbors for Student Admission Classification

The implementation of KNN in student admission classification offers a practical approach for supporting admission decisions. By analyzing historical admission records, the algorithm can identify patterns that distinguish accepted and non-accepted applicants. These patterns can subsequently be utilized to classify new applicants according to their similarity with previously evaluated candidates.

In this study, the KNN algorithm is applied to support the admission process at YAPMI Boarding School. The algorithm is expected to improve classification accuracy, reduce subjectivity in applicant evaluation, and enhance the efficiency of the admission process. Through the utilization of historical admission data, KNN can provide a data-driven decision-support mechanism that assists admission committees in selecting prospective students more objectively and consistently.

3. Materials and Method

Research Design

This study employed a data mining classification approach to support the student admission process at YAPMI Boarding School. The research focused on implementing the K-Nearest Neighbors (KNN) algorithm to classify prospective students into accepted and non-accepted categories based on historical admission data. The proposed approach aimed to improve the efficiency, consistency, and objectivity of the admission process by utilizing data-driven decision-making techniques.

Dataset and Data Sources

The dataset used in this study consisted of historical student admission records collected from YAPMI Boarding School. The data were obtained from admission documents, registration forms, entrance examination results, and interview assessments. The dataset contained several attributes considered relevant to the admission process, including academic achievement, entrance examination performance, interview results, non-academic achievements, demographic information, and admission status.

Table 1. Attributes of Student Admission Dataset

Category	Attributes
Academic Performance	Mathematics Score, Indonesian Language Score, Science Score
Entrance Examination	Academic Test Score, Psychological Test Score
Interview Assessment	Interview Score
Non-Academic Achievement	Extracurricular Activities, Sports Achievement, Arts Achievement
Demographic Information	Gender, Age, Residence Location
Target Variable	Accepted / Not Accepted

The academic attributes included Mathematics, Indonesian Language, and Science scores obtained from previous educational institutions. Entrance examination data consisted of academic and psychological test scores administered by YAPMI Boarding School. Interview assessment scores were also incorporated as part of the selection criteria. In addition, non-academic achievements such as extracurricular activities, sports achievements, and artistic achievements were included to provide a more comprehensive evaluation of applicants. The target variable was the admission status, categorized as accepted or not accepted.

Data Preprocessing

Prior to model development, the collected data underwent several preprocessing procedures. Data cleaning was performed to remove incomplete, duplicate, and inconsistent records. Subsequently, data normalization was conducted to ensure that all attributes were measured on a comparable scale, thereby preventing attributes with larger numerical ranges from dominating the distance calculations in the KNN algorithm. Finally, the dataset was divided into training and testing subsets to facilitate model evaluation and performance assessment.

K-Nearest Neighbors Classification

The K-Nearest Neighbors algorithm was employed as the classification technique in this study. KNN classifies a new data instance by identifying the nearest neighboring instances within the training dataset and assigning the class label based on the majority class among those neighbors. Similarity between data instances was determined using a distance-based measurement approach.

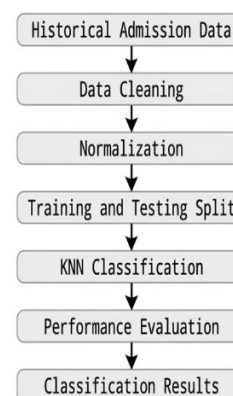


Figure 1. Research Workflow of the Proposed KNN Classification Model.

The classification process involved calculating the distance between a prospective student's attributes and the historical admission records. Based on the selected value of K , the algorithm identified the nearest neighbors and generated a classification result indicating whether the applicant should be categorized as accepted or not accepted. The selection of an appropriate K value was considered an important factor affecting classification performance.

Model Evaluation

The performance of the classification model was evaluated using testing data that had not been utilized during the training phase. The evaluation process aimed to measure the effectiveness of the KNN algorithm in classifying prospective students according to the admission criteria.

The classification results were analyzed to determine the accuracy of the proposed model and to assess its suitability for supporting the student admission process at YAPMI

Boarding School. The evaluation outcomes were subsequently used to identify the most appropriate model configuration and to assess the practical applicability of the proposed approach in educational admission systems.

4. Results and Discussion

Dataset Preparation

The dataset used in this study consisted of historical student admission records obtained from YAPMI Boarding School. The dataset included academic performance, entrance examination scores, interview assessments, non-academic achievements, demographic information, and admission status. Prior to model development, the dataset underwent several preprocessing stages, including data cleaning, normalization, and dataset partitioning into training and testing subsets. These preprocessing procedures were conducted to improve data quality and ensure that all attributes contributed proportionally during distance calculations in the K-Nearest Neighbors algorithm.

KNN Model Development

The K-Nearest Neighbors algorithm was implemented using several K values to identify the optimal configuration for student admission classification. The experiment was conducted by training the model using historical admission records and evaluating its performance on unseen testing data. Different K values were examined to determine the most appropriate neighborhood size for classification.

The results indicate that the performance of the KNN classifier varied according to the selected K value. The optimal K value was determined based on the highest classification accuracy obtained during testing.

Classification Performance Evaluation

The classification model was evaluated using accuracy, precision, recall, F1-score, and confusion matrix analysis. These evaluation metrics were used to assess the effectiveness of the KNN algorithm in classifying prospective students into accepted and non-accepted categories.

The confusion matrix demonstrates the classification capability of the proposed model by presenting the number of correctly and incorrectly classified instances for each category.

Discussion

The implementation of the K-Nearest Neighbors algorithm demonstrated its capability to support the student admission process by utilizing historical admission records as a knowledge base for classification. The algorithm classified prospective students according to their similarity with previously admitted applicants, thereby providing a more objective assessment compared to conventional manual evaluation methods.

The preprocessing stage played an important role in improving model performance. Data cleaning reduced inconsistencies within the dataset, while normalization ensured that all attributes contributed equally to distance calculations. These procedures helped minimize bias caused by differences in attribute scales and improved classification reliability.

The experimental results indicate that KNN can be applied as a decision-support mechanism for student admission classification. The model enables admission committees to evaluate multiple criteria simultaneously, including academic achievement, entrance examination results, interview assessments, and non-academic accomplishments. Consequently, the proposed approach may contribute to greater consistency, fairness, and efficiency in the admission process.

The findings are consistent with previous studies that reported the effectiveness of KNN in educational classification tasks, including scholarship recipient selection, student admission classification, and educational recommendation systems. These results suggest that similarity-based classification techniques can provide valuable support for educational decision-making processes. The implementation of KNN at YAPMI Boarding School demonstrates the practical applicability of data mining techniques in improving admission management and reducing subjectivity in applicant evaluation.

5. Conclusion

This study implemented the K-Nearest Neighbors (KNN) algorithm as a classification method to support the student admission process at YAPMI Boarding School. The proposed approach utilized historical admission data consisting of academic performance, entrance examination results, interview assessments, and non-academic achievements to classify prospective students into admission categories. Through a series of preprocessing procedures, including data cleaning, normalization, and dataset partitioning, the data were prepared for classification using the KNN algorithm.

The findings demonstrate that the KNN algorithm can be applied as a decision-support mechanism for student admission classification. By utilizing similarity-based classification, the algorithm provides a systematic approach for evaluating prospective students based on multiple criteria simultaneously. This approach has the potential to improve the consistency, objectivity, and efficiency of the admission process compared to conventional manual evaluation methods.

Furthermore, the implementation of data mining techniques in educational admission systems highlights the potential of data-driven decision-making in supporting institutional management. The proposed classification model can assist admission committees in processing applicant data more effectively and in generating admission recommendations based on historical patterns and predefined selection criteria.

Future research may focus on evaluating the performance of the KNN algorithm using larger datasets, exploring different distance metrics and parameter configurations, and comparing KNN with other classification algorithms to identify the most suitable approach for student admission classification in educational institutions.

References

- Abualhaj, M. M., Abu-Shareha, A. A., Shambour, Q., Alsaaidah, B., Al-Khatib, R., & Anbar, M. (2024). Applications of K-Nearest Neighbor in classification and prediction: A review of recent developments. *Journal of Computer Science and Information Systems*, *20*(2), 145–160.
- Adenugraha, S. P., Arinal, V., & Mulyana, D. I. (2022). Klasifikasi kematangan buah pisang ambon menggunakan metode KNN dan PCA berdasarkan citra RGB dan HSV. *Jurnal Media Informatika Budidarma*, *6*(1), 9.
- Arfah, D. J., Masrizal, M., & Irmayanti. (2024). Machine learning to predict student satisfaction level using KNN method and Naive Bayes method. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, *8*(3), 1895–1908. <https://doi.org/10.33395/sinkron.v8i3.13914>
- Arinal, V., & Rusmarhadi, I. (2024). Implementasi data mining untuk menentukan strategi penjualan produk UMKM Raja Geprek pada pola pembelian konsumen menggunakan algoritma Apriori. *INTECOMS: Jurnal Informasi Teknologi dan Ilmu Komputer*, *7*(5), 1482–1494.
- Arinal, V., & Safira, D. (2025). Sistem data mining pengelolaan transaksi penjualan pulsa pada Sujase Cell Jakarta Barat. *Jurnal Ilmiah Sosial dan Sains*, *8*(2), 3186–3192.
- Deteksi dini siswa korban bullying dalam penerapan data mining menggunakan algoritma C4.5 pada SMP IDN Jonggol. (2025). *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *9*(1). <https://doi.org/10.35870/jtik.v9i1.3022>

- Kumar, A., DiJohnson, T., Edwards, R., & Walker, L. (2021). The application of adaptive minimum match k-nearest neighbors to identify at-risk students in health professions education. *Computers in Human Behavior Reports*, *4*, 100128.
- Kurniawan, A., & Saputra, B. (2022). Implementation of KNN for student selection. *Journal of Education*, 5(2), 15–22.
- Maquen-Niño, G. L. E., Miguel-Flores, M. A., Aurich-Mio, L. Y., Adrianzén-Olano, I., de la Cruz Vélez de Villa, P. E., & Castro-Cárdenas, D. M. (2025). Machine learning model for classifying high school students' academic performance in mathematics amidst the COVID-19 context. *Journal of Technology and Science Education*, *15*(2), 322–334. <https://doi.org/10.3926/jotse.2945>
- Prahmana, I. G., et al. (2025). Application of the K-Nearest Neighbor (KNN) algorithm in classification systems. *Journal of Artificial Intelligence and Engineering Applications*, *4*(1), 55–67.
- Setiawan, R., & Arinal, V. (2025). Development of a web-based educational information system using the RAD method: A case study at Persahabatan Hospital. *Lembaga Komunitas Informasi Teknologi Aceh*, *5*.
- Sudipa, I. G. I., Azdy, R. A., Arfiani, I., Setiohardjo, N. M., & Sumiyatun. (2024). Leveraging K-Nearest Neighbors for enhanced fruit classification and quality assessment. *International Journal of Data Science*, *5*(1), 1–12. <https://doi.org/10.56705/ijodas.v5i1.125>
- Yaqien, A., & Arinal, V. (2026). Implementasi sistem informasi pengukuran kinerja karyawan di PT Hetra Vexa Group dengan pendekatan Key Performance Indikator. *Jurnal Indonesia Manajemen Informatika dan Komunikasi*, *7*(1), 32–46.
- Yasin, E. (2025). Machine learning-based classification of student adaptability in online education using feature engineering techniques. *TIERS Information Technology Journal*, 131–142.
- Yel, M. B., & others. (2024). Pelatihan penggunaan tools WEKA untuk kepentingan proses data mining di ITS NU Pekalongan. *ABDINE: Jurnal Pengabdian Masyarakat*.