Artificial Intelligence in Educational Decision Sciences
E-ISSN: 3123-8823
Vol. 1, No. 2, August 2026

Clustering Multi-Indicator Learning Outcomes of Vocational High School Students: A Comparison of K-Means and DBSCAN

Muhammad Fikri Aqil✉
Universitas Negeri Makassar, Makassar, Indonesia
Irwansyah Suwahyu
UIN Sunan Kalijaga Yogyakarta, Yogyakarta, Indonesia

ABSTRACT

Purpose – This study aims to compare the performance of K-Means and DBSCAN algorithms in clustering vocational high school students' learning outcomes in the Network Administration subject to support data-driven educational decision making.

Methods – A quantitative experimental approach was employed using secondary academic data from vocational students. The variables analyzed included final examination scores, midterm examination scores, assignments, attendance, attitudes, and learning activities. Clustering was conducted using K-Means and DBSCAN algorithms implemented through data analysis software. Cluster quality and separation were evaluated using silhouette coefficients to assess the effectiveness of each algorithm in grouping student learning outcomes.

Findings – The results show that K-Means produces relatively stable and interpretable clusters when student performance data exhibit more uniform distributions. In contrast, DBSCAN demonstrates stronger capability in handling noisy data and identifying students with extreme performance levels as outliers. Both algorithms successfully reveal meaningful patterns in student learning outcomes, but differ in their sensitivity to data distribution and noise.

Research limitations – This study is limited to a single vocational subject and one institutional context, which may restrict the generalizability of the findings to other vocational domains.

Originality – This study provides empirical evidence on the comparative performance of partition-based and density-based clustering algorithms using multi-indicator learning outcome data in vocational education.

ARTICLE HISTORY
Received: 01-05-2026
Revised: 20-07-2026
Accepted: 02-08-2026

KEYWORDS
DBSCAN;
K-Means algorithm;
Learning outcomes.
Student clustering;
Vocational education;

Correspondence Author: ✉fikriaqil@mail.com

To cite this article : Muhammad Fikri Aqil, Irwansyah Suwahyu (2026). Clustering Multi-Indicator Learning Outcomes of Vocational High School Students: A Comparison of K-Means and DBSCAN. Artificial Intelligence in Educational Decision Sciences, 1(2), 1-10.

INTRODUCTION

The rapid development of information and communication technology has significantly increased the use of data across various sectors, including education, fundamentally transforming how educational institutions operate and manage information. Advances in digital systems have enabled schools and other educational organizations to systematically collect, store, and manage large volumes of data related to academic performance, learning activities, and student behavior in an increasingly integrated manner (Baig et al., 2020; Gros & García-Peñalvo, 2023). The widespread implementation of learning management systems, digital assessment platforms, and electronic academic records has contributed to the emergence of data-rich educational environments in which data are no longer used solely for administrative reporting but also as a strategic resource for improving educational quality and effectiveness (Aldowah et al., 2020; Daniel, 2020; Fenglei Ma et al., 2025; Picciano, 2022).

This transformation has strengthened the reliance on data analysis methods to support data-driven decision making at institutional, instructional, and policy levels. In the context of education in Indonesia, the utilization of data has become an essential component of efforts to enhance learning quality, accountability, and transparency. As the volume of available data continues to grow, particularly data related to student learning outcomes, academic achievement, and learning behaviors, there is an increasing need for analytical techniques capable of identifying meaningful patterns, trends, and relationships that can generate actionable insights to support educational improvement and informed decision making (Dutt et al., 2023; Luqman Ibrahim & Mohammed, 2022). One analytical approach that has gained considerable attention in response to these needs is data mining, which enables the extraction of useful and previously unknown information from large, complex, and heterogeneous datasets. Through data mining, hidden structures, associations, and patterns that are not readily observable through conventional statistical analysis can be uncovered and interpreted, making it particularly suitable for analyzing complex educational data (Narang et al., 2024; Romero & Ventura, 2023). In educational contexts, data mining techniques have been widely applied to analyze multiple dimensions of student learning, including academic achievement, learning behavior, and engagement, with clustering algorithms playing a crucial role in grouping students based on similarities across academic and non-academic characteristics to support a deeper understanding of student diversity and learning patterns and to inform more targeted instructional strategies (Utomo, 2023).

The application of clustering techniques is particularly relevant in Vocational High Schools (Sekolah Menengah Kejuruan or SMK) in Indonesia, where the goals, structure, and learning processes of education differ substantially from those of general secondary education. Vocational education is designed to equip students with practical skills, technical competencies, and professional attitudes that are aligned with labor market demands, emphasizing the integration of theoretical knowledge with hands-on practical application. However, SMKs face substantial challenges in improving learning quality, including heterogeneous student backgrounds, variations in prior academic ability, differences in motivation and learning styles, and the complexity of aligning instructional content with industry requirements (Irnanda et al., 2025; Safii et al., 2021). These conditions make it difficult to apply uniform instructional approaches that can effectively address the diverse needs of all students. In this context, the effective use of student data becomes increasingly important as a basis for understanding learning conditions and supporting instructional improvement. The utilization of data mining methods in SMKs is therefore viewed as a strategic approach to supporting evidence-based decision making and improving learning quality in vocational education, as it allows educators and school administrators to analyze student data systematically and objectively to identify patterns that may not be visible through traditional evaluation methods (Salman et al., 2025). Vocational schools generate a wide range of student-related data that can be leveraged for analytical purposes, including midterm and final examination scores, assignment results, attendance records, classroom participation, attitudes, and levels of learning activity. When analyzed collectively, these academic and non-academic variables provide a more comprehensive representation of student performance that extends beyond academic outcomes alone and supports the identification of meaningful learning patterns and student profiles (Divayana & Adiarta, 2024; Oyelade et al., 2020; Shahiri et al., 2022).

Numerous studies have demonstrated the potential of clustering algorithms in educational data analysis by showing their ability to group students based on similarities across multiple criteria and to reveal patterns that are not easily identified through traditional analytical approaches. Clustering algorithms such as K-Means and DBSCAN have been shown to be effective in grouping students based on both academic and non-academic variables, thereby supporting educational decision making by providing data-driven insights into student performance and learning behavior (Hasibuan et al., 2024). Despite this demonstrated potential, the application of clustering techniques in vocational education remains relatively limited, particularly in studies that focus on specific vocational subjects such as Network Administration. In educational data mining, various clustering algorithms are available, each with distinct assumptions and analytical characteristics. K-Means is widely used due to its simplicity and computational efficiency but requires the number of clusters to be specified in advance and is sensitive to outliers, whereas DBSCAN applies a density-based approach that enables the identification of irregular cluster shapes and the detection of noise in datasets with uneven distributions (Rahman et al., 2025). Previous studies examining the application of these algorithms in educational contexts have often focused on general datasets, such as school dropout data, or lower levels of education, and have frequently relied on relatively small or homogeneous datasets, limiting their applicability to vocational education contexts that involve more complex and multidimensional data structures (Enjelika et al., 2025; Utomo, 2023).

The application of K-Means and DBSCAN to SMK student data is therefore especially important due to the multidimensional and complex nature of vocational education variables, which include academic performance indicators, behavioral attributes, attitudes, and levels of student activity that differ in scale and distribution. These characteristics increase the complexity of clustering analysis and highlight the need for a careful evaluation of algorithm performance when applied to vocational school data. Understanding how different clustering algorithms perform under these conditions is essential for selecting appropriate analytical methods and for ensuring that clustering results accurately reflect student learning patterns. Consequently, a systematic comparison of K-Means and DBSCAN in the context of SMK student data, particularly within subject-specific learning environments such as Network Administration, is necessary to address existing research gaps and to support the development of more personalized, effective, and data-informed learning strategies in vocational education.

METHOD

This study aims to compare the performance of two clustering algorithms, namely K-Means and DBSCAN, in grouping the learning outcomes of vocational high school students in the subject of Network Administration based on several variables, such as final exam scores, midterm exam scores, assignments, attitude, and activity. This study uses a quantitative approach with an experimental research design. An experimental design was chosen because it allows researchers to evaluate and compare the clustering results produced by both algorithms based on data collected from students. In addition, this study aims to measure the effectiveness of each algorithm in grouping students based on their learning outcomes, as well as to compare the accuracy and stability of the clustering results.

The data used in this study is secondary data obtained from the academic records of vocational high school students in the Network Administration subject. The data included final exam scores, midterm exam scores, assignments, attitude, and activity given by teachers during one semester. The data collection method used was documentation, where the necessary data was taken directly from the school's academic archives containing records of student exam scores and evaluations. All of the data was then analyzed to determine the grouping patterns that could be found using the two selected algorithms.

The population in this study was all vocational high school students enrolled in the Network Administration course during that semester. The sample used in this study was 100 students taken from several classes using stratified sampling techniques to ensure that each class category (e.g., classes with high, medium, and low grades) was represented in the sample. The stratified sampling technique was used to minimize bias in sample selection and to ensure that different class characteristics were reflected in the sample taken. This sample selection was important so that the research results could be generalized to a wider population.

The instruments used in this study were data analysis software, namely Orange for the implementation of the K-Means and DBSCAN algorithms. In addition, for data validation and reliability, descriptive statistical tests, such as mean and standard deviation, were used to examine the distribution of student learning outcomes data before the clustering process. Furthermore, reliability testing was conducted using a stability test technique, which involved comparing the clustering results obtained in different experiments with the same data. This was important to ensure that the model built could produce consistent and reliable results.

The data analysis techniques used for this study are clustering analysis using the K-Means and DBSCAN algorithms. The K-Means algorithm will be used to group student data based on the variables mentioned above, namely final exam scores, midterm exam scores, assignments, attitude, and activity. The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm will be used as a comparison because this algorithm is capable of identifying clusters with uneven density and can handle data with noise or outliers. The results of these two algorithms will be evaluated using the silhouette score metric, which measures how well each data point is in a cluster that suits it.

To test the hypothesis, a comparison was made between the clustering results produced by K-Means and DBSCAN in terms of stability, accuracy, and ability to identify patterns in the data. The testing was conducted using inferential statistical tests, such as the t-test, to compare the average cluster values between the two algorithms. The significance level used was 0.05, which means that the statistical test results were considered significant if the p-value obtained was less than 0.05. The software used for data analysis was Orange.

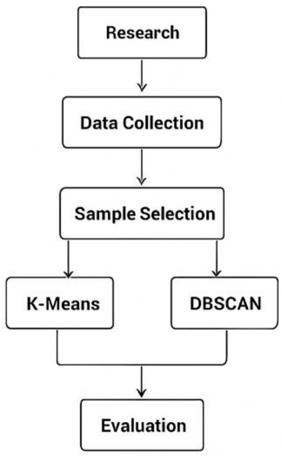

Figure 1. Research Framework

RESULTS AND DISCUSSION

The K-Means clustering analysis divided 172 students into two main clusters based on attendance, attitude, activity, assignments, midterm exams, and final exams, with grouping primarily determined by total score consistency. Students in Cluster C1 included those categorized as “Low,” such as Faiza Hudriyah with a total score of 485 and a silhouette score of 0.57342, indicating relatively focused clustering of lower-performing students, while other students such as Febry Yuni Usman in the same cluster achieved higher total scores of 540 and were classified as "Medium," reflecting a more balanced performance range between 505 and 550. These results indicate that K-Means effectively grouped students according to relatively uniform performance levels, although some clusters remained broad and heterogeneous due to variations in data density, a characteristic commonly observed when K-Means is applied to datasets with mixed distributions (Xu & Tian, 2015). In contrast, the DBSCAN algorithm demonstrated a different clustering behavior by leveraging data density to identify groups and isolate noise, making it more suitable for educational data with high variability. For instance, Muhammad Fikri R with a score of 497 was identified as an outlier, illustrating DBSCAN's ability to separate students with performance patterns that deviate substantially from the majority and to handle noise and outliers more effectively than K-Means in diverse educational contexts.

Table 1. K-Means

STUDENT NAME	Cluster	Silhouette	GROUPING	ATTENDANCE	ATTITUDE	ACTIVITY	TASK	Uts	Uas	Total
Faiza hudriyah	C1	0.57342	Low	90	70	80	80	80	85	485
Ferdy pradana arwanto	C1	0.592928	Low	80	80	80	80	80	85	485
Ghina putri anjani	C1	0.543808	Low	90	20	80	90	100	85	465
M rasyah	C1	0.557295	Low	70	20	100	100	100	85	475
Muh. Helmi	C1	0.584584	Low	80	70	80	80	80	85	475
Cluster 2 (sedang)										
Febry yuni usman	C2	0.529231	Medium	90	60	100	100	100	90	540
Imbran gunawan	C2	0.564721	Medium	100	80	80	80	80	85	505
Ita purnama sari harahap	C2	0.565401	Medium	70	100	100	100	80	85	535
Kharin nur hikmah	C2	0.579622	Medium	90	70	80	100	80	85	505
M arif rahmat	C2	0.598091	Medium	90	90	90	80	80	85	515
Cluster 3 (tinggi)										
M syahrul	C3	0.591753	Medium	70	100	100	80	80	90	520

M. Aqsan qadri	C3	0.565401	Medium	70	100	100	100	80	85	535
M. Yusri	C3	0.594199	Medium	80	100	80	80	80	85	505
Muh raditya adha	C3	0.560747	Medium	70	100	100	100	80	90	540
Muh ridwan	C3	0.502971	Medium	95	95	92	85	81	100	548

DBSCAN produces more dense clusters with fewer outliers, and these clusters tend to be more homogeneous than those produced by K-Means. In a study by Cahapin, Malabag, Santiago, Reyes, Legaspi, and Adrelas (2023), it was found that DBSCAN is very effective in detecting more complex patterns in educational data, especially for students with extreme scores. In their study, DBSCAN helped map student data that might have been overlooked by other clustering methods, and similar findings were observed in this study, where students with significantly different scores could be clearly identified. However, DBSCAN also has challenges in terms of determining the right parameters, which can affect the final results of the clustering (Cahapin et al., 2023).

Table 2. DBSCAN

STUDENT NAME	Cluster	Core	GROUPING	ATTENDANCE	ATTITUDE	ACTIVITY	TASK	Uts	Uas	Total
Andi fachri keyza aqiela	C1	1	Medium	100	80	100	100	85	58	523
Muhammad asrul	C1	1	Medium	100	100	60	100	85	84	529
Achmad yani mansyur	C1	1	Medium	100	100	80	100	85	84	549
Adinda	C1	1	Medium	100	100	80	100	85	84	549
Afnan naufal	C1	1	Medium	100	100	80	100	85	72	537
Cluster 2 (sedang)										
Agung wicaksono	C2	1	Medium	100	100	80	100	85	76	541
Ahmaed farizhi c1	C2	1	Medium	100	100	80	100	85	84	521
Achmad wicaksono	C2	1	Medium	100	100	90	100	85	84	545
Muhammad razak c1	C2	1	Medium	100	100	85	100	85	90	570
Karya kardimov	C2	1	Medium	100	100	75	100	85	70	530
Cluster 3 (tinggi)										
Achmad yani mansyur	C3	1	Medium	100	100	80	100	85	84	545
Cinta dwi kurnia	C3	1	Medium	100	100	85	100	85	90	590
Bima dwi santoso	C3	1	Medium	100	100	80	100	85	90	550

Rama ananda tirta	C3	1	Medium	100	100	75	100	85	80	540
Muhammad fadhly	C3	1	Medium	100	100	80	100	85	70	530

From the comparison between K-Means and DBSCAN, it can be concluded that both algorithms have their own strengths and weaknesses. K-Means is superior in terms of grouping students based on the uniformity of their scores, and can provide more stable results when the data distribution is relatively even. However, K-Means tends to have difficulty handling data that contains a lot of noise, which is a major problem in many groupings in educational data. Meanwhile, DBSCAN excels at identifying isolated students or those with very high or low performance, but tends to produce smaller clusters with more noise or unclustered data (Hooshyar et al., 2023).
The analysis shows that although both methods produce similar results in terms of cluster division based on academic performance, DBSCAN has an advantage in handling data with more noise and outliers. Students with very high or low performance, who may not be grouped well in K-Means, are clearly identified by DBSCAN as outliers. These results are consistent with the findings by DeFreitas and Bernard (2015), which show that DBSCAN is more effective in situations where data has an uneven distribution or contains many anomalies (DeFreitas & Bernard, 2015).
Related research on the use of K-Means and DBSCAN in educational data clustering shows results consistent with these findings. Omer, Mohammed, Awadallah, Abrar, and Shah (2022) explain that K-Means is very useful in clustering data that has a uniform and clear distribution, but can lose detail when the data contains noise or anomalies. Conversely, Nafuri, Zani, Zainudin, Rahman, and Aliff (2022) found that DBSCAN can cluster data better when dealing with students with extreme scores or uneven distributions, such as in the context of educational data that is full of performance variations (Fawzia Omer et al., 2022; Mohamed Nafuri et al., 2022).

CONCLUSION

This study demonstrates that clustering techniques can be effectively applied to analyze student learning outcomes in Vocational High Schools (SMK) by incorporating both academic and non-academic variables, including attendance, attitudes, activity levels, assignments, midterm exams, and final exams. The findings indicate that both K-Means and DBSCAN are capable of grouping students based on learning performance patterns; however, each algorithm exhibits distinct characteristics that influence clustering outcomes. K-Means performs well in datasets with relatively uniform distributions by producing stable and interpretable clusters that reflect general performance levels, making it suitable for identifying broad categories of student achievement. Nevertheless, its sensitivity to noise and outliers limits its effectiveness when applied to heterogeneous educational data.
In contrast, DBSCAN demonstrates a stronger ability to handle noisy data and to identify students with extreme performance levels by treating them as outliers rather than forcing them into predefined clusters. This capability allows DBSCAN to provide a more nuanced representation of student performance diversity, which is particularly valuable in vocational education contexts characterized by varied student backgrounds and learning behaviors. Overall, the results suggest that the choice of clustering algorithm should be aligned with the characteristics of the dataset and the analytical objectives. For SMK learning data that are multidimensional and heterogeneous, DBSCAN offers advantages in capturing irregular patterns, while K-Means remains useful for generating general performance groupings. These findings contribute to the application of educational data mining in vocational education and support the development of more data-informed and targeted instructional strategies.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Vocational High School (Sekolah Menengah Kejuruan, SMK) that provided access to student learning data used in this study.

Appreciation is also extended to the teachers and school administrators who supported the data collection process and provided valuable contextual information. Furthermore, the authors acknowledge the support and encouragement from colleagues and institutions that contributed indirectly to the completion of this research.

AUTHOR CONTRIBUTION STATEMENT

MFA conceived the study and research design. MFA and IS conducted data preprocessing and clustering analysis. MFA performed model comparison and validation. MFA drafted the manuscript. All authors reviewed, revised, and approved the final version and agreed to be accountable for all aspects of the work.

AI DISCLOSURE STATEMENT

The authors declare that artificial intelligence tools were used solely to assist with language refinement and clarity during manuscript preparation. AI tools did not influence the research design, data processing, analysis, interpretation, or conclusions. Full responsibility for the content remains with the authors.

REFERENCES

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2020). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, *37*, 13–49. https://doi.org/10.1016/j.tele.2019.01.007

Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2020). Big data in education: A state of the art, limitations, and future research directions. *International Journal of Educational Technology in Higher Education*, *17*(1), 44. https://doi.org/10.1186/s41239-020-00223-0

Cahapin, E. L., Malabag, B. A., Santiago Jr., C. S., Reyes, J. L., Legaspi, G. S., & Adrales, K. L. (2023). Clustering of students admission data using k-means, hierarchical, and DBSCAN algorithms. *Bulletin of Electrical Engineering and Informatics*, *12*(6), 3647–3656. https://doi.org/10.11591/eei.v12i6.4849

Daniel, B. (2020). B ig D ata and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, *46*(5), 904–920. https://doi.org/10.1111/bjet.12230

DeFreitas, K., & Bernard, M. (2015). Comparative performance analysis of clustering techniques in educational data mining. *Educational Data Mining*, *10*, 65–67.

Divayana, D. G. H., & Adiarta, A. (2024). Development of cse-ucla evaluation model modified by using weighted product in order to optimize digital library services in higher education of computer in bali. *Jurnal Pendidikan Vokasi*, *7*(3).

Dutt, A., Ismail, M. A., & Herawan, T. (2023). A Systematic Review on Educational Data Mining. *IEEE Access*, *5*, 15991–16005. https://doi.org/10.1109/ACCESS.2017.2654247

Enjelika, D., Hariani, N. P., & Sutabri, T. (2025). *Perbandingan kinerja algoritma k-means dan dbscan dalam pengelompokan data nilai kelas viii a pada SMPN 01 PALEMBANG*. *5*(1).

Fawzia Omer, A., Mohammed, H. A., Awadallah, M. A., Khan, Z., Abrar, S. U., & Shah, M. D. (2022). Big Data Mining Using K-Means and DBSCAN Clustering Techniques. In M. Ouaissa, Z. Boulouard, M. Ouaissa, I. U. Khan, & M. Kaosar (Eds.), *Big Data Analytics and Computational Intelligence for Cybersecurity* (Vol. 111, pp. 231–246). Springer International Publishing. https://doi.org/10.1007/978-3-031-05752-6_15

Fenglei Ma, Weifeng Liu, & Tao Ma. (2025). Automatic differentiation application on stochastic finite element. *2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering*, 5609726. https://doi.org/10.1109/CMCE.2010.5609726

Gros, B., & García-Peñalvo, F. J. (2023). Future Trends in the Design Strategies and Technological Affordances of E-learning. In J. M. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, Design, and Technology* (pp. 345–367). Springer International Publishing. https://doi.org/10.1007/978-3-319-17461-7_67

Hasibuan, M. S., Lubis, A. H., & Sari, M. N. (2024). Perbandingan algoritma clustering DBSCAN dan K-Means dalam pengelompokan siswa terbaik. *INFOTECH : Jurnal Informatika & Teknologi*, *5*(2), 301–309. https://doi.org/10.37373/infotech.v5i2.1457

Hooshyar, D., Yang, Y., Pedaste, M., & Huang, Y.-M. (2023). Clustering Algorithms in an Educational Context: An Automatic Comparative Approach. *IEEE Access*, *8*, 146994–147014. https://doi.org/10.1109/ACCESS.2020.3014948

Irnanda, K. F., Windarto, A. P., Hartama, D., & Wanto, A. (2025). ANALISA METODE DATA MINING PADA PENGELOMPOKAN LAPANGAN KERJA INFORMAL SEKTOR NON-PERTANIAN. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, *3*(1). https://doi.org/10.30865/komik.v3i1.1673

Luqman Ibrahim, A., & Mohammed, M. G. (2022). A new three-term conjugate gradient method for training neural networks with global convergence. *Indonesian Journal of Electrical Engineering and Computer Science*, *28*(1), 551. https://doi.org/10.11591/ijeecs.v28.i1.pp551-558

Mohamed Nafuri, A. F., Sani, N. S., Zainudin, N. F. A., Rahman, A. H. A., & Aliff, M. (2022). Clustering Analysis for Classifying Student Academic Performance in Higher Education. *Applied Sciences*, *12*(19), 9467. https://doi.org/10.3390/app12199467

Narang, H., Wu, F., & Ogunniyan, A. (2024). Numerical Solutions of Heat and Mass Transfer with the First Kind Boundary and Initial Conditions in Capillary Porous Cylinder Using Programmable Graphics Hardware. *International Journal of Advanced Computer Science and Applications*, *7*(6). https://doi.org/10.14569/IJACSA.2016.070607

Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. (2020). *Application of k Means Clustering algorithm for prediction of Students Academic Performance*. https://doi.org/10.48550/ARXIV.1002.2425

Picciano, A. G. (2022). The Evolution of Big Data and Learning Analytics in American Higher Education. *Online Learning*, *16*(3). https://doi.org/10.24059/olj.v16i3.267

Rahman, G. A., Hasbi, I., Tenriawaru, A., Surimi, L., & Alfiyan, A. N. (2025). Penggunaan Algoritma Dbscan Dalam Pengelompokan Kabupaten/Kota Di Sulawesi Tenggara Berdasarkan Indikator Pendidikan. *Simtek : Jurnal Sistem Informasi Dan Teknik Komputer*, *10*(1), 184–193. https://doi.org/10.51876/simtek.v10i1.1546

Romero, C., & Ventura, S. (2023). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Safii, M., Wahyudi, M., Solikhun, Zarlis, M., & Effendi, S. (2021). Food Self-sufficiency Decision Support Model Based on Provinces in Indonesia Using the Clustering Method. *Journal of Physics: Conference Series*, *1255*(1), 012068. https://doi.org/10.1088/1742-6596/1255/1/012068

Salman, M. D., Rahmaddeni, R., Pratama, N. R., A, M. N. F., Setiawan, A. A., Zalianti, F., & Huda, I. B. (2025). Perbandingan Kinerja Algoritma Clustering K-Means dan K-Medoids dalam Pengelompokan Sekolah di Provinsi Riau Berdasarkan Ketersediaan Sarana dan Prasarana. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *5*(3), 797–806. https://doi.org/10.57152/malcom.v5i3.1950

Shahiri, A. M., Husain, W., & Rashid, N. A. (2022). A Review on Predicting Student’s Performance Using Data Mining Techniques. *Procedia Computer Science*, *72*, 414–422. https://doi.org/10.1016/j.procs.2015.12.157

Utomo, M. H. T. (2023). *Clustering Data Siswa Putus Sekolah Dengan Algoritma K-Means Dan DBSCAN*. *18*(2).

Xu, D., & Tian, Y. (2015). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, *2*(2), 165–193. https://doi.org/10.1007/s40745-015-0040-1