The IJICS (International Journal of Informatics and Computer Science)
Vol 10 No 2, July 2026, Page 125-129
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v10i2.9806

K-Means Clustering for Adaptive Learning Recommendations Based on Student Academic Performance in Vocational Education

Petti Indrayati Sijabat*, Eviyanti Br Barus, Kristin Lourensi Sitompul, Sinta Suwanda, Maira Mau Lydia

Informatics Engineering, STMIK Pelita Nusantara, Medan, Indonesia
Email: 1,*pettisijabat.mamaliam@gmail.com, 2eviyanti1794@gmail.com, 3sriwijaya11121987@gmail.com, 4Sintasuwanda21@gmail.com, 5mairamau@gmail.com
(*: pettisijabat.mamaliam@gmail.com)
Submitted: **16/05/2026**; Accepted: **02/07/2026**; Published: **29/07/2026**

Abstract- Differences in students' academic abilities create challenges in the learning process because many schools still apply uniform learning strategies without considering individual academic characteristics. This study aims to cluster students based on academic performance using the K-Means Clustering algorithm and to generate adaptive learning recommendations for each cluster. The dataset consisted of academic records from 50 students of SMK Swasta 2 Delima Sari in the 2024/2025 academic year, including assignment scores, Midterm Examination (UTS) scores, Final Examination (UAS) scores, and attendance rates. Data preprocessing was conducted through data cleaning and Min-Max normalization to ensure that each variable contributed proportionally to the clustering process. The K-Means algorithm was implemented with k = 3 to form high, medium, and low academic performance groups. The results show that 12 students (24%) were classified into the High Cluster, 28 students (56%) into the Medium Cluster, and 10 students (20%) into the Low Cluster. The clustering quality was supported by a WCSS value of 2.3714 and a Silhouette Score of 0.6843, indicating reasonably well-separated clusters. Based on the cluster profiles, students in the High Cluster are recommended to receive enrichment activities, students in the Medium Cluster receive regular instruction with periodic feedback, and students in the Low Cluster receive remedial learning and structured mentoring. These findings indicate that K-Means Clustering can support data-driven educational decision-making and provide a practical basis for adaptive learning recommendations in vocational education.

Keywords: K-Means Clustering; Educational Data Mining; Adaptive Learning; Student Academic Performance; Vocational Education

1. INTRODUCTION

The development of information technology has driven significant transformations in the world of education, particularly in the management of academic data and the development of more effective learning strategies. Although various educational information systems have been implemented in schools, the learning process in practice still often uses a one-size-fits-all approach, providing the same learning treatment to all students without considering differences in individual academic characteristics. This situation has the potential to cause students with low academic ability to experience difficulties in understanding learning materials, while students with high academic ability do not receive learning challenges appropriate to their potential [1],[2]

In classroom practice, many schools still apply a uniform learning approach to all students. This approach can create problems because students have different levels of academic ability, learning readiness, and learning support needs. Students with low academic performance may need remedial assistance, whereas students with high academic performance may require enrichment activities to develop their potential. Therefore, a data-driven approach is needed to identify student academic profiles and support more targeted learning recommendations.

Educational Data Mining (EDM) provides techniques for extracting meaningful patterns from educational datasets [2]. One method that is widely used for grouping data is K-Means Clustering. K-Means is an unsupervised learning algorithm that groups data based on similarity so that objects within the same cluster have closer characteristics than objects in different clusters [1], [3]. In the educational context, K-Means has been used to group students based on achievement, learning behavior, and academic performance [4],[5],[6],[7],[8],[9],[10].

SMK Swasta 2 Delima Sari has academic data that include assignment scores, Midterm Examination (UTS) scores, Final Examination (UAS) scores, and attendance rates. Nevertheless, these data have not been fully utilized to inform the design of adaptive learning strategies. Teachers generally rely on classroom observations, which are useful but may not always yield measurable, consistent information about students' academic patterns.

Based on this condition, this study applies the K-Means Clustering algorithm to group students by academic performance and to translate the clustering results into adaptive learning recommendations. The research focuses on three pedagogically meaningful categories: high, medium, and low academic performance. The contribution of this study lies in linking cluster profiles directly to practical learning pathways that teachers in vocational education can implement.

2. RESEARCH METHODOLOGY

2.1 Research approach

The IJICS (International Journal of Informatics and Computer Science)
Vol 10 No 2, July 2026, Page 125-129
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v10i2.9806

This study used a quantitative approach within the framework of Educational Data Mining. The objective was to identify patterns in student academic performance through clustering analysis and to convert the resulting clusters into adaptive learning recommendations. The research was conducted using academic data from 50 students of SMK Swasta 2 Delima Sari in the 2024/2025 academic year. The sample was selected purposively, including only students with complete academic records across all measured variables.

Four variables were used in the clustering process: assignment scores, Midterm Examination (UTS) scores, Final Examination (UAS) scores, and attendance rates. These variables were selected because they represent students' cognitive achievement and participation in learning. Personal identifiers, such as student names and identification numbers, were anonymized before analysis to maintain confidentiality.

2.2 Research Procedure

The research procedure consisted of seven main stages: data collection, data preprocessing, Min-Max normalization, K-Means Clustering implementation, cluster evaluation, adaptive learning recommendation development, and report writing. The research procedure is shown in Figure 1.

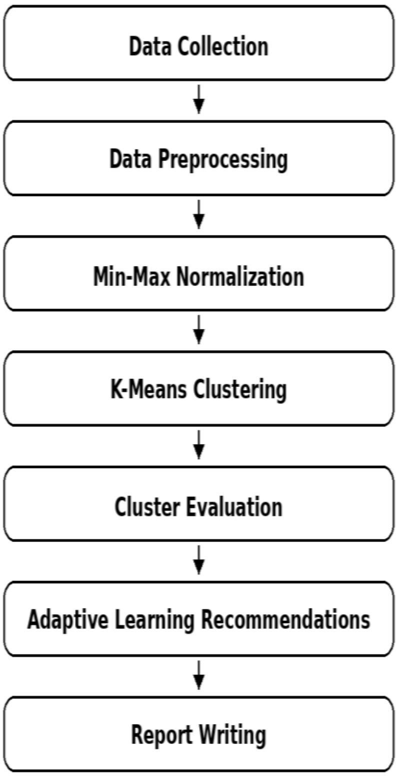

Figure 1. Research Procedure

2.3 Data Preprocessing and Normalization

Before clustering, the dataset was checked to identify incomplete data, duplicate records, and logically inconsistent values. Records that did not meet the completeness criteria were excluded from the analysis [1]. Because the variables had different value ranges, Min-Max normalization was applied to transform all attributes into the range [0, 1]. This normalization prevented variables with larger numeric ranges from dominating the distance calculation in K-Means [4].

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where X' is the normalized value, X is the original value, X_min is the minimum value of the attribute, and X_max is the maximum value of the attribute.

2.4 K-Means Clustering Algorithm

K-Means Clustering is an unsupervised learning algorithm that partitions data into k clusters based on the distance between each data point and the cluster centroid [3]. In this study, k was set to 3 to represent high, medium, and low academic performance groups. The algorithm began by initializing the centroid positions, calculating the distance between

The IJICS (International Journal of Informatics and Computer Science)
Vol 10 No 2, July 2026, Page 125-129
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v10i2.9806

each student record and each centroid, assigning each record to the nearest centroid, and updating the centroid until convergence was reached.

The Euclidean Distance formula used to calculate the distance between a student data point and a centroid is expressed as follows:

$$d(x, c_j) = [\sum_{i=1}^{n}(x_i - c_{ji})^2]^{1/2} \quad (2)$$

where d(x, c_j) is the distance between data point x and centroid c_j, x_i is the value of the i-th attribute in the student data, c_ji is the value of the i-th attribute in centroid j, and n is the number of attributes.

After all data points were assigned to clusters, the centroid of each cluster was recalculated using the average value of all members in the cluster:

$$C_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_i \quad (3)$$

where C_j is the centroid of cluster j, n_j is the number of data points in cluster j, and x_i is the data vector of each cluster member.

2.5 Cluster Evaluation

The clustering results were evaluated using Within Cluster Sum of Squares (WCSS) and the Silhouette Score. WCSS was used to measure cluster compactness, while the Silhouette Score was used to measure the degree of separation between clusters [5]. The Elbow Method was applied by comparing WCSS values for several k values to determine whether k = 3 provided an appropriate cluster structure.

2.6 Adaptive Learning Recommendation Mapping

After the final clusters were obtained, each cluster was interpreted based on its academic profile. The High Cluster was mapped to an enrichment pathway, the Medium Cluster to a regular learning pathway with periodic feedback, and the Low Cluster to a remedial and mentoring pathway. This mapping was designed to make the clustering results practically useful for teachers in planning learning interventions.

3. RESULT AND DISCUSSION

3.1 Dataset and Preprocessing Results

The dataset consisted of 50 student records containing assignment scores, UTS scores, UAS scores, and attendance rates. After data inspection, the complete and valid records were processed further. Each variable was normalized using the Min-Max Scaling method so that all attributes had equal contribution to the clustering process. This step was important because K-Means uses distance calculation, which is sensitive to differences in data scale.

3.2 K-Means Clustering Results

The clustering process was implemented using the Scikit-Learn library [4]. The number of clusters was set to k = 3 based on the research objective and supported by cluster evaluation. The final clustering result produced three student groups representing high, medium, and low academic performance categories. The summary of cluster variables is presented in Table 1.

Table 1. Cluster Variable Summary

Variable	Cluster 1 (High)	Cluster 2 (Medium)	Cluster 3 (Low)
Assignment Score	88.5	76.2	60.3
Midterm Examination Score (UTS)	90.1	74.8	58.6
Final Examination Score (UAS)	89.4	75.1	55.9
Attendance Rate (%)	97.5	88.2	72.4
Number of Students	12	28	10

Vol 10 No 2, July 2026, Page 125-129
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v10i2.9806

Table 1 shows that Cluster 1 has the highest average scores across all academic indicators, followed by Cluster 2 and Cluster 3. This pattern indicates that the clusters formed by K-Means are consistent with the intended academic performance categories.

Table 2. Cluster Distribution

Cluster	Performance Category	Number of Students	Percentage (%)
Cluster 1	High	12	24%
Cluster 2	Medium	28	56%
Cluster 3	Low	10	20%
Total	**-**	**50**	**100%**

As shown in Table 2, most students were classified into the Medium Cluster, consisting of 28 students or 56% of the dataset. The High Cluster consisted of 12 students or 24%, while the Low Cluster consisted of 10 students or 20%. These results show that a meaningful proportion of students require additional academic support.

3.3 Cluster Evaluation

Cluster evaluation was conducted using WCSS and the Silhouette Score. The Elbow Method indicated that k = 3 was appropriate because further increases in the number of clusters produced only marginal reductions in WCSS. The Silhouette Score of 0.6843 indicates that the clusters were reasonably well separated and internally consistent. The evaluation results are presented in Table 3.

Table 3. Clustering Evaluation Results

Evaluation Metric	Result
Number of Clusters (k)	3
Evaluation Method	Elbow Method and Silhouette Score
WCSS (k = 3)	2.3714
Silhouette Score (k = 3)	0.6843
Interpretation	Reasonably well-separated clusters

3.4 Analysis of Student Academic Characteristics

Cluster 1 represents students with high academic performance. Students in this cluster had strong achievement in assignments, UTS, UAS, and attendance. This indicates that they have stable academic performance and high learning participation. Therefore, these students are suitable for enrichment activities and more challenging learning tasks.

Cluster 2 is the largest group and represents students with medium academic performance. Students in this cluster showed relatively stable achievement but still require regular monitoring and feedback. They do not require intensive intervention, but continuous formative feedback is needed to prevent performance decline and support gradual improvement.

Cluster 3 represents students with low academic performance. The average scores and attendance rates in this cluster were lower than those of the other clusters. Students in this group require remedial learning, structured mentoring, and closer academic monitoring. Early identification of this group is important so that learning difficulties can be addressed before they become more serious.

3.5 Adaptive Learning Recommendations

The clustering results were converted into adaptive learning recommendations. Students in Cluster 1 are recommended to receive enrichment programs, such as project-based tasks, advanced exercises, and independent exploration activities. Students in Cluster 2 are recommended to receive regular instruction supported by periodic feedback and formative assessment. Students in Cluster 3 are recommended to receive remedial programs, simplified learning materials, step-by-step guidance, and intensive mentoring.

These recommendations show that clustering results can be used not only as statistical outputs but also as practical guidance for teachers. By using academic data, schools can identify student learning needs more objectively and allocate

ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v10i2.9806

learning support more effectively. This finding supports previous studies showing that K-Means Clustering can assist educational institutions in grouping students and supporting academic decision-making [6]–[12].

4. CONCLUSION

This study successfully implemented the K-Means Clustering algorithm to classify student academic performance based on assignment scores, UTS scores, UAS scores, and attendance rates. The clustering process produced three student groups: the High Cluster with 12 students (24%), the Medium Cluster with 28 students (56%), and the Low Cluster with 10 students (20%). The cluster evaluation results, with WCSS of 2.3714 and a Silhouette Score of 0.6843, indicate that the three-cluster structure is appropriate for representing student academic performance categories. The clustering results were then translated into adaptive learning recommendations: enrichment activities for the High Cluster, regular instruction with periodic feedback for the Medium Cluster, and remedial learning with structured mentoring for the Low Cluster. These findings demonstrate that K-Means Clustering can support data-driven educational decision-making and provide a practical foundation for adaptive learning recommendations in vocational education. However, this study is limited by the relatively small dataset and the use of only four academic variables. Future research should include larger datasets and additional variables, such as learning motivation, socioeconomic background, learning behavior, and digital literacy, to improve clustering accuracy and the generalizability of adaptive learning recommendations.

REFERENCES

[1] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2012.

[2] C. Romero and S. Ventura, “Educational data mining: A review of the state of the art,” IEEE Transactions on Systems, Man, and Cybernetics, Part C, vol. 40, no. 6, pp. 601–618, 2010, doi: 10.1109/TSMCC.2010.2053532.

[3] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297.

[4] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[5] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” Journal of Computational and Applied Mathematics, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[6] A. Akram, N. Risal, D. D. Andayani, M. I. Suherman, and A. B. 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. 5, no. 1, pp. 60–67, 2024, doi: 10.59562/jessi.v5i1.2178.

[7] J. Hutagalung, Y. H. Syahputra, and Z. P. Tanjung, “Pemetaan siswa kelas unggulan menggunakan algoritma K-Means Clustering,” JATISI (Jurnal Teknik Informatika dan Sistem Informasi), vol. 9, no. 1, pp. 606–620, 2022, doi: 10.35957/jatisi.v9i1.1516.

[8] E. A. Saputra and Y. Nataliani, “Analisis pengelompokan data nilai siswa untuk menentukan siswa berprestasi menggunakan metode clustering K-Means,” Journal of Information Systems and Informatics, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.

[9] C. Azzahra and Sriani, “Clustering of high school students academic scores using K-Means Algorithm,” Journal of Information Systems and Informatics, vol. 7, no. 1, pp. 572–586, 2025, doi: 10.51519/journalisi.v7i1.1029.

[10] S. J. S. Alawi and I. N. M. Shaharanee, “Clustering student performance data using K-Means algorithms,” Journal of Computing Innovation and Analytics, vol. 2, no. 1, pp. 41–55, 2023.

[11] Tundo, S. Raihanah, T. Wahyudi, and Sugiyono, “K-Means Clustering of Social Studies Performance at Junior High School,” IJID (International Journal on Informatics for Development), vol. 13, no. 2, pp. 460–472, 2024, doi: 10.14421/ijid.2024.4632.

[12] A. A. N. Risal, D. Maryani, N. Fadillah, and A. A. Risal, “Implementasi K-Means Clustering untuk rekomendasi kelas unggulan di SMK 1 Teknologi dan Rekayasa Mimika,” Journal of Embedded Systems, Security and Intelligent Systems, vol. 5, no. 3, pp. 255–261, 2024.

[13] A. Paramythis and S. Loidl-Reisinger, “Adaptive learning environments and e-learning standards,” Electronic Journal of e-Learning, vol. 2, no. 1, pp. 181–194, 2004.