Vol.7 No.1 Tahun 2026
Jurnal Kecerdasan Buatan, Komputasi dan Teknologi Informasi
https://ejournal.unuja.ac.id/index.php/core
E-ISSN : 2774-7875 and P-ISSN : 2775-0124

K-Means Clustering for Prospective Student Interest Segmentation Based on Study Program Preferences

Riszki Fadillah[1], Intan Nur Fitriyani[2], Desi Irfan[3]
[1] Ika Bina Institute of Technology and Health, Labuhanbatu, Indonesia

Article Info

Article history:

Received : 05 08, 2026
Revised : 05 21, 2026
Accepted : 06 07, 2026

Keywords:

K-Means;
Interest Segmentation;
Data Mining;
Clustering;

Abstract

Higher education institutions in the Labuhanbatu Raya region face challenges in understanding the interests of prospective students. Data collected from 1,722 prospective students at 29 high schools has been stored but has not yet been optimally utilized to support student recruitment and admissions strategies. This study aims to apply the K-Means algorithm to cluster prospective students' interests based on their preferred degree programs. The research methodology included data collection, feature selection, data cleaning, transforming categorical data to numerical using label encoding, determining the optimal number of clusters using the Elbow method, clustering with the K-Means algorithm, evaluating cluster quality using the Silhouette Score, and visualizing and interpreting the clustering results. The findings indicate that the optimal number of clusters is K = 3, with a Silhouette Score of 0.3695, indicating fairly good clustering quality. The cluster distribution shows that Cluster 0 consists of 19 schools (65.52%) with relatively homogeneous interest patterns, Cluster 1 consists of 9 schools (31.03%) with more diverse interest patterns, and Cluster 2 consists of only 1 school (3.45%) with a very distinct interest profile. The K-Means algorithm proved effective in clustering schools based on prospective students' program preferences.

Corresponding Author:

Riszki Fadillah,
fadillahriszki@gmail.com.

INTRODUCTION

Higher education plays a vital role in producing competent, innovative, and adaptable human resources capable of responding to advances in science and technology. Universities not only provide formal education but also serve as institutions that develop competencies aligned with labor market demands and regional development. In recent years, competition among higher education institutions has intensified, requiring universities to gain a deeper understanding of the characteristics and preferences of prospective students(Syafrudin et al., 2025). Such understanding is essential for developing effective promotional strategies, improving study programs, and formulating data-driven student admission policies (Renita et al., 2024).

The Labuhanbatu Raya region, comprising Labuhanbatu, North Labuhanbatu, and South Labuhanbatu Regencies in North Sumatra Province, produces a substantial number of graduates from Senior High Schools (SMA), Vocational High Schools (SMK), and Islamic Senior High Schools (MA) each year. These graduates represent the primary target for university admission campaigns. However, prospective students exhibit diverse characteristics

DOI: https://doi.org/10.33650/coreai.v6i2

and preferences when selecting study programs. While some are interested in Information Technology, others prefer Health Sciences, Education, Economics, or other academic disciplines. Understanding these differences is essential for universities to formulate appropriate marketing strategies and academic development plans (Wahyudin & Dikananda, 2024).

In practice, data collected from school visits, educational exhibitions, and student interest surveys are often used solely for administrative purposes. Nevertheless, these datasets contain valuable information that can reveal patterns of student interest based on school origin, geographical location, and preferred study programs (Tuti Hartati; Nurdiawan Odi; Wiyandi Eko;, 2022). Analyzing such data enables universities to identify groups of prospective students with similar characteristics, thereby supporting more focused and effective promotional activities (Bellanov & Nurhayati, 2023).

Advances in information technology have encouraged organizations to utilize data as a foundation for decision-making. One widely adopted approach for extracting knowledge from large datasets is data mining. Data mining is the process of discovering patterns, relationships, and hidden information from large collections of data to generate meaningful knowledge that supports organizational decision-making (Hendrastuty, 2024). In education, data mining has been extensively applied to analyze student behavior, evaluate learning outcomes, and support data-driven institutional planning (Hilyawan et al., 2025).

Among various data mining techniques, clustering is one of the most widely used unsupervised learning methods. Clustering aims to group objects according to the similarity of their characteristics without requiring predefined class labels, making it particularly suitable for analyzing prospective student data that have not been previously categorized (Setiawan et al., 2024). Through clustering, objects with similar characteristics are assigned to the same group, whereas those with different characteristics are placed into separate groups (Sofia & Hema, 2024).

The K-Means algorithm is one of the most popular clustering methods because of its simplicity, ease of implementation, and ability to produce reliable clustering results (Maukar et al., 2022). The algorithm partitions data into several clusters by assigning each observation to the nearest cluster centroid. The centroids are then iteratively updated until convergence is achieved (Alfitra et al., 2025). Numerous studies have demonstrated that K-Means offers high computational efficiency and performs well when clustering large datasets (Simbolon et al., 2026).

In the educational domain, K-Means has been applied in various studies. Hendrastuty utilized K-Means to cluster student learning evaluation results and identify patterns that could improve instructional quality (Hendrastuty, 2024). Other studies have shown that K-Means can identify patterns in student learning behavior and attendance, enabling educators to better understand student characteristics (Renita et al., 2024).

K-Means has also been widely employed for student and prospective student segmentation. Wahyudin and Dikananda successfully grouped students according to their study program preferences based on shared interests (Wahyudin & Dikananda, 2024). Nopriandi and Haswan applied K-Means to analyze the tendencies of new students in selecting study programs, providing useful information for academic development strategies (Nopriandi & Haswan, 2022). Similarly, Maukar et al. demonstrated that clustering admission data can generate valuable insights for designing university promotional strategies (Maukar et al., 2022).

Furthermore, Syafrudin et al. reported that clustering-based segmentation of new student data assists universities in developing more effective and targeted promotional strategies (Syafrudin et al., 2025). Hartati et al. applied K-Means to identify prospective student characteristics as a basis for campus marketing strategies (Tuti Hartati; Nurdiawan Odi; Wiyandi Eko;, 2022). Bellanov and Nurhayati also showed that clustering results can be

used to identify potential geographical areas contributing significantly to student enrollment (Bellanov & Nurhayati, 2023).

More recent studies have integrated K-Means with other techniques to improve clustering performance. Rahmawati et al. combined Principal Component Analysis (PCA) with K-Means to obtain a more comprehensive analysis of prospective student characteristics (Rahmawati et al., 2024). Meanwhile, Hilyawan et al. emphasized the importance of cluster validation methods such as the Elbow Method and Silhouette Score for determining the optimal number of clusters (Hilyawan et al., 2025).

Although numerous studies have investigated student and prospective student segmentation, research specifically focusing on the segmentation of prospective students' interests based on study program preferences in the Labuhanbatu Raya region remains limited. Considering the region's significant potential as a source of university applicants in North Sumatra, there is a need to transform existing student interest data into meaningful information that can support institutional decision-making.

Therefore, this study implements the K-Means algorithm to segment prospective students according to their preferred study programs in the Labuhanbatu Raya region. The selection of K-Means is based on its effectiveness in identifying similarity patterns within datasets and producing clusters that are easy to interpret (Alfitra et al., 2025), (Setiawan et al., 2024). The resulting segmentation is expected to provide insights into the distribution of prospective students' interests across different study programs, thereby supporting the development of targeted promotional strategies, student admission planning, and study program development.

Accordingly, this study, entitled "K-Means Clustering for Prospective Student Interest Segmentation Based on Study Program Preferences," aims to provide valuable information that supports data-driven decision-making in higher education institutions while enriching the literature on the application of data mining, particularly the K-Means clustering algorithm, in higher education research (Ananda et al., 2025; Arifin et al., 2026).

METHOD

Establishing a robust methodology for collecting and analyzing data ensures that the process is reliable and reproducible. This workflow culminates in the interpretation of the analysis results, the drawing of objective conclusions, and the preparation of scientific papers or reports to share these new insights with the scientific community.

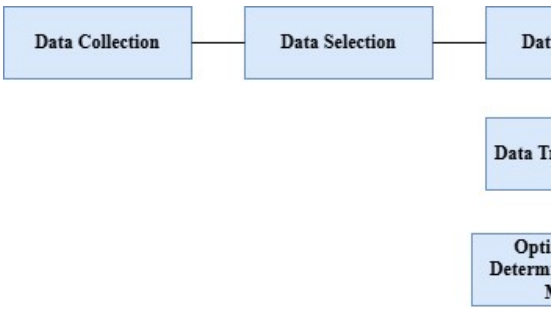

Figure 1. Research Workflow

Data Collection the initial stage involves collecting information on prospective students through school promotion activities and interest surveys. The collected information is stored in spreadsheet format and utilized as the dataset for this study. The data used consist of information on prospective students obtained from promotional activities and interest surveys conducted at Senior High Schools (SMA), Vocational High Schools (SMK), and Islamic Senior High Schools (MA) located in Labuhanbatu Regency, North Labuhanbatu Regency,

DOI: https://doi.org/10.33650/coreai.v6i2

and South Labuhanbatu Regency. The dataset consists of 1,722 prospective student records (Syafrudin et al., 2025).

Data Selection in this phase, feature selection is performed based on the research objectives. The variables used are: School, Address, Interest (Study Program)

Meanwhile, attributes such as the student's name, parent's name, and phone number are excluded because they do not influence the interest clustering process (Hendrastuty, 2024).

Data Cleaning the selected data are then cleaned to improve data quality. The data cleaning process includes the following steps (Tuti Hartati; Nurdiawan Odi; Wiyandi Eko;, 2022):

a. Removing duplicate records.
b. Handling missing values.
c. Standardizing school names.
d. Standardizing address names.
e. Normalizing study program names.

Data Transformation since the K-Means algorithm can only process numerical data, categorical data must first be converted into numerical format using the Label Encoding method (Rahmawati et al., 2024).

Optimal Cluster Determination (Elbow Method) before performing the clustering process, the optimal number of clusters is determined using the Elbow Method. This method is carried out by calculating the Within-Cluster Sum of Squares (WCSS) for several cluster values (K) (Hilyawan et al., 2025).

The WCSS equation is as follows:

$$WCSS = \sum_{i=1}^{k} \sum_{x \in C_i} (x - \mu_i)^2 \quad \text{..... (1)}$$

The value of K that forms the elbow point on the graph is selected as the optimal number of clusters (Renita et al., 2024).

K-Means Clustering after determining the optimal value of K, the clustering process is performed using the K-Means algorithm. The K-Means algorithm consists of the following steps (Wahyudin & Dikananda, 2024):

a. Determine the number of clusters (K).
b. Select the initial centroids.
c. Calculate the distance between each data point and the centroids.
d. Assign each data point to the nearest cluster.
e. Recalculate the centroids.
f. Repeat the process until the centroids no longer change.
g. Distance is measured using the Euclidean Distance:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad \text{..... (2)}$$

Cluster Evaluation (Silhouette Score) the evaluation stage is conducted to assess the quality of the resulting clusters using the Silhouette Score. The Silhouette Score is calculated using the following equation (Arifin et al., 2026):

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad \text{..... (3)}$$

Results Visualization and Interpretation the final stage involves visualizing and analyzing the clustering results. The visualization is performed using:

a. Elbow Chart.
b. Cluster Distribution Plot.
c. Study Program Distribution Bar Chart.

d. Cluster Profile Table.

The analysis is conducted by examining the characteristics of each cluster based on school of origin, region, and the most preferred study programs.

RESULTS AND ANALYSIS

This study employs the K-Means algorithm to cluster prospective students' interests based on their preferred study programs in the Labuhanbatu Raya region. The dataset consists of 1,722 prospective student records collected through interest surveys conducted at Senior High Schools (SMA), Vocational High Schools (SMK), and Islamic Senior High Schools (MA) across Labuhanbatu Regency, North Labuhanbatu Regency, and South Labuhanbatu Regency.

The research procedure includes data collection, data selection, data cleaning, data transformation, determination of the optimal number of clusters using the Elbow Method, clustering using the K-Means algorithm, cluster evaluation using the Silhouette Score, and visualization and interpretation of the results.

Data Collection and Data Selection

The research data were collected from surveys of students' interests in enrolling at higher education institutions conducted across several Senior High Schools (SMA), Vocational High Schools (SMK), and Islamic Senior High Schools (MA) in Labuhanbatu

Regency, North Labuhanbatu Regency, and South Labuhanbatu Regency. The initial dataset consisted of 1,722 prospective student records containing information on the school of origin, address, and preferred study program.

These data served as the basis for segmenting prospective students' interests using the K-Means algorithm. During the data selection stage, only attributes relevant to the research objectives were retained. The selected attributes are presented below.

Table 1. Initial Data Attributes

Attribute	Description
School	Prospective student's school of origin
Address	Prospective student's place of residence
Interest	Preferred study program

Identity attributes were excluded because they do not influence the clustering process.

Table 2. Initial Dataset

No.	School	Student Name	Parent's Name	Address	Phone Number	Preferred Study Program
1	SMA N 1 Kota Pinang	Grace Dora Novelia Hutabarat	Denni Natal Hutabarat	PT. Nubika Jaya	–	Nursing
2	SMA N 1 Kota Pinang	Bayu Ananda	Jimamri	Simanggir, Kota Pinang	085262922032	Information Technology
3	SMA N 1 Kota Pinang	Pandu Wira	Ridanu	Simaninggir, Kota Pinang	082163147131	Information Technology
4	SMA N 1 Kota Pinang	Frananda Korintus Simbolon	Pangela Tua Simbolon	Lubuk Panjang	081386854821	Information Technology
5	SMA N 1 Kota Pinang	Nadia Siregar	J. Siregar	Blok Songo	–	Nursing
6	SMA N 1 Kota	Ahmad Riski Abli	Mariani Siregar	Sialang Bujing	081370575859	Health Administration

DOI: https://doi.org/10.33650/coreai.v6i2

	Pinang	Harahap				
7	SMA N 1 Kota Pinang	Ranti Ajeng Pratiwi	Wagiran	Dsn. Sijambu	083162851844	Information Technology
…	…	…	…	…	…	…
1721	SMK N 3 Rantau Utara	Suci Dwi Hardani	Safrianto	Jl. Kampung Sawah Gg. Amal II	08386899793	Nursing
1722	SMK N 3 Rantau Utara	Nur Kholiza	Sukardi	Sidodadi PNKA	085165241334	Health Administration

Data Cleaning

The data cleaning stage was performed to improve data quality before the analysis. The following steps were carried out:

a. Removing records with missing values.
b. Removing duplicate records.
c. Standardizing study program names.
d. Correcting spelling errors.

Table 3. Data Cleaning Results

No.	School	Student Name	Parent's Name	Address	Phone Number	Preferred Study Program	Study Program Code
1	SMA N 1 Kota Pinang	Grace Dora Novelia Hutabarat	Denni Natal Hutabarat	PT. Nubika Jaya	–	Nursing	3
2	SMA N 1 Kota Pinang	Bayu Ananda	Jimamri	Simanggir, Kota Pinang	085262922032	Information Technology	10
3	SMA N 1 Kota Pinang	Pandu Wira	Ridanu	Simaning gir, Kota Pinang	082163147131	Information Technology	10
4	SMA N 1 Kota Pinang	Frananda Korintus Simbolon	Pangela Tua Simbolon	Lubuk Panjang	081386854821	Information Technology	10
5	SMA N 1 Kota Pinang	Nadia Siregar	J. Siregar	Blok Songo	–	Nursing	3
6	SMA N 1 Kota Pinang	Ahmad Riski Abli Harahap	Mariani Siregar	Sialang Bujing	081370575859	Health Administration	0
7	SMA N 1 Kota Pinang	Ranti Ajeng Pratiwi	Wagiran	Dsn. Sijambu	083162851844	Information Technology	10
…	…	…	…	…	…	…	…
1627	SMK N 3 Rantau Utara	Suci Dwi Hardani	Safrianto	Jl. Kampung Sawah Gg. Amal II	08386899793	Nursing	3
1628	SMK N 3 Rantau Utara	Nur Kholiza	Sukardi	Sidodadi PNKA	085165241334	Health Administration	0

Data Transformation

The cleaned data were then converted into numerical format using an encoding technique. Subsequently, the number of prospective students interested in each study program was aggregated for each school, resulting in a numerical dataset ready for processing using the K-Means algorithm.

The transformation process produced 29 schools as the clustering objects, with attributes representing the number of prospective students interested in each study program.

Table 4. Data Transformation Results

School	Health Administration	Keb	Midwifery	Information Technology
MAN 1 N 2 Labura	12	0	7	0
MAS Al-Amiin Kp. Pajak	1	0	0	0
MAS Al-Washliyah Marbau	10	0	6	0
SMA/SMK Yapim Taruna Pinang Awan	1	0	1	0
SMA N 1 Kota Pinang	30	0	12	0
SMA N 1 Kualuh Selatan	3	0	9	0
SMA N 2 Pangkatan	8	0	3	0
…	…	…	…	…
SMAN 1 Aek Natas	23	0	17	0
Yayasan Pondok Pesantren Ridho Allah	0	1	1	0

Determination of the Optimal Number of Clusters Using the Elbow Method

The optimal number of clusters was determined using the Elbow Method by calculating the Within-Cluster Sum of Squares (WCSS) for a range of K values.

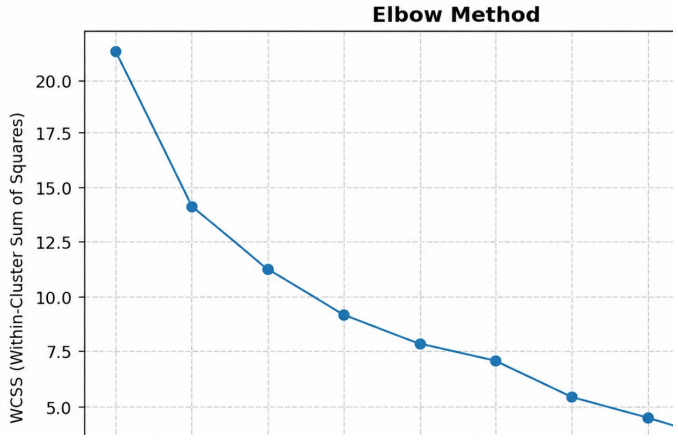

Figure 2. Elbow Method Graph

As shown in Figure 2, the WCSS value decreases substantially across the first few clusters, after which the rate of decrease begins to level off between K = 3 and K = 5. Therefore, clustering was performed using the K-Means algorithm, followed by cluster evaluation to determine the most representative number of clusters.

The evaluation results indicate that K = 3 is the optimal number of clusters for this study, resulting in three groups of schools with distinct prospective student interest characteristics.

K-Means Clustering Results

After determining the optimal number of clusters, the clustering process was performed using the K-Means algorithm with K = 3. The clustering results show that 29 schools were successfully grouped into three clusters based on prospective students' preferences for study programs.

DOI: https://doi.org/10.33650/coreai.v6i2

Table 5. Cluster Distribution

Cluster	Number of Schools	Percentage
Cluster 0	19	65.52%
Cluster 1	9	31.03%
Cluster 2	1	3.45%
Total	**29**	**100%**

The results indicate that the majority of schools belong to Cluster 0, while Cluster 2 contains only a single school with a prospective student interest profile that differs significantly from those of the other schools.

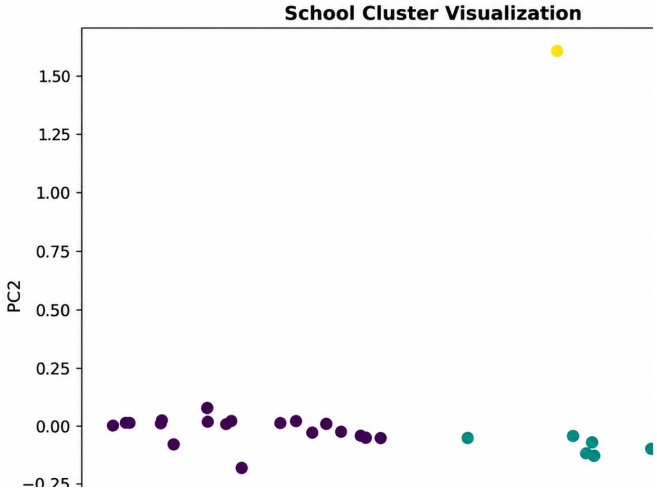

Figure 3. Visualization of School Clustering Results Using K-Means

As shown in Figure 3, the schools are grouped into three distinct clusters. Cluster 0 is the largest group, represented by the collection of points on the left side of the graph. Cluster 1 is located on the right side of the graph and consists of nine schools with interest patterns that differ from those in Cluster 0. Meanwhile, Cluster 2 contains only a single school that is clearly separated from the other groups, forming a unique cluster.

The visualization demonstrates that the K-Means algorithm effectively identifies groups of schools based on prospective students' study program preference patterns. The differences among the clusters are clearly distinguishable, particularly for Cluster 2, which exhibits the most distinct characteristics compared with the other clusters.

Cluster Evaluation Using the Silhouette Score

To assess the quality of the resulting clusters, the Silhouette Score was used. The evaluation produced a score of 0.3695, which falls within the range of 0.26–0.50, indicating that the clustering quality is reasonably good. This result suggests that the objects within each cluster have a relatively high degree of similarity, while the separation between clusters is sufficiently distinct. Therefore, the segmentation results can be used to support the analysis of prospective students' interests in the Labuhanbatu Raya region.

Based on the clustering results obtained using the K-Means algorithm, a total of 29 schools were successfully grouped into three clusters: Cluster 0 consists of 19 schools (65.52%), Cluster 1 consists of 9 schools (31.03%), and Cluster 2 consists of 1 school (3.45%). These findings indicate that the majority of schools exhibit relatively similar interest patterns and are therefore grouped into Cluster 0.

The Silhouette Score of 0.3695 indicates that the clustering quality is reasonably good, suggesting that the resulting segmentation is suitable for describing prospective students' interest patterns based on their preferred study programs.

Overall, the results demonstrate that the K-Means algorithm effectively groups schools according to the characteristics of prospective students' study program preferences. This information can assist higher education institutions in developing more targeted promotional strategies, identifying potential schools, and supporting data-driven decision-making in the student admission process within the Labuhanbatu Raya region.

CONCLUSIONS

Based on the findings of this study, it can be concluded that the K-Means algorithm successfully grouped 29 schools in the Labuhanbatu Raya region into three clusters according to prospective students' study program preference patterns. The optimal number of clusters was determined using the Elbow Method (K = 3), while the clustering quality was evaluated using the Silhouette Score, which achieved a value of 0.3695, indicating reasonably good clustering performance. This result suggests that the objects within each cluster are relatively similar and that clear distinctions exist among the clusters. The cluster distribution shows that Cluster 0 is the largest group, consisting of 19 schools (65.52%), followed by Cluster 1 with 9 schools (31.03%), and Cluster 2 with 1 school (3.45%), representing a school with a distinct and unique interest profile. These findings demonstrate that the K-Means algorithm is effective in identifying groups of schools based on prospective students' study program preferences. The resulting segmentation can assist higher education institutions in developing more targeted promotional strategies, identifying potential schools, and supporting data-driven decision-making in the student admission process. For future work, higher education institutions are encouraged to use the segmentation results as a reference for designing promotional activities, outreach programs, and student recruitment strategies tailored to the characteristics of each school and the identified interest patterns. Future studies are also recommended to utilize larger datasets and incorporate additional relevant attributes, such as students' academic majors in high school, academic achievement, gender, and other demographic factors, to improve clustering accuracy. Furthermore, comparisons with other clustering algorithms, such as K-Medoids, DBSCAN, and Hierarchical Clustering, are recommended to obtain more optimal segmentation results. Such efforts will further enhance the application of data mining in higher education by supporting effective, objective, and data-driven decision-making in student admissions and study program development.

REFERENCES

Alfitra, A., Nurdin, & Meiyanti, R. (2025). Comparison of K-Means and K-Medoids Methods in Clustering High Population Density Areas in Bireuen Regency. *JITE (Journal of Informatics and Telecommunication Engineering)*, *9*(1), 292–302. https://doi.org/10.31289/jite.v9i1.15602

Ananda, M. D., Malik, K. N., Masruriyah, A. F. N., & Mardiah. (2025). Studi Komparatif Algoritma K-Means dan K-Medoids untuk Segmentasi Informasi Kesehatan. *Computer Science (CO-SCIENCE)*, *5*(2), 103–112. https://doi.org/10.31294/coscience.v5i2.9207

Arifin, M., Prastya, I. W. D., & Budiani, J. R. (2026). Evaluating K-Means and K-Medoids Using Silhouette Score for Eysenck Personality-Based Clustering of Prospective Students. *Journal of Applied Informatics and Computing (JAIC)*, *10*(2), 1818–1827. https://doi.org/10.30871/jaic.v10i2.12338

Bellanov, A., & Nurhayati, L. (2023). K-Means Clustering Analysis Untuk Menentukan Strategi Promosi Kampus. *Jurnal Teknik Industri*, *9*(1), 259–268. https://doi.org/10.24014/jti.v9i1.22492

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Juenal Ilmiah Informatika Dan*

Ilmu Komputer (JIMA-ILKOM), *3*(1), 46–56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Hilyawan, F. R., Syafrullah, M., & Sudija, I. (2025). A Systematic Literature Review on the Optimization of K-Means and Agglomerative Clustering for Student Performance Segmentation : A Comparative Analysis of Elbow and Silhouette Methods. *Edcomtech: Jurnal Kajian Teknologi Pendidikan*, *9879*(1), 73–88. https://doi.org/10.17977/um039v10i12025p73-88

Maukar, A. L., Marisa, F., Widodo, A. A., Kamilaningtyas, N., & Nugraha, N. D. (2022). Analisis data penerimaan mahasiswa baru berbasis k-means. *JIKO (Jurnal Informatika Dan Komputer)*, *6*(2), 142–147. https://doi.org/10.26798/jiko.v6i2.558

Nopriandi, H., & Haswan, F. (2022). Analisis Klasterisasi Mahasiswa Baru dalam Memilih Program Studi dengan Menggunakan Algoritma K-Means. *Journal of Information System Research (JOSH)*, *3*(4), 666–671. https://doi.org/10.47065/josh.v3i4.1986

Rahmawati, D. P., Hidayati, S., Arismawati, P., & Johan, A. W. S. B. (2024). Prospective New College Student Dashboard : Insights from K-Means Clustering with Principal Component Analysis. *Inform : Jurnal Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, *9*(2), 137–144. https://doi.org/10.25139/inform.v9i2.8462

Renita, C., Padilah, T. N., & Wahyu, M. A. E. (2024). IDENTIFIKASI POLA BELAJAR DAN ABSENSI SISWA DENGAN TEKNIK CLUSTERING DAN ANALISIS KORELASI. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(5), 10286–10292. https://doi.org/10.36040/jati.v8i5.11057

Setiawan, D., Arsa, D., Fitri, L. E., & Zahardy, F. F. P. (2024). Comparative Analysis of Clustering Approaches in Assessing ChatGPT User Behavior. *COGITO Smart Journal*, *10*(2), 366–379. https://doi.org/10.31154/cogito.v10i2.661.366-379

Simbolon, Y., Giovani, A., & Sipayung, S. P. (2026). Analisis Pengelompokan Minat Belajar Mahasiswa Menggunakan Algoritma K-Means. *Jurnal Ilmu Komputer Dan Informatika*, *2*(3), 108–112. https://doi.org/10.62379/jiki.v3i1

Sofia, G., & Hema, D. (2024). Clustering-based Recommendations for Enhancing Students' Academic Performance by Recognizing Prevalent Assessment Method using Exploratory Data Analysis. *Indian Journal of Science and Technology*, *17*(6), 503–513. https://doi.org/10.17485/IJST/v17i6.2175

Syafrudin, T., Hermawan, A., Avianto, D., & Maulana, I. (2025). Analisis Komparasi Algoritma K-Means Dan K-Medoids Dalam Segmentasi Data Untuk Strategi Promosi Mahasiswa Baru Di Universitas X. *Komputika: Jurnal Sistem Komputer*, *14*(2), 203–211. https://doi.org/10.34010/komputika.v14i2.16698

Tuti Hartati; Nurdiawan Odi; Wiyandi Eko; (2022). Analisis dan penerapan algoritma k-means dalam strategi promosi kampus akademi maritim suaka bahari. *Jurnal Sains Teknologi Transportasi Maritim*, *3*(1), 1–7. https://doi.org/10.51578/j.sitektransmar.v3i1.30

Wahyudin, E., & Dikananda, F. (2024). Segmentasi Minat Mahasiswa Terhadap Program Studi Menggunakan Algoritma K-Means Clustering. *BULLET : Jurnal Multidisiplin Ilmu*, *2*(01), 271–276.