KOMTEKINFO JOURNAL
LPPM Universitas Putra Indonesia YPTK Padang
Jl. Raya Lubuk Begalung, Padang, West Sumatera, Indonesia, Zip Code: 25221
Volume: 13, Issue: 2, Page: 66 - 73, June 30th, 2026, e-ISSN: 2502-8758
Available online at website: https://jkomtekinfo.org/ojs/index.php/komtekinfo

Analysis of Student Selection Models Using K-Means Clustering and K-Nearest Neighbor Classification Algorithms

Imam Fakhri Muhammad[1]✉, Syafri Arlis[2] , Musli Yanto[3]
[1,2,3] Fakultas Ilmu Komputer, Universitas Indonesia YPTK, Padang, 25221, Indonesia
Imamfakhrimuhammad@gmail.com

Abstract

The high level of student interest in the selection process poses challenges, including student admission management. The selection process generally consists of several stages, ranging from administrative tests, academic tests, psychological tests, and physical fitness tests. Based on this, the purpose of this study is to develop an approach that can help evaluate student readiness objectively and based on data. This study aims to analyze student selection by applying the concept of data mining using the K-Means and K-Nearest Neighbor (KNN) algorithms. The K-Means algorithm is used to group student data into several clusters based on the similarity of characteristics. Meanwhile, the K-Nearest Neighbor algorithm works by classifying new data based on similarity or the closest distance. The research dataset consists of 124 student data points obtained from the Arka Padang tutoring center headquarters. Based on this study, the results show that the application of the K-Means and K-Nearest Neighbor (KNN) algorithms demonstrates that both methods are capable of processing student data to identify patterns and levels of readiness for selection, achieving an accuracy of 92.10%. Thus, this method is considered reliable in supporting the process of evaluating student readiness. This research contributes to the understanding of the application of data mining concepts to evaluate and analyze student readiness levels and demonstrates how the K-Means and KNN algorithms can be used in the process of classifying students objectively and based on data.

Keywords: Data Mining, Student Selection, K-Means Clustering, K-Nearest Neighbor (KNN), Student Readiness Evaluation

1. Introduction

The high level of interest among Indonesian students in participating in the selection process for civil service educational institutions demonstrates the appeal of a guaranteed career path and the certainty of appointment as a civil servant through a service bond, resulting in increasingly high levels of competition in the selection process [1]. Given the high level of student interest in the selection process, government educational institutions need to ensure that the selection is conducted objectively and efficiently in order to select the best candidates, considering the intense competition with the number of applicants exceeding the available quota each year [2]. Challenges in this selection process include managing student admissions at tutoring institutions in an efficient, professional, and open manner to ensure that the quality of students accepted meets the criteria set by the organizers [3]. Student readiness in facing the police service selection process is one of the main indicators of the success of pre-service training and education, as it reflects the ability of prospective participants to meet the physical, academic, and mental requirements set by the police institution. This readiness is not only related to mastery of theory or general knowledge, but also includes physical readiness and endurance to participate in all stages of the selection process, from administrative, academic, and psychological tests to physical fitness tests [4]. Coaching through physical training, academic potential test simulations, and administrative assistance has been proven to significantly improve participants' readiness for police recruitment selection [5]. Institutional support in the form of community service activities and physical training guidance is also an important factor in preparing prospective participants to optimally meet the selection standards for civil service schools. Thus, comprehensive readiness involving cognitive, affective, and psychomotor aspects is the main determinant of prospective students' success [6]. A study by Hefny et al. (2024) shows that assessments that incorporate aspects of motivation and psychological resilience increase the validity of selection results [7]. Meanwhile, the results of a study (Seyfried et al., 2024) reveal that an adaptive selection

Submitted: 16-03-2026 | Revised: 15-04-2026 | Accepted: 12-05-2026 | Published: 30-06-2026

approach that considers individual potential gives students the opportunity to demonstrate their optimal abilities after being accepted [8]. In this context, data mining plays a strategic role in analyzing student behavior, performance, and characteristics, thereby supporting academic selection and evaluation [9]. In the context of education, the application of data mining plays an important role in analyzing student behavior, performance, and characteristics so that it can be used to support the academic selection and evaluation process [10]. Research by Staneviciene et al. (2024 used data mining techniques such as classification and regression to predict student academic performance. This study successfully identified students at risk of failure early on, enabling more timely intervention. The results contributed to improving the quality of education by improving student management and increasing student retention [11]. Research conducted by M ydyti et al., 2023 developed a data mining-based system to predict student academic performance. This system helps educational institutions identify students who need attention early on and design more effective interventions, thereby improving overall academic outcomes and student retention [12].

Clustering techniques in data mining are analysis methods that aim to group objects in a dataset based on the similarity of their characteristics [13]. In the field of education, clustering can be used to group students based on attributes such as academic achievement, learning engagement, and attendance, so that institutions can gain a more comprehensive understanding of the variations in student behavior and learning characteristics [14]. One method that can be used to analyze the factors that influence students' readiness to participate in civil service selection is the K-Means Clustering algorithm [15]. This method is included in unsupervised learning techniques in data mining, which functions to group data into several clusters based on the level of similarity between data characteristics [16]. In the context of education, K-Means Clustering can be used to group students based on attributes such as academic grades, physical abilities, motivation, and psychological test results, thereby producing groups of students with different levels of readiness—for example, ready, somewhat ready, and not yet ready. This method has been widely used to support decision-making in education because it simplifies complex data into more meaningful information [17]

The research conducted by Fang applied the K-Means and enhanced Apriori algorithms to analyze student education and employment data. The results of this study showed an increase of more than 90% in understanding of student work practices and their readiness for the world of work, which supports more sustainable education management [18].

Furthermore, research conducted by Andi Akram Nur Risal et al. shows that the application of a data mining approach with the K -Means Clustering algorithm is an effective solution for objectively mapping student readiness by grouping them into homogeneous clusters based on similar attributes such as academic grades, tryout results, and physical and mental readiness, so that educational institutions can adjust their learning programs in a more targeted manner and improve the effectiveness of the teaching and learning process [19].

After the student data has been grouped through preliminary analysis, the next step is to predict their readiness to participate in the police selection process using the K-Nearest Neighbor (KNN) algorithm. This prediction is made by utilizing the similarities between student attributes, such as academic grades, psychological test results, physical abilities, and selection test simulations, to estimate each student's potential to pass the selection process or require additional training. The K-Nearest Neighbor (KNN) algorithm is one of the most popular classification methods in data mining due to its simplicity and effectiveness in recognizing data patterns [20]. KNN works by classifying new data based on similarities or closest distances to a number of neighbors in the training data, so that previously established student readiness patterns can be used to predict each individual's potential for success [21]. Research conducted by (Atmaja et al., 2023) found that the application of the K-Nearest Neighbor (KNN) algorithm was able to accurately predict student graduation based on academic attributes, thereby helping educational institutions to evaluate and provide more targeted guidance [22]. Meanwhile, research by Situmeang et al., 2025) shows that the use of the KNN algorithm in predicting stunting rates in children in East Nusa Tenggara provides fairly good prediction results, so it can be used as a basis for determining health intervention priorities efficiently [23].

Furthermore, research by (Putra, 2025) reveals that the KNN algorithm performs better than the NBC and C4.5 methods in classifying prospective recipients of the Indonesia Pintar (PIP) program, demonstrating the algorithm's great potential for application in the civil service selection process, particularly in assessing the eligibility of prospective participants objectively, measurably, and based on data [24]. Based on this background, this study is a strategic step to provide a more objective and comprehensive picture of the readiness level of police selection participants at Bimbel Arka. Through the application of the K-Means algorithm for segmentation and K-Nearest Neighbor (KNN) for predicting student readiness levels, this study is expected to offer a more accurate and useful analysis model for improving the coaching system.

2. Method

Research methodology is a series of systematic steps used to design, conduct, and evaluate scientific research in order to achieve predetermined objectives. This

methodology includes the selection of research types and approaches, data collection techniques, data analysis methods, and procedures for testing and validating research results.

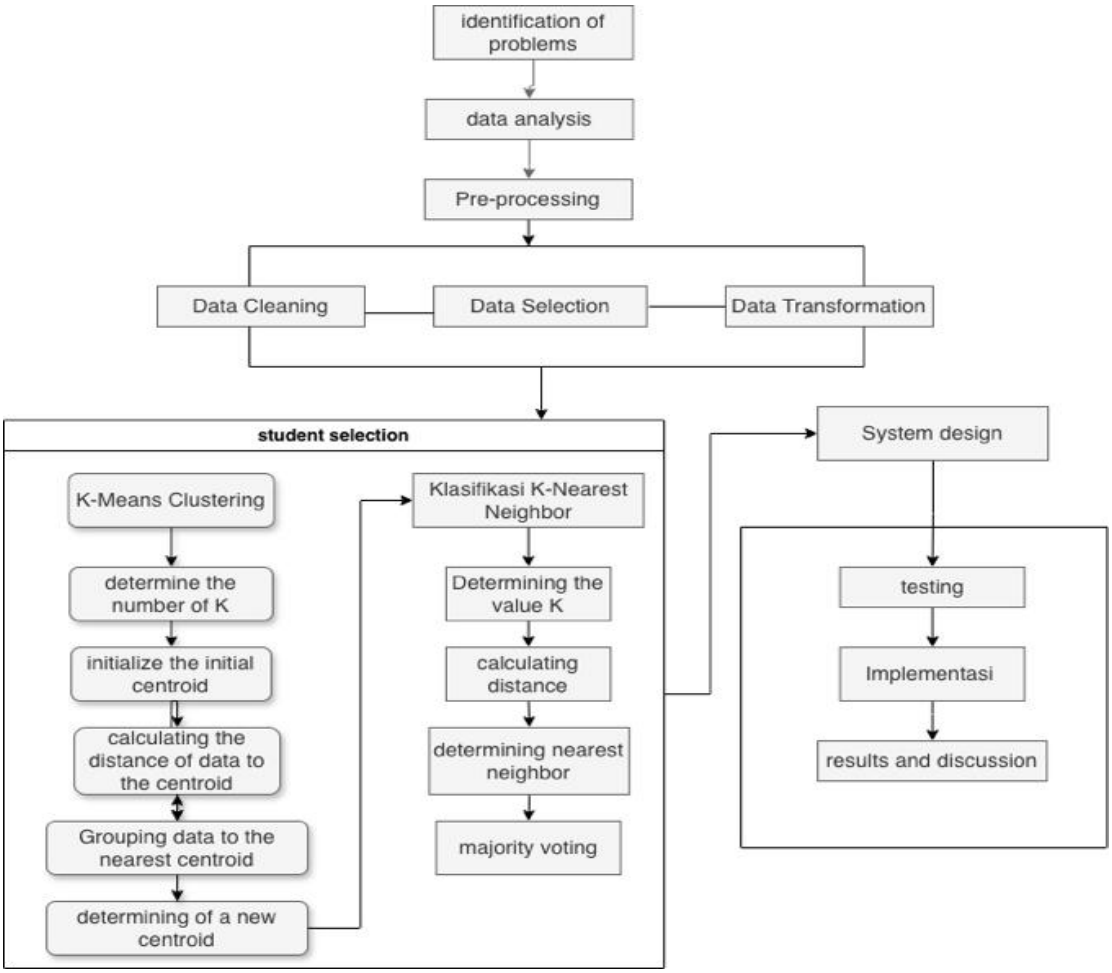

Figure 1. Research Framework

The research process begins with problem identification, data collection, and pre-processing, which includes data cleaning, data selection, and data transformation to ensure that the data is ready for analysis. After that, the K-Means algorithm is used to form clusters through determining the number of clusters, centroid initialization, distance calculation, data grouping, and centroid updating, which then becomes the basis for the classification process using KNN. The final stage includes system design, testing, implementation, and compilation of results and discussion as a form of overall process evaluation.

2.1 K-Means Clustering

The steps in the K-Means Clustering method are an important process for producing optimal data clustering. Each stage must be well understood so that the analysis results are in line with the research objectives. The stages of the K-Means Clustering method are described as follows [25]:

a. determining the value of k

   The first step in the K-Means method is to determine the value of k, which is the number of clusters to be formed. Determining this value is the basis for the data clustering process [26]

b. Determining the Initial Cluster Centers (k-Centroid)

   The next step is to determine the initial cluster center or centroid. The selection of this starting point is very important because it will affect the final results of the clustering process [26].

c. Calculating the Distance of Each Data Point to Each Centroid

After the cluster center point is determined, the distance of each data point to the centroid is calculated. This calculation uses Euclidean Distance, which is a formula for measuring the distance between objects and the cluster center, as shown in Formula 1.

$$D_{ij} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

Where D(X,Y) is the Euclidean distance between data X and centroid Y, x_i represents the value of variable i in data X, y_i represents the value of variable i in centroid Y, and n is the number of variables used in the distance calculation process.

d. Data Grouping

Based on the distance calculation results, each data point is then grouped into clusters according to the closest distance between the data point and the centroid.

e. Determining the New Centroid

The new centroid is calculated based on the average value of the data within the same cluster. This process is repeated until there are no further changes in cluster position. Thus, data with similar characteristics will be in the same group, while different data will be separated into different groups. The new centroid can be determined using Equation 2.2 [26].

$$D_{ij} = \frac{x1 + x2 + x3 + \cdots + xn}{\sum x} \quad (2)$$

At this stage, X_n is the value of the nth record, while ∑x represents the total number of data records used in the calculation. The grouping process is carried out gradually and repeatedly through several iterations to obtain the most optimal results, where each iteration aims to refine the cluster division to be more accurate so that internal variation within each group can be minimized optimally [27].

2.2 K-Nearest Neighbor

K-Nearest Neighbor (K-NN) is a simple yet efficient classification technique that is widely used in various machine learning and data processing applications. Its working principle is based on the assumption that similar objects tend to be close to each other in feature space [28]. The following is a complete explanation of the K-NN algorithm stages [29]:

a. Determining the Value of K

Determining the value of k is important because the larger the value of k, the more information from neighbors is considered. However, a value of k that is too large can cause the model to become more general and less sensitive to noise in the data. Therefore, choosing the right k greatly affects the classification results [29].

b. Calculating Distance.

After the k value is determined, the next step is to calculate the distance between the test data (the data to be predicted) and all the training data. This process is done using a distance metric to measure how close the test data is to each piece of training data. One of the most commonly used metrics is Euclidean distance, which is calculated using equation (2.3) [29]:

$$D = \sqrt{\sum_{i=1}^{n}(x_1 - y_2)^2} \quad (3)$$

The above formula explains that D is the Euclidean distance between two data points, n represents the number of attributes or dimensions used in the calculation, x_1 and y_2 are the coordinate values or attributes at the two points being compared, while the symbol Σ (sigma) indicates the summation process of the squared difference of each attribute in the distance calculation.

c. Determining the Closest Neighbors.

After calculating the distance, we select the K closest neighbors based on the smallest distance that has been calculated. In this step, we select the K training data that has the smallest distance from the test data. By selecting the closest neighbors, we ensure that the classification is based on the most relevant data [29]

d. Majority Voting

The final step in the K-NN algorithm is to perform majority voting to determine the class of the test data. After finding the K nearest neighbors, we will look at the classes of those neighbors and select the class that appears most frequently. The class that appears most frequently among the nearest neighbors will be the prediction for the test data [29].

3. Results and Discussions

This study aims to analyze the application of the K-Means Cluster and K-Nearest Neighbor (KNN) algorithms in the student selection process. In general, the study begins with the collection and preprocessing of student data to ensure data quality before further processing.

Next, the data is grouped using the K-Means algorithm to form clusters based on student characteristics. The cluster results are then used in the classification process using the KNN algorithm, which ends with a testing and evaluation stage to assess the performance of the student selection model.

3.1 Prepocessing

Data preprocessing is the initial stage in research that aims to prepare data so that it can be processed optimally in the next stage of analysis. The data used in this study was sourced from a dataset of students at the Arka Padang tutoring institution. This stage plays a very important role in ensuring data quality, so that the data used is accurate, consistent, and free from errors. The data pre-processing process in this study consists of three main stages, namely data cleaning, data selection, and data transformation.

Table 1 Dataset

No	X1	X2	X3	X4
1	91,05	39	75,179	46
2	37,1	39	75,391	29
3	56,347	53	68,143	23
4	74,429	41	73,731	44
5	47,992	18	69,03	23
6	66,338	12,5	72,933	29
7	73,475	45	74,692	48
8	61,038	33,5	71,882	51
9	68,917	61	74,08	52
10	91,05	39	75,179	46

The table shows a student dataset consisting of four variables, namely X1, X2, X3, and X4, which represent scores for each aspect of student readiness assessment. Each row of data describes one student object used as input data in the study. This dataset is then used as initial data before pre-processing and analysis using data mining methods.

3.2 Analysis / Process of the K-means Clustering Algorithm Method

The first method used in this study was the K-Means Clustering algorithm, which aims to group student data into several clusters based on the similarity of their characteristics. This process was carried out to obtain an initial grouping of students before the classification stage. Data that had undergone the preprocessing stage was then analyzed using the K-Means algorithm by determining the number of clusters (K) appropriate to the research needs.

The K-Means algorithm begins by randomly determining the initial cluster centers (centroids). Next, the distance between each student data point and the centroid is calculated using Euclidean distance. The data is assigned to the cluster with the closest distance. This process is carried out iteratively until the position of the centroid does not change or the cluster results have converged.

Table 2. k-means clustering results

No	X1	X2	X3	X4	Y
1	91,05	39	75,179	46	2
2	37,1	39	75,391	29	3
3	56,347	53	68,143	23	3
4	74,429	41	73,731	44	2
5	47,992	18	69,03	23	3
6	66,338	12,5	72,933	29	2
7	73,475	45	74,692	48	2
8	61,038	33,5	71,882	51	2
9	68,917	61	74,08	52	2
10	91,05	39	75,179	46	2

Of the total 124 data points, 70% or 86 data points were selected as training data, while the remaining 30% were used as test data. This separation stage aims to train the K-Nearest Neighbor (KNN) model to recognize patterns in known data, while also evaluating the model's ability to classify new data that is not included in the training data. With this division, the testing process becomes more objective, so that the results of classifying student readiness based on physical, academic, psychological, and cognitive variables can be measured accurately.

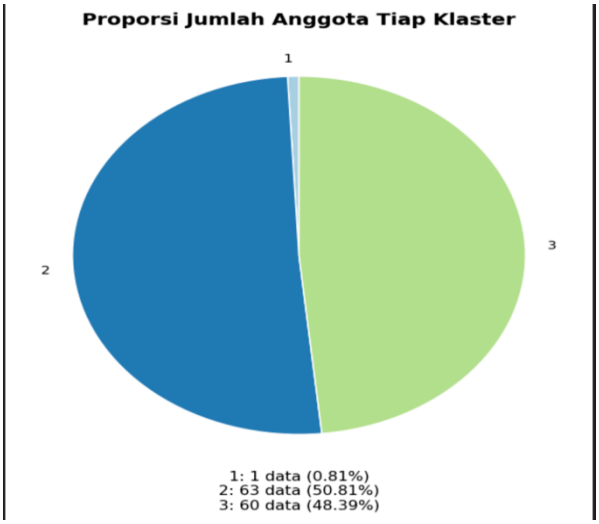

Figure 3 pie chart

The proportion diagram shows that Cluster 2 has the largest number of members with 63 data points or around 50.81% of the total data. Cluster 3 is in second place with 60 data points, or about 48.39%, which shows that the number of members is almost equal to Cluster 2. Meanwhile, Cluster 1 consists of only 1 data point, or about 0.81%, so it can be said to be the cluster with the fewest members.

3.3 Analysis / Process k-nearest neighbor method

The K-Nearest Neighbor (KNN) method was used in this study to classify students based on the similarity of data characteristics. This method works by comparing student data to be classified with a number of the closest training data based on a certain distance value. The purpose of applying KNN is to determine the student selection category objectively based on existing data patterns.

Table 1. KNN results

No	1	2	3	Actual	Preduction
87	0	5	0	2	2
88	0	0	5	3	3
89	0	1	4	3	3
90	0	0	5	3	3
91	0	5	0	2	2
92	0	5	0	2	2
93	0	5	0	2	2
94	0	5	0	2	2
95	0	4	1	2	2
96	0	2	3	3	3

The KNN process begins by determining the value of K as the number of closest neighbors used in the classification process. Next, the distance between the test data and the training data is calculated using a distance calculation method, such as Euclidean Distance. Student categories are determined based on the class that appears most frequently among the K closest neighbors, resulting in a classification decision that corresponds to the characteristics of the student data.

```
Accuracy : 0.9210526315789473
Precision: 0.9220109096270088
Recall   : 0.9210526315789473
F1-Score : 0.9208875848156978
```

Figure 2 performance evaluation results of the classification model

The evaluation results show that the model has an accuracy rate of 0.92, indicating excellent classification capabilities. Precision and recall values, both in the range of 0.92, show a good balance between prediction accuracy and the model's ability to recognize data correctly. In addition, an F1-Score value of 0.92 indicates that the overall performance of the model is optimal and consistent.

4. Conclusions

The evaluation stage was conducted to measure the performance of the K-Nearest Neighbor (KNN) classification model in predicting student readiness levels based on K-Means clustering results. The data was divided into training data and test data with a ratio of 70% and 30%, and the K parameter value was set at 5 with three class categories, namely very ready, quite ready, and not yet ready.

The model's performance was evaluated using accuracy, precision, and recall metrics. Accuracy was used to assess the overall accuracy of the model, while precision and recall were used to measure the accuracy and ability of the model in recognizing each category of student readiness.

Table 4.

Experiment	Akurasi	Presisi	Recall
124 data	92%	92%	92%

The evaluation stage was conducted to measure the performance of the K-Nearest Neighbor (KNN) classification model in predicting student readiness levels based on K-Means clustering results. The data was divided into training data and test data with a ratio of 70% and 30%, and the K parameter value was set at 5 with three class categories, namely very ready, quite ready, and not yet ready.
The model's performance was evaluated using accuracy, precision, and recall metrics. Accuracy was used to assess the overall accuracy of the model, while precision

and recall were used to measure the accuracy and ability of the model in recognizing each category of student readiness.

References

[1] F. Hermawan Hadi Prasetyo, D. Prajna, A. Prasetyo, dan Shinta Widyaningtyas, J. Teknologi Pertanian, and P. Negeri Jember, "Pengembangan Minat Studi Lanjut: Sosialisasi Strategi Masuk Sekolah Kedinasan dan Simulasi Try out Berbasis CAT (Computer Assisted Test) di MA Al Ishlah Jenggawah Jember," *CEMARA: Jurnal Pengabdian Masyarakat Multidisiplin*, vol. 2, no. 2, 2024.

[2] F. A. N. Rabani, "Analisis Minat Siswa Melanjutkan Studi Ke Perguruan Tinggi Sebagai Bentuk Investasi Pendidikan Untuk Meningkatkan Perekonomian," *Jurnal Pendidikan Sultan Agung*, vol. 3, no. 2, pp. 113–122, 2023.

[3] F. Maulana, J. Jaenudin, M. Muhaimin, A. A. Maulana, and R. G. Suyatna, "Analisis Strategi Manajemen Penerimaan Peserta Didik Baru di Madrasah Al-Irfan, Kesawon, Serang," *Profit: Jurnal Manajemen, Bisnis dan Akuntansi*, vol. 4, no. 1, pp. 245–254, 2025.

[4] Sannishara, "TINGKAT KESAMAPTAAN JASMANI SISWA SEKOLAH POLISI NEGARA (SPN) SELOPAMIORO TAHUN 2021," 2023.

[5] S. A. Almalki, A. H. AlJameel, Z. Alghomlas, T. Alothman, and F. Alhajri, "Assessing the predictive validity of pre-admission criteria on dental students' academic performance: a cross-sectional study," *BMC Oral Health*, vol. 24, no. 1, p. 90, 2024.

[6] "formoza.ris".

[7] A. F. Hefny *et al.*, "Relationship between admission selection tools and student attrition in the early years of medical school," *J. Taibah Univ. Med. Sci.*, vol. 19, no. 2, pp. 447–452, 2024.

[8] M. Seyfried, S. Hollenberg, and J. Heße-Husain, "Student selection in higher education–the organisational performance dilemma," *Journal of Higher Education Policy and Management*, vol. 46, no. 6, pp. 671–686, 2024.

[9] Satrio Junaidi, R. Valicia Anggela, and D. Kariman, "Klasifikasi Metode Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naïve Bayes, Random Forest, Support Vector Machine (SVM) dan Artificial Neural Nerwork (ANN)," *Journal of Applied Computer Science and Technology*, vol. 5, no. 1, pp. 109–119, Jun. 2024, doi: 10.52158/jacost.v5i1.489.

[10] D. Khairy, N. Alharbi, M. A. Amasha, M. F. Areed, S. Alkhalaf, and R. A. Abougalala, "Prediction of student exam performance using data mining classification algorithms," *Educ. Inf. Technol. (Dordr).*, vol. 29, no. 16, pp. 21621–21645, Nov. 2024, doi: 10.1007/s10639-024-12619-w.

[11] E. Staneviciene, D. Gudoniene, V. Punys, and A. Kukstys, "A Case Study on the Data Mining-Based Prediction of Students' Performance for Effective and Sustainable E-Learning," *Sustainability (Switzerland)*, vol. 16, no. 23, Dec. 2024, doi: 10.3390/su162310442.

[12] H. Mydyti, A. Kadriu, and M. Pejic Bach, "Using Data Mining to Improve Decision-Making: Case Study of A Recommendation System Development," *Organizacija*, vol. 56, no. 2, pp. 138–154, May 2023, doi: 10.2478/orga-2023-0010.

[13] A. BODAGHI, B. C. M. FUNG, and K. A. SCHMITT, "Publishered in ACM Transactions on the Web (TWEB)," 2025.

[14] A. Mühling and G. Große-Bölting, "Novices' conceptions of machine learning," *Computers and Education: Artificial Intelligence*, vol. 4, p. 100142, 2023.

[15] W. Herulambang, M. Mahaputra Hidayat, N. A. Putri, and M. M. Hidayat, "Educational data mining for mapping the comprehension of personality subject using K-Means algorithm (case study of SD Insan Mulya)," 2023.

[16] J. makmun Jemakmun and R. A. Dicky Syarief Purboyo, "Data Clustering Recommendations For Selection Student Majors To Higher Edication Using The K-Means Method (Case Study of SMAN 2 Palembang)," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 6, no. 2, pp. 367–377, Jan. 2023, doi: 10.31289/jite.v6i2.7911.

[17] M. Qusyairi, Zul Hidayatullah, and Arnila Sandi, "Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow," *Infotek: Jurnal Informatika dan Teknologi*, vol. 7, no. 2, pp. 500–510, Jul. 2024, doi: 10.29408/jit.v7i2.26375.

[18] F. Fang, "A Study on the Application of Data Mining Techniques in the Management of Sustainable Education for Employment," *Data Sci. J.*, vol. 22, no. 1, 2023, doi: 10.5334/dsj-2023-023.

[19] Andi Akram Nur Risal, Dyah Darma Andayani, Muh Ilham Suherman, and Andi Baso Kaswar, "Utilizing the K-Means Clustering Algorithm for Analyzing Student Achievement Assessment at SMK Negeri 1 Gowa," *Journal of Embedded Systems, Security and Intelligent Systems*, vol. 05, no. March, pp. 60–67, 2024, doi: 10.59562/jessi.v5i1.2178.

[20] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *J. Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-024-00973-y.

[21] N. Ahmad, S. Hafizh, and R. Sulthanah, "Prediksi kelulusan mata kuliah mahasiswa Teknologi Informasi menggunakan algoritma K-Nearest Neighbor," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 14, no. 2, pp. 135–149, 2024.

[22] I. P. Y. P. Atmaja, G. I. Setiawan, I. W. Dika, and I. A. P. F. Imawati, "DATA MINING MEMPREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN METODE K-NEAREST NEIGHBORS (KNN) STUDI KASUS UNIVERSITAS PGRI MAHADEWA INDONESIA," *Jurnal Manajemen dan Teknologi Informasi*, vol. 13, no. 2, pp. 86–94, 2023.

[23] N. Situmeang, E. A. Komarsyah, A. Fauzi, and J. Sagala, "Penggunaan Algoritma K-Nearest Neighbor (KNN) Untuk Mendeteksi Stunting Pada Anak," *Innovative: Journal Of Social Science Research*, vol. 5, no. 3, pp. 8401–8417, 2025.

[24] T. A. Putra, "KLASIFIKASI PENERIMA PROGRAM INDONESIA PINTAR MENGGUNAKAN ALGORITMA KNN, NBC, DAN C4. 5," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 4, 2025.

[25] L. P. K. Maharani, A. N. P. Angin, P. R. Wardhana, and F. Achmad, "Evaluasi Performa Mahasiswa pada Pembelajaran Mata Kuliah Data Analitik Menggunakan K-Means Clustering: Studi Kasus di Telkom University," *Jurnal Konatif: Jurnal Ilmiah Pendidikan*, vol. 1, no. 2, 2023.

[26] M. P. A. Ariawan, I. B. A. Peling, and G. B. Subiksa, "Prediksi Nilai Akhir Matakuliah Mahasiswa Menggunakan Metode K-Means Clustering (Studi Kasus: Matakuliah Pemrograman Dasar)," *J. Nas. Teknol. dan Sist. Inf*, vol. 9, no. 2, pp. 122–131, 2023.

[27] I. Mikrou and N. S. Sapidis, "A Systematic Evaluation of Clustering Algorithms Against Expert-Derived Clustering," in *Operations Research Forum*, Springer, 2025, p. 55.

[28] S. Daulay and R. Wandri, "Sistemasi: Jurnal Sistem Informasi Integrating K-Means Clustering and K-Nearest Neighbor Classification for Effective Scholarship Recipient Selection." [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[29] H. Zhou, *Learn Data Mining Through Excel*, 2nd ed. Apress Berkeley, CA, 2023. doi: 10.1007/978-1-4842-9771-1_10.

[30] A. Ramadhanu, H. Hendri, F. Hadi, D. Guswandi, D. M. Putra, R. Hardianto, and S. D. Rizki, "Hybrid decision support system and image processing for classifying priority applications in the Padang Government," CSRID (Computer Science Research and Its Development Journal), vol. 18, no. 1, pp. 178–191, 2026.

[31] A. Ramadhanu, H. Hendri, and F. Hadi, "Organic fertilizer content detection based on image segmentation and texture analysis," in Proc. 2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Dec. 2025, pp. 50–55.

Biographies of Authors

	**Imam Fakhri Muhammad** Imam Fakhri Muhammad was born on May 3, 1998. He graduated from the Computer Science Underggrade Program (S1) at STMIK Indonesia Padang and completed his education in 2021. Currently, he is pursing a Master's degree (S2) in the Infformation Technology Jakarta on July 07, 1989. He is a lecturer at Universitas Putra Indonesia YPTK Padang.
	Musli Yanto was born in Jakarta on July 07, 1989. He is a lecturer at Universitas Putra Indonesia YPTK Padang. Educational background since 2014 has completed postgraduate studies in the field of informatics engineering. His skills include data science analysis, algorithms and programming, big data and artificial intelligence. She can be contacted at email: musli_yanto@upiyptk.ac.id.
	**Syafri Arlis** Syafri Arlis was born in Padang on October 23, 1986. His undergraduate study was completed in 2009 at Universitas Putra Indonesia YPTK. He completed his master's degree at Putra Indonesia University, YPTK Padang. He currently serves as a lecture in the Informatics Engineering study program at the Universitas Putra Indonesia YPTK Padang. Teaching history that has been carried out starting from 2011 until now, such as databases and digital image processing. Published research history places more emphasis on the atificial neural network and digital image processing. He can be contacted at email: syafri_arlis@upiyptk.ac.id.