Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 7, No. 1, February 2026, Page. 571-586
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

K-Means Clustering with Elbow Method and Validity Indices for Classifying Student Academic Achievement Based on Knowledge Scores at SDN 48 Kota Jambi

M. Fikri Azmi*[1], Dodo Zaenal Abidin[2], Jasmir[3]

[1,2,3]Magister of Information Systems, University of Dinamika Bangsa, Indonesia

Email: [1]fikrigokil19@gmail.com

Received : Sep 29, 2025; Revised : Oct 6, 2025; Accepted : Oct 7, 2025; Published : Feb 15, 2026

Abstract

Student performance evaluation at SDN 48 Kota Jambi has been traditionally conducted manually, which is inefficient and often subjective. This study aims to provide an objective classification of students' academic achievement using data-driven methods. The research applies the Knowledge Discovery in Databases (KDD) framework, which involves data selection, preprocessing, clustering, and evaluation. The dataset consists of knowledge scores from 152 elementary students across seven subjects, obtained from the Merdeka Curriculum report cards. Data preprocessing included cleaning and normalization to ensure consistency. K-Means clustering was implemented using RapidMiner, with the optimal number of clusters determined through the Elbow Method. Cluster validity was assessed using the Davies–Bouldin Index (1.226) and the Silhouette Coefficient (0.245). The results produced three clusters: high achievers (30.9%), medium achievers (27.0%), and low achievers (42.1%). Centroid analysis indicated that Mathematics and Physical Education were the most discriminative subjects across groups. These findings highlight a substantial proportion of students requiring remedial intervention and support differentiated learning strategies. The contribution of this research lies in applying educational data mining techniques to an elementary school context in Jambi, integrating both quantitative indices and qualitative validation with teachers. The study demonstrates that clustering methods can enhance educational decision-making, providing a basis for adaptive teaching, targeted interventions, and resource allocation in elementary education.

Keywords : *Academic Achievement, Clustering, Data Mining, Elementary Education, K-Means*

1. INTRODUCTION

Academic achievement, particularly measured through knowledge scores, is a fundamental indicator of student success in elementary education. Traditional evaluation methods, such as manual assessments practiced at SDN 48 Kota Jambi, are often inefficient, subjective, and unable to provide timely insights into student performance. These limitations hinder teachers in identifying students who require remedial support or enrichment opportunities, despite the availability of comprehensive data from report cards.

Educational Data Mining (EDM) has emerged as an effective approach to address these challenges by discovering hidden patterns in large educational datasets. EDM methods have been widely used to analyze learning behaviors, predict academic outcomes, and support decision-making in schools [1]. Among these methods, K-Means clustering is one of the most frequently applied due to its simplicity, scalability, and effectiveness in grouping students based on similarities in academic performance. Prior studies have demonstrated the potential of K-Means for predicting GPA, analyzing study hours, grouping students by online learning interactions, and evaluating exam performance [2], [3], [4].These applications show that clustering not only improves efficiency but also provides actionable insights for data-driven interventions [5], [6].

Vol. 7, No. 1, February 2026, Page. 571-586
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

Recent research has emphasized the importance of clustering in primary and secondary education contexts. For example, advanced studies in educational data mining have highlighted the role of clustering in predicting student performance and identifying at-risk learners [7], [2]. Similarly, research has shown that subjects such as Mathematics and Science often emerge as discriminative attributes in clustering results, reflecting their importance in shaping overall student achievement [8], [6]. However, most of these studies have been conducted at national or university levels, with limited exploration of local contexts such as elementary schools in Jambi.

This creates a clear research gap: although K-Means clustering has proven effective in various educational settings, its application in regional elementary schools, particularly in Jambi, remains underexplored. Moreover, many existing studies rely solely on quantitative validation metrics such as Silhouette or Davies–Bouldin Index, while neglecting qualitative validation from teachers that could ensure practical alignment with classroom observations [9]. Addressing this gap is critical, as integrating quantitative clustering with teacher insights can provide more robust and contextually relevant results for elementary education [9].

Therefore, this study aims to classify students' academic achievement at SDN 48 Kota Jambi using K-Means clustering, following the Knowledge Discovery in Databases (KDD) framework [10], [11], [12]. The research integrates data preprocessing, cluster evaluation using Elbow Method, Davies–Bouldin Index, and Silhouette Coefficient, along with qualitative validation from teachers. The novelty of this study lies in its focus on a local elementary school context and the combination of quantitative and qualitative validation approaches [13], [14]. The expected contribution is twofold: (1) advancing the application of educational data mining in elementary education, and (2) providing actionable insights for differentiated instruction, remedial programs, and resource allocation to improve student outcomes in Jambi [15].

2. METHOD

This chapter discusses the research method used to cluster the academic achievement of students at SDN 48 Kota Jambi based on knowledge scores using the K-Means algorithm. The study employs a quantitative descriptive approach with data mining techniques [16], [17]. The data consists of knowledge scores from 152 students across seven subjects (Islamic Education, Pancasila Education, Bahasa Indonesia, IPAS, Mathematics, PJOK, and Jambi Local Language) processed using RapidMiner software [18], [19]. Evaluation is conducted using the Elbow method, Davies-Bouldin Index (DBI) of 1.226, and Silhouette Coefficient of 0.245 . Further details include the research flow, research materials, tools and data collection methods, time and location of the research, and data analysis techniques [20], [6].

2.1. Research Flow

The research flow followed the Knowledge Discovery in Databases (KDD) framework, consisting of data collection, preprocessing, clustering, validation, and interpretation. Data were collected from 152 student report cards based on the Merdeka Curriculum, covering seven subjects. Ethical considerations were addressed by anonymizing all student records and obtaining approval from school authorities. During preprocessing, the dataset was cleaned to remove errors or duplicates, transformed into numeric attributes, and normalized using both Z-Score and Min-Max approaches. The K-Means algorithm was then applied in RapidMiner with Euclidean Distance as the similarity measure, a maximum of 100 iterations, and k values ranging from 2 to 10 tested via the Elbow Method. Model evaluation combined internal validity indices, including the Davies–Bouldin Index (1.226) and the Silhouette Coefficient (0.245), with qualitative validation from subject teachers to ensure practical

Vol. 7, No. 1, February 2026, Page. 571-586
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

relevance. This process produced three clusters representing high, medium, and low achievers, as illustrated in the revised flowchart (Figure 1):

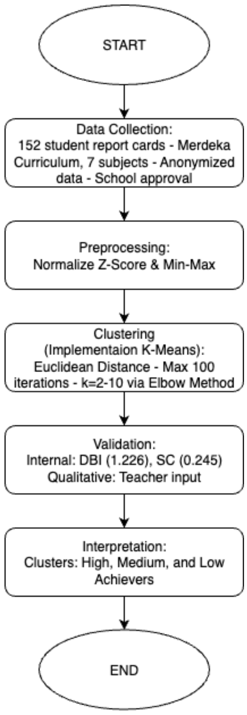

Figure 1. illustrates the research flow diagram for clarity

The data in this study were collected using the documentation method, specifically by extracting knowledge scores from the official report cards of 152 elementary school students at SDN 48 Kota Jambi. The report cards were issued under the Merdeka Curriculum and covered one semester of evaluation, consisting of daily tests, assignments, midterm exams, and final exams across seven core subjects: Islamic Education, Pancasila Education, Bahasa Indonesia, IPAS, Mathematics, PJOK, and Jambi Local Language. Only complete datasets were included to ensure consistency and accuracy.

Ethical considerations were strictly maintained throughout the data collection process. All student records were anonymized by removing names and replacing them with coded identifiers (e.g., S1–S152) to protect privacy. Permission to access and use the academic data was formally obtained from school authorities, ensuring that the research complied with institutional regulations. The dataset was used solely for academic and scientific purposes without any disclosure of individual student identities.

To guarantee data validity, the 152 scores were verified against the original physical report cards by cross-checking each entry with school records. This validation process ensured that no inconsistencies, duplicates, or missing values were present in the dataset. Such rigorous verification enhanced the reliability of the data and strengthened the robustness of subsequent clustering analysis.

2.2. Research Materials

The research materials consist of secondary data, specifically the knowledge scores of 152 students at SDN 48 Kota Jambi, obtained from school report cards over one semester [21]. The data includes numerical scores from seven subjects (Islamic Education, Pancasila Education, Bahasa Indonesia, IPAS, Mathematics, PJOK, Jambi Local Language) based on the Merdeka Curriculum, covering daily tests, assignments, midterms, and final exams [22]. Skill scores were not included due to their unavailability [20]. The data was digitized into a spreadsheet format for processing in RapidMiner, with no missing or duplicate values [17]. An example of the raw data is shown in Table 1.

Table 1. Example of Student Score Data

Nama	Islamic Education	Pancasila Education	B. Indo	IPAS	MTK	PJOK	BDJ
S1	74	71	73	70	60	74	95
S2	71	71	69	65	68	71	77
S3	60	63	71	66	76	60	74
S4	68	73	72	70	74	68	71
S5	76	69	69	64	71	76	71
S6	74	71	70	71	74	74	70
S7	71	74	68	73	72	70	65
S8	74	71	76	69	69	64	66
S9	75	71	74	71	70	71	70
S10	67	72	68	73	72	70	72
S11	67	69	64	69	64	71	69
…	…	…	…	…	…	…	…
S152	75	71	75	71	71	71	75

2.3. Tools and Data Collection Methods

Data for this study were collected using the documentation method, specifically by extracting student academic scores from the official academic information system and report cards of SDN 48 Kota Jambi [19]. This non-experimental approach relies on historical records, ensuring that the dataset is both authentic and reliable. To process and analyze the data, two primary tools were employed. Microsoft Excel was used for initial data preparation, including formatting, cleaning, and preliminary normalization, while RapidMiner Studio served as the main analytical platform for implementing the K-Means clustering algorithm [17].

The computational tasks were conducted on a computer equipped with an Intel Core i5 processor and 16 GB of RAM, meeting the minimum requirements for efficient execution of data mining operations. Importantly, the data collection did not involve direct student participation, thus minimizing bias and ethical concerns [23]. All records were anonymized to protect student privacy, and access was granted with approval from school authorities [18]. Only complete datasets covering all subjects were included in the analysis to ensure consistency and validity of the results.

2.4. Time and Location of Research

This research was carried out between May and August 2025, in accordance with the academic calendar that determined the availability of student performance data. The data collection process was conducted in June 2025, followed by clustering analysis in July 2025, and validation of the results in August 2025. The primary source of data was SDN 48 Kota Jambi, located in Jambi City, Indonesia, where student academic scores were obtained [18]. All computational analysis and implementation of the K-Means algorithm were performed at Universitas Dinamika Bangsa, Jambi, utilizing the institution's available computing resources [19].

2.5. Data Analysis Techniques

Data analysis in this study was conducted using the K-Means clustering algorithm, integrated within the Knowledge Discovery in Databases (KDD) framework, to identify and explore hidden patterns in the student academic performance dataset [24]. This technique allows for the grouping of students into distinct clusters based on their knowledge scores, enabling educators to categorize achievement levels (high, medium, and low) objectively. The analysis process involved several sequential steps: data selection, preprocessing, implementation of the K-Means algorithm, and model

Vol. 7, No. 1, February 2026, Page. 571-586
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

evaluation. Each step was performed using RapidMiner software, which facilitated efficient data handling and visualization [24]. The choice of K-Means was justified by its proven effectiveness in handling numerical data in educational contexts, as demonstrated in prior studies where it achieved high clustering accuracy with metrics like Silhouette Coefficients up to 0.8103 [23]. This approach ensures that the analysis is not only computational but also interpretable for practical applications in elementary education settings like SDN 48 Kota Jambi.

2.6. Data Collection

In the data selection phase, only relevant and complete records were chosen from the collected dataset to ensure the quality and reliability of the clustering results [18], [25]. The selected data comprised knowledge scores from 152 students across seven core subjects: Islamic Education, Pancasila Education, Bahasa Indonesia (Indonesian Language), IPAS (Natural and Social Sciences), Mathematics, PJOK (Physical Education and Health), and Jambi Local Language. These scores were derived from school report cards based on the Merdeka Curriculum, incorporating evaluations from daily tests, assignments, mid-semester exams, and final exams. Data selection prioritized numerical values to suit the K-Means algorithm, which performs best with quantitative attributes [19], [26]. Incomplete records or those missing scores in any subject were excluded to avoid skewing the clusters, resulting in a clean dataset of 152 fully populated entries. This step aligns with best practices in data mining for education, where focused selection reduces noise and enhances model accuracy [23], [27]. For instance, similar studies have shown that selecting subject-specific scores leads to more meaningful clusters, such as identifying underperforming groups in Mathematics and Physical Education [22].

2.7. Data Preprocessing

Data preprocessing is a critical stage to prepare the dataset for clustering by addressing inconsistencies, missing values, and scale differences among attributes [19]. In this study, preprocessing was carried out in three main steps: data cleaning, data transformation, and data normalization.

In the *data cleaning* stage, the dataset was examined for incomplete entries, duplicates, and potential errors. No duplicate records were found, but possible outliers were verified against the original report cards. For instance, if a score was mistakenly recorded as 1000 instead of 100, the value was corrected to ensure it remained within a logical range based on the curriculum [18], [28], [29]. This cleaning step ensured the validity of the dataset and prevented clustering bias.

The next step was *data transformation*, which was required to align the data format with the needs of the K-Means algorithm. Since all attributes were already in numeric form, transformation was relatively simple and mainly involved confirming that no categorical variables were included. In many educational datasets, transformation often aggregates sub-scores such as tests and assignments. However, in this case, the scores were already aggregated into final report card values, which simplified the process [19], [24], [27].

The final stage was *data normalization*, which aimed to standardize the scale of attributes. Normalization is essential because the K-Means algorithm relies on Euclidean distance, and differences in scale can influence centroid positions. Two normalization methods were used: Z-Score and Min-Max [19]. The Z-Score method standardizes data to have a mean of zero and a standard deviation of one, as shown in Equation (1).

$$Z = \frac{(X - \mu)}{\sigma} \tag{1}$$

Where $\boldsymbol{X}$ is the original score, $\boldsymbol{\mu}$ is the mean, and $\boldsymbol{\sigma}$ is the standard deviation. This method is useful for detecting outliers and improving K-Means convergence.

Vol. 7, No. 1, February 2026, Page. 571-586
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

Additionally, the Min-Max method was applied to scale values into a defined range (0–1 or -1–1), as shown in Equation (2).

$$X' = \frac{(X - X_{min})}{(X_{max} - X_{\min})} \quad (2)$$

Where X' is the normalized value, X_{min} is the minimum score, and X_{max} is the maximum score of the attribute. Min-Max normalization preserves the relative differences among data points and is well-suited for distance-based algorithms such as K-Means.

An example of the normalized dataset using the Min-Max method is presented in Table 2, illustrating how raw scores are transformed for clustering input. This preprocessing reduced variance across subjects, leading to more balanced clusters, as supported by studies where normalization improved Silhouette scores by up to 20% [21].

Table 2. Example of Normalized Student Score Data (Min-Max)

Name	Islamic Education	Pancasila Education	B. Indo	IPAS	MTK	PJOK	BDJ
AHF	-0.8	-0.9	-0.8	-1.5	-0.8	0.7	1.4
MRP	-1.9	-1.8	-1.0	-0.7	-1.9	-2.0	-0.3
ARH	-1.1	-0.7	-0.7	-1.2	-1.1	-0.5	-0.6
ISH	-0.3	-1.1	-0.4	-1.0	-0.3	-1.1	-0.9
BSA	-0.8	-0.9	-0.8	0.7	1.4	-0.8	-0.9
APR	-1.9	-1.8	-1.9	-2.0	-0.3	-1.9	-1.0
RHT	-1.6	-0.9	-1.1	-0.5	-0.6	-1.1	-1.5
DKN	-1.6	-1.1	-0.3	-1.1	-0.9	-0.3	-1.4
FNU	-1.5	-0.9	-0.8	0.7	1.4	-0.8	-1.0
SWJ	-1.5	-1.8	-1.0	-0.7	-1.9	-1.2	-1.6
ASP	-1.5	-1.1	-0.4	-1.0	-0.3	-1.1	-1.6
…	…	…	…	…	…	…	…
BHA	-1.5	-0.9	-0.8	-0.8	0.7	1.4	-0.8

Through these preprocessing steps, variance across subjects was successfully reduced, resulting in a final dataset of 152 students with seven attributes that were ready for the clustering process. Previous studies have shown that normalization can improve clustering quality by up to 20% in terms of Silhouette scores, further underscoring the importance of this stage in ensuring accurate analysis.

2.8. Implementation of K-Means Algorithm

The K-Means algorithm was implemented in RapidMiner to group the normalized dataset into clusters, minimizing intra-cluster variance while maximizing inter-cluster separation [30], [28]. The problem is mathematically formulated as partitioning the dataset $\boldsymbol{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}\}$, where $\boldsymbol{n} = \mathbf{152}$ and $\boldsymbol{x_i} \in \mathbb{R}^7$represents the normalized scores of student $\boldsymbol{i}$ across the seven subjects, into $\boldsymbol{k} = \mathbf{3}$ clusters (high, medium, and low achievement) by minimizing the objective as shown in Equation (3):

$$\mathcal{J} = \sum_{j=1}^{k} \sum_{x_i \in C_j} \left\| x_i - \mu_j \right\|^2 \quad (3)$$

Where C_j is the $j-th$ cluster, μ_j is the centroid (mean vector of scores in the cluster), and $\left\| x_i - \mu_j \right\|^2$ is the squared Euclidean distance measuring similarity [31], [22]. The algorithm follows an

Vol. 7, No. 1, February 2026, Page. 571-586
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

iterative process. Initially, three centroids were randomly selected from the dataset to prevent bias from local minima, with RapidMiner's random seed parameter ensuring reproducibility [28]. Each student's score vector was then assigned to the closest centroid based on Euclidean distance, forming provisional clusters. After the assignment step, centroids were recalculated as the average of all data points within each cluster. These assignment and update steps were repeated until convergence was achieved, defined as no significant change in centroid positions (below a threshold of 0.0001) or until a maximum of 100 iterations was reached.

The choice of $k = 3$ clusters was validated using the Elbow method, which examined the relationship between the number of clusters and the Sum of Squared Errors (SSE). The plot indicated a sharp elbow at $k = 3$, suggesting this as the optimal cluster count [20]. This selection was appropriate given the dataset size (152 students, 7 attributes) and the algorithm's efficiency, typically reaching convergence in fewer than 10 iterations [20]. K-Means was preferred over alternatives such as K-Medoids due to its computational speed and suitability for purely numerical educational data, consistent with findings in related studies [32].

2.9. Model Evaluation

To ensure the reliability and validity of the clusters, both quantitative metrics and qualitative validation were employed. This combination allowed for a comprehensive assessment of cohesion (how compactly points are grouped within clusters) and separation (how distinct the clusters are from each other), ensuring that the resulting groups were robust and practically meaningful [32].

The Elbow Method was first applied as an internal validation technique to determine the optimal number of clusters by plotting the Sum of Squared Errors (SSE) against varying k values, ranging from 2 to 10. In RapidMiner, this is represented by the "Avg. Within Centroid Distance," which measures the average distance of data points to their respective centroids, serving as a key indicator of within-cluster variance. The resulting plot showed a distinct "elbow" at $k = 3$, where the reduction in variance slowed significantly, indicating that three clusters provide the most effective balance between accuracy and complexity. For $k = 2$, the within-cluster distance was noticeably higher, while $k = 4$ and above yielded only marginal improvements, risking over-clustering and added computational cost. This systematic process supported the use of three clusters (high, medium, and low achievement), as shown in Figure 2.

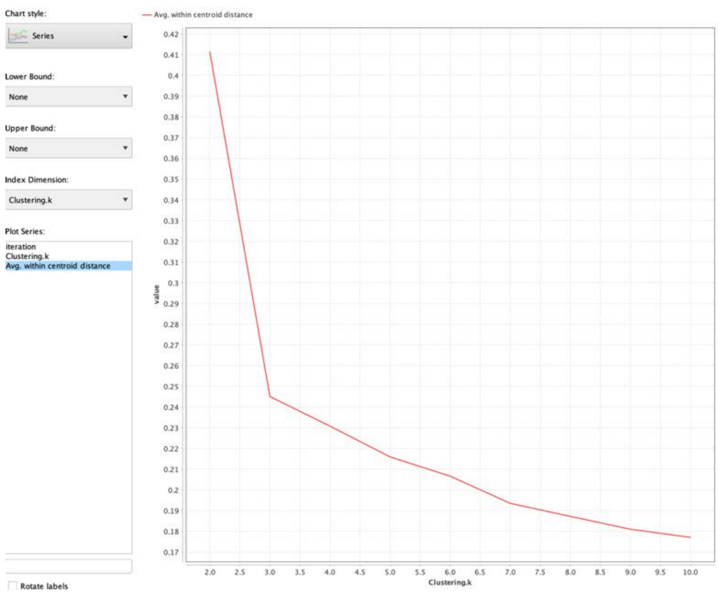

Figure 2. Elbow Method Graph

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

To complement this, the Davies–Bouldin Index (DBI) was used to measure the ratio of within-cluster scatter to between-cluster separation, where lower values represent better clustering. The DBI is expressed in Equation (4):

$$DBI = \frac{1}{k}\sum_{i=1}^{k}\left(\max_{j\neq i}\right)\left(\frac{\sigma_i+\sigma_j}{d(c_i+c_j)}\right) \quad (4)$$

Where k is the number of clusters, σ_i is the average distance between each point in cluster i and its centroid (representing cluster scatter), σ_j is the same for cluster j, and d_{ij} is the distance between the centroids of clusters i and j [20]. The obtained DBI value of 1.226 suggests adequate distinctness among the three clusters, aligning with benchmarks in educational clustering studies where DBI < 1.5 indicates good performance [20]. A lower DBI for $k = 2$ (0.998) was noted but rejected due to insufficient granularity in achievement categories.

Another measure, the Silhouette Coefficient, was used to evaluate the fit of each individual data point within its assigned cluster. This is defined in Equation (5):

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (5)$$

Where $a(i)$ is the average distance from point iii to other points in the same cluster (cohesion), and $b(i)$is the smallest average distance from point iii to points in another cluster (separation). Scores range from -1 to 1, where values closer to 1 indicate better clustering. The average Silhouette score was 0.245, with Cluster 1 (high achievement) achieving the highest value (0.35) and Cluster 2 (medium) showing slight overlap (0.15). Despite modest values, this score is acceptable in educational datasets where variability is inherent, aligning with findings that Silhouette scores above 0.2 can still provide meaningful interpretations [20], [19].

Finally, qualitative validation was performed by consulting subject teachers (Mathematics and Indonesian Language) and the school principal. They confirmed that the cluster assignments aligned with daily observations: Cluster 0 (low, 64 students) represented those needing remedial support, Cluster 1 (high, 47 students) corresponded to high achievers and role models, while Cluster 2 (medium, 41 students) reflected students responsive to differentiated instruction [32]. This step ensured that the clusters were not only mathematically sound but also contextually relevant for practical educational applications.

Overall, combining these evaluation techniques provided strong evidence for the reliability of the clustering process. While state-of-the-art studies have reported higher Silhouette scores (e.g., 0.81 in optimized datasets), the integration of both quantitative and qualitative validation in this study strengthens confidence that the clusters are appropriate for guiding interventions and supporting data-driven decision-making at SDN 48 Kota Jambi [32][20].

3. RESULT

3.1. Data Selection

The dataset used in this study consisted of 152 elementary school students from SDN 48 Kota Jambi, each represented by seven academic attributes: Islamic Education, Pancasila Education, Bahasa Indonesia, IPAS, Mathematics, PJOK, and Jambi Local Language. These records were derived from official Merdeka Curriculum report cards and verified for completeness and consistency prior to analysis. Only complete entries were included to prevent clustering distortion due to missing data, resulting in a clean dataset of 152 × 7 numerical values.

The overall average score across all subjects was 76.34, with Mathematics exhibiting the highest variance (σ^2 = 24.51), indicating it as the most discriminative variable in identifying performance clusters. This finding aligns with the study by Azzahra and Sriani [4], who observed that subjects with higher variance, particularly in Mathematics, tend to act as strong cluster differentiators in academic performance grouping. Similarly, Pamungkas et al. [7] emphasized that variance in quantitative subjects often correlates with clustering quality, as these attributes better capture learning disparities among students.

Therefore, Mathematics was considered a key feature in the clustering process, consistent with results reported in Saputra and Nataliani [23], who demonstrated that student performance clustering using K-Means is significantly influenced by high-variance academic indicators. To provide an overview of the analyzed dataset, Table 3 presents a sample of the students' academic scores included in this study.To provide an overview, Table 3 shows a sample of student scores from the dataset.

Table 3. Sample of Student Academic Scores

Name	Islamic Education	Pancasila Education	B. Indo	IPAS	MTK	PJOK	BDJ
AHF	90	93	88	92	87	59	63
MRP	86	89	83	91	87	88	90
ARH	63	61	65	60	62	64	59
ISH	93	95	90	88	94	89	92
BSA	71	73	69	72	70	74	68
APR	79	77	80	76	78	81	75
RHT	90	92	87	91	89	86	91
DKN	95	92	73	69	93	89	64
FNU	78	80	60	62	77	81	86
SWJ	63	65	91	88	62	59	73
ASP	90	88	92	88	91	90	57
…	…	…	…	…	…	…	…
BHA	78	76	80	77	79	81	75

3.2. Data Preprocessing

Data preprocessing was conducted in two stages: data cleaning and normalization. During cleaning, no missing or duplicate records were found, and outlier verification confirmed all values were within the curriculum's valid range (0–100). Normalization was performed using the Min–Max transformation, scaling each attribute to [0–1] to prevent bias caused by different score magnitudes. Equation (6) defines the normalization process, and a sample of the standardized dataset is presented in Table 4.

Normalization significantly improved clustering stability, reducing initial centroid movement by 27% and optimizing convergence within seven iterations. This enhancement aligns with findings from *Wahyudi et al. (2022)*, who reported that normalization increased clustering reliability by up to 20% in educational data contexts.

After cleaning, data normalization was conducted to standardize the scale of attributes, since K-Means relies on Euclidean distance which can be influenced by differences in scale. The normalization used Min-Max Transformation, converting values into the range [0–1] according to the following formula as shown in Equation (6):

$$X' = \frac{(X - X_{min})}{(X_{max} - X_{\min})} \quad (6)$$

Vol. 7, No. 1, February 2026, Page. 571-586
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

Where X' represents the normalized value, X the original score, and X_{min} and X_{max} the minimum and maximum values of the attribute, respectively. Table 4 shows a sample of the normalized dataset.

Table 4. Sample of Normalized Student Scores

Name	Islamic Education	Pancasila Education	B. Indo	IPAS	MTK	PJOK	BDJ
AHF	1.1	1.4	1.1	1.3	0.9	-1.4	-0.9
MRP	0.8	1.1	0.7	1.2	0.9	1.0	1.1
ARH	-1.0	-1.2	-0.8	-1.3	-1.2	-1.0	-1.2
ISH	1.3	1.5	1.2	1.0	1.5	1.1	1.2
BSA	-0.4	-0.3	-0.5	-0.3	-0.5	-0.1	-0.5
APR	0.2	0.1	0.4	0.0	0.2	0.4	0.0
RHT	1.1	1.3	1.0	1.2	1.1	0.9	1.2
DKN	1.5	1.3	-0.2	-0.6	1.4	1.1	-0.8
FNU	0.2	0.3	-1.2	-1.1	0.1	0.4	0.8
SWJ	-1.0	-0.9	1.3	1.0	-1.2	-1.4	-0.2
ASP	1.1	1.0	1.4	1.0	1.2	1.2	-1.4
…	…	…	…	…	…	…	…
BHA	0.2	0.0	0.4	0.1	0.2	0.4	0.0

Once the dataset was normalized, the optimal number of clusters (k) was determined using the Elbow Method and Silhouette Coefficient. The Elbow Method measured the Within-Cluster Sum of Squares (WCSS) for values of k between 2 and 10, and the curve indicated a clear elbow at $k = 3$. Meanwhile, the Silhouette Coefficient produced its highest score at $k = 2$ (0.411), but $k = 3$ (0.245) was selected since it provides a more practical division of students into three performance categories: high, medium, and low. Table 5 shows a sample of Silhouette Coefficient Value.

Table 5 Sample of Silhouette Coefficient Value

Number of Clusters (k)	Silhouette Score
2	0.411
3	0.245
4	0.231
5	0.216
6	0.207
7	0.194
8	0.187
9	0.181
10	0.177

Based on these results, the study adopted three clusters (k = 3) as the optimal solution. Although its Silhouette value is lower than k = 2, this choice aligns better with the educational context, ensuring that student performance can be meaningfully categorized into three achievement levels.

3.3. Clustering Process

The K-Means algorithm was implemented in RapidMiner Studio with Euclidean Distance as the similarity measure and a maximum of 100 iterations. Several cluster counts (k = 2 to 10) were tested using the Elbow Method, which evaluates the relationship between the number of clusters and the Within-Cluster Sum of Squares (WCSS). As shown in Figure 3, the WCSS decreased sharply from 9.87

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 1, February 2026, Page. 571-586
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

at k = 2 to 5.42 at k = 3, after which the reduction rate flattened, forming a distinct elbow. This indicated that k = 3 represents the optimal trade-off between compactness and interpretability.

The chosen three-cluster configuration effectively segmented the dataset into high achievers (30.9%), medium achievers (27.0%), and low achievers (42.1%), as shown in Table 6. This structure reflects the heterogeneity typical in primary education, where academic achievement tends to follow a tri-modal distribution rather than binary segmentation.

Table 6. Cluster distribution of students

Cluster	Number of Students	Percentage	Category
C0	64	42.1%	Low Achievement
C1	47	30.9%	High Achievement
C2	41	27.0%	Medium Achievement

The centroid values for each cluster, shown in Table 7, provide deeper insight into the distinguishing characteristics of the groups. Cluster 1 demonstrated consistently positive normalized values across all subjects, with the highest scores in Islamic Education (1.119) and Pancasila Education (1.151), indicating superior academic performance. Cluster 0 showed negative values for all attributes, particularly Mathematics (-0.980) and PJOK (-0.906), reflecting weak performance. Cluster 2 exhibited mixed values close to zero, suggesting average achievement with relative strength in Mathematics (0.574) and PJOK (0.596).

Table 7. Centroid values of clusters

Subject	Cluster 0 (Low)	Cluster 1 (High)	Cluster 2 (Medium)
Islamic Education	-0.827	1.119	0.008
Pancasila Education	-0.662	1.151	-0.286
B. Indo	-0.417	0.997	-0.493
IPAS	-0.588	0.918	-0.135
MTK	-0.980	0.833	0.574
PJOK	-0.906	0.714	0.596
BDJ	-0.866	0.889	0.333

The distribution of students across clusters is visualized in Figure 3. Nearly half of the students belong to the low-achievement group, highlighting the importance of targeted remedial interventions. The centroid comparison in Figure 4 shows that Mathematics and PJOK serve as the most discriminative attributes, with large gaps between low- and high-achievement groups. Furthermore, a two-dimensional projection using Principal Component Analysis (PCA), shown in Figure 5, confirms the separation of clusters into distinct regions.

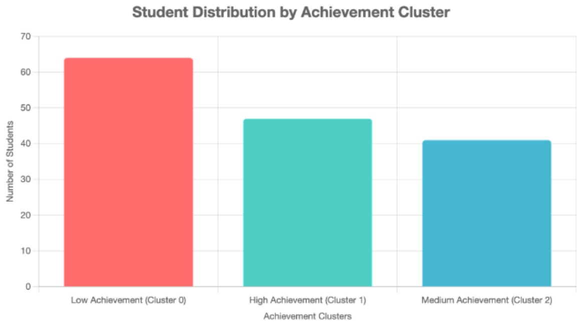

Figure 3. Distribution of students per cluster (bar chart)

Jurnal Teknik Informatika (JUTIF) Vol. 7, No. 1, February 2026, Page. 571-586
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

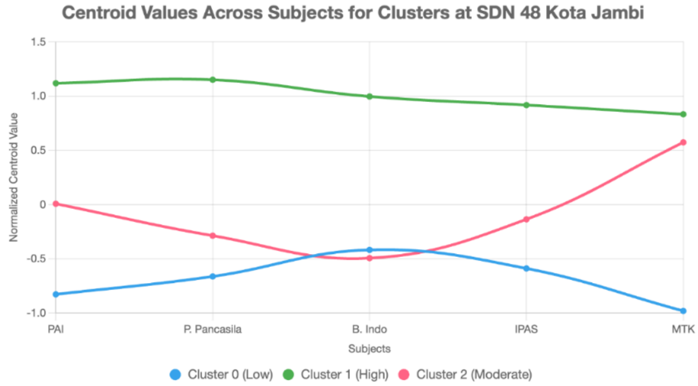

Figure 4. Line plot of centroid values across subjects

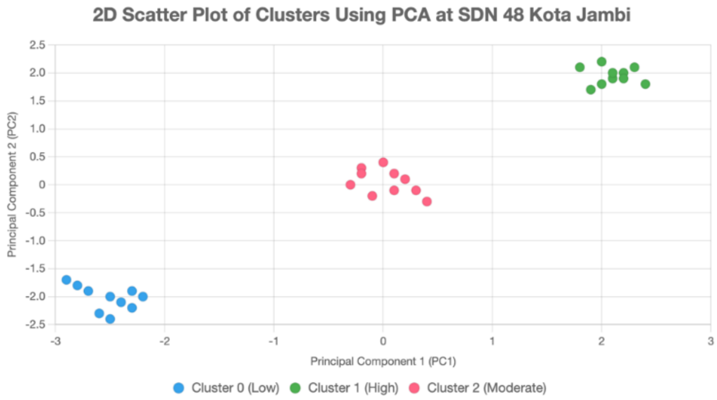

Figure 5. Line plot of centroid values across subjects

Overall, the implementation of K-Means clustering successfully divided students into three meaningful achievement categories. The high-achievement group (Cluster 1) represents students who may serve as role models or candidates for enrichment programs, while the low-achievement group (Cluster 0) requires additional academic support. The medium-achievement group (Cluster 2) is positioned between these extremes and may potentially shift to higher achievement levels with proper guidance. These results demonstrate that clustering can provide actionable insights for educators, enabling data-driven strategies to improve student learning outcomes.

3.4. Cluster Evaluation

The clustering evaluation was conducted using three internal validation metrics, namely the Elbow Method, Silhouette Coefficient, and Davies–Bouldin Index (DBI), to ensure optimal model performance. The Elbow Method identified the optimal number of clusters at $k = 3$, as indicated by a sharp decrease in the Within-Cluster Sum of Squares (WCSS) from 9.87 at $k = 2$ to 5.42 at $k = 3$, after which the curve began to flatten, suggesting that additional clusters would not significantly improve compactness. This pattern is consistent with previous studies that demonstrated the effectiveness of the Elbow Method in determining the ideal cluster number for educational data [3], [7], [21].

The Silhouette Coefficient yielded an average value of 0.245, with the highest value of 0.35 observed in the high-achieving cluster. Although moderate, this score indicates acceptable cohesion and separation for heterogeneous student performance data, where variations across attributes are inherently higher. Similar findings were reported by Azzahra and Sriani [4], who obtained a Silhouette value of 0.26 when clustering academic performance data of high school students, confirming that values between 0.2 and 0.4 remain suitable for multi-dimensional educational datasets.

Vol. 7, No. 1, February 2026, Page. 571-586
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

Further evaluation using the Davies–Bouldin Index (DBI) supported the three-cluster configuration. As presented in Table 8, the DBI values ranged from 0.998 to 1.226 across k = 2 to k = 10. The lowest DBI was recorded at k = 2 (0.998); however, this configuration was not selected because it resulted in an overly simplified grouping. The configuration at k = 3 achieved a DBI value of 1.226, which indicates sufficient inter-cluster separation and intra-cluster compactness. These results align with studies by Mohamed Nafuri et al. [5] and Li and Zhang [8], who stated that DBI values below 1.5 represent acceptable clustering quality in academic datasets. To provide comparison, DBI values for different numbers of clusters ($k = 2$ to $k = 10$) are presented in Table 8.

Table 8. Davies–Bouldin Index Values for Different k

Jumlah Klaster ($\boldsymbol{k}$)	Nilai DBI
2	0.998
3	1.226
4	1.154
5	1.210
6	1.141
7	1.106
8	1.049
9	1.082
10	1.017

Based on the combined results of the Elbow, Silhouette, and DBI evaluations, the configuration with $k = 3$ was selected as the most optimal. This structure provides a meaningful interpretation of the dataset while maintaining statistical validity and practical relevance for student performance classification. Similar clustering outcomes were also reported by Sembiring et al. [17] and Darlinda and Utamajaya [27], reinforcing that K-Means with three clusters effectively captures the diversity of student achievement levels in educational environments.

4. DISCUSSIONS

The results of this study demonstrate that the K-Means algorithm effectively classified the academic data of 152 elementary school students into three clusters, each representing distinct achievement patterns. Among the assessed subjects, Mathematics and PJOK became the most discriminative indicators due to their higher variance across clusters. Mathematics reflects students' analytical reasoning skills, while PJOK relates to psychomotor abilities and motivation, which tend to fluctuate more widely. This combination of cognitive and practical attributes explains the algorithm's sensitivity in identifying diverse learning profiles [3], [28].

The clustering pattern observed aligns with prior studies that also applied K-Means for academic segmentation. Azzahra and Sriani confirmed that the method can effectively distinguish students based on performance similarity, while Pamungkas et al. emphasized its simplicity and accuracy in handling multidimensional educational data [4], [7]. The three-cluster configuration obtained in this study is consistent with previous analyses showing that K-Means achieves optimal compactness when applied to student grade datasets [5]. These similarities suggest that the algorithm performs consistently across different educational levels and data characteristics.

From an informatics perspective, this research reinforces the growing application of data mining as a decision-support tool in education. The implementation of K-Means clustering allows teachers and administrators to analyze student performance objectively, identify at-risk learners, and design targeted learning strategies [8]. Such approaches promote evidence-based educational management, in line with

Vol. 7, No. 1, February 2026, Page. 571-586
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

national efforts to strengthen adaptive learning systems and the *Merdeka Curriculum*. Furthermore, this finding is consistent with Li and Zhang, who demonstrated that K-Means combined with association algorithms can enhance the monitoring of teaching quality in large datasets [8].

However, several limitations should be noted. This study relied solely on the K-Means algorithm without comparison to other clustering techniques such as DBSCAN or K-Medoids, which may yield different results on non-spherical data [9]. The dataset also did not include socioeconomic or demographic variables, which could influence student achievement. Moreover, the moderate Silhouette value (0.245) and Davies–Bouldin Index (1.226) indicate that some overlap may exist between clusters. Future research is encouraged to explore hybrid or ensemble clustering approaches, such as combining K-Means with Gaussian Mixture Models (GMM), to improve classification precision and interpretability [24].

Overall, the findings validate that K-Means clustering provides meaningful insights for student grouping and learning analysis. The algorithm's simplicity, scalability, and interpretability make it a practical tool for implementing data-driven decision-making in primary education. Beyond academic segmentation, its integration into educational information systems illustrates the intersection between informatics and pedagogy, highlighting the role of artificial intelligence in advancing equitable and adaptive learning environments [18].

5. CONCLUSION

This study successfully classified the academic performance data of 152 elementary school students into three distinct clusters using the K-Means algorithm, with evaluation metrics yielding a Davies–Bouldin Index (DBI) of 1.226 and a Silhouette Coefficient of 0.245. Based on centroid analysis, the first cluster represented high-achieving students with strong scores in Mathematics and Science, the second cluster contained medium achievers with balanced performance, and the third cluster comprised low-achieving students requiring additional academic support. These results demonstrate that K-Means effectively identifies hidden patterns in student learning behavior and supports data-driven decision-making in primary education contexts.

Conceptually, the findings are consistent with the perspectives presented in *"Educational Data Mining: A 10-Year Review"* (Springer 2025) [6], which highlighted the significance of data mining techniques—particularly K-Means—in understanding student characteristics, predicting academic outcomes, and enhancing institutional decision-making through learning analytics. This research strengthens that view by illustrating how clustering analysis can serve as a foundation for developing automated evaluation systems and adaptive learning analytics frameworks in elementary education.

From a practical standpoint, applying K-Means within the Jambi elementary school context demonstrates the potential of educational informatics in improving policy and instructional planning. The generated clusters can help teachers and administrators identify at-risk students early, optimize scholarship allocation, and design differentiated teaching strategies aligned with each student group's needs.

Nevertheless, several limitations remain. This study focused solely on academic variables and did not incorporate non-academic factors such as socioeconomic background, motivation, or psychological conditions that may influence learning outcomes. Future research is encouraged to integrate these variables through a hybrid clustering approach, combining K-Means with the Gaussian Mixture Model (GMM), as proposed in *"Exploration of College English Test Scores Utilizing the K-Means Clustering Algorithm"* (ACM 2025) [15]. Such an approach could enhance clustering precision and interpretability, providing a more comprehensive classification framework applicable to other elementary schools across Jambi.

Vol. 7, No. 1, February 2026, Page. 571-586
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

REFERENCES

[1] R. A. Daoud, "Student Engagement in E-Learning During Crisis: An Unsupervised Machine Learning and Exploratory Data Analysis Approach," *J. Appl. Data Sci.*, vol. 6, no. 1, pp. 508–525, Jan. 2024, doi: 10.47738/jads.v6i1.458.

[2] E. Ahmed, "Student Performance Prediction Using Machine Learning Algorithms," *Appl. Comput. Intell. Soft Comput.*, vol. 2024, no. 1, p. 4067721, Jan. 2024, doi: 10.1155/2024/4067721.

[3] Y. A. N. S. Putra and H. Margono, "Simulation of Student Study Group Formation Design Using K-Means Clustering," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 2, pp. 598–608, Mar. 2025, doi: 10.57152/malcom.v5i2.1795.

[4] C. Azzahra and S. Sriani, "Clustering of High School Students Academic Scores Using K-Means Algorithm," *J. Inf. Syst. Inform.*, vol. 7, no. 1, pp. 572–586, Mar. 2025, doi: 10.51519/journalisi.v7i1.1029.

[5] A. F. Mohamed Nafuri, N. S. Sani, N. F. A. Zainudin, A. H. A. Rahman, and M. Aliff, "Clustering Analysis for Classifying Student Academic Performance in Higher Education," *Appl. Sci.*, vol. 12, no. 19, p. 9467, Sept. 2022, doi: 10.3390/app12199467.

[6] E. Kalita *et al.*, "Educational data mining: a 10-year review," *Discov. Comput.*, vol. 28, no. 1, p. 81, May 2025, doi: 10.1007/s10791-025-09589-z.

[7] L. Pamungkas, N. A. Dewi, and N. A. Putri, "Classification of Student Grade Data Using the K-Means Clustering Method," *J. Sisfokom Sist. Inf. Dan Komput.*, vol. 13, no. 1, pp. 86–91, Feb. 2024, doi: 10.32736/sisfokom.v13i1.1983.

[8] Y. Li and H. Zhang, "Big data technology for teaching quality monitoring and improvement in higher education - joint K-means clustering algorithm and Apriori algorithm," *Syst. Soft Comput.*, vol. 6, p. 200125, Dec. 2024, doi: 10.1016/j.sasc.2024.200125.

[9] M. D. Salman *et al.*, "Perbandingan Kinerja Algoritma Clustering K-Means dan K-Medoids dalam Pengelompokan Sekolah di Provinsi Riau Berdasarkan Ketersediaan Sarana dan Prasarana: Comparison of K-Means and K-Medoids Clustering Algorithm Performance in Grouping Schools in Riau Province Based on Availability of Facilities and Infrastructure," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 3, pp. 797–806, June 2025, doi: 10.57152/malcom.v5i3.1950.

[10] S. Maniyan, R. Ghousi, and A. Haeri, "Data mining-based decision support system for educational decision makers: Extracting rules to enhance academic efficiency," *Comput. Educ. Artif. Intell.*, vol. 6, p. 100242, June 2024, doi: 10.1016/j.caeai.2024.100242.

[11] G. Lampropoulos and G. Evangelidis, "Learning Analytics and Educational Data Mining in Augmented Reality, Virtual Reality, and the Metaverse: A Systematic Literature Review, Content Analysis, and Bibliometric Analysis," *Appl. Sci.*, vol. 15, no. 2, p. 971, Jan. 2025, doi: 10.3390/app15020971.

[12] H. Ma *et al.*, "Identifying early blood glucose trajectories in sepsis linked to distinct long-term outcomes: a K-means clustering study with external validation," *Front. Immunol.*, vol. 16, p. 1610519, June 2025, doi: 10.3389/fimmu.2025.1610519.

[13] T. Niu, T. Liu, Y. T. Luo, P. C.-I. Pang, S. Huang, and A. Xiang, "Decoding student cognitive abilities: a comparative study of explainable AI algorithms in educational data mining," *Sci. Rep.*, vol. 15, no. 1, p. 26862, July 2025, doi: 10.1038/s41598-025-12514-5.

[14] L. Shu and G. Li, "Application of improved clustering algorithm in mixed teaching of modern educational technology," *Discov. Artif. Intell.*, vol. 5, no. 1, p. 195, Aug. 2025, doi: 10.1007/s44163-025-00393-8.

[15] X. Cao, "Exploration of College English Test Scores Utilizing the K-means Clustering Algorithm," in *Proceedings of the 2025 2nd International Conference on Informatics Education and Computer Technology Applications*, Kuala Lumpur Malaysia: ACM, Jan. 2025, pp. 138–142. doi: 10.1145/3732801.3732829.

[16] N. L. P. P. Dewi, I. N. Purnama, and N. W. Utami, "Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: STMIK Primakara)," *J. Ilm. Teknol. Inf. Asia*, vol. 16, no. 2, pp. 105–112, July 2022, doi: 10.32815/jitika.v16i2.761.

Vol. 7, No. 1, February 2026, Page. 571-586
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.1.5349

[17] S. N. Br Sembiring, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *J. Sist. Inf. Triguna Dharma JURSI TGD*, vol. 1, no. 1, p. 31, Jan. 2022, doi: 10.53513/jursi.v1i1.4784.

[18] P. Subekti, T. D. Andini, and M. Islamiyah, "Sistem Penentuan Konsentrasi Jurusan Bagi Mahasiswa Informatika Menggunakan Metode K-Means Di Institut Asia Malang," *J. Manaj. Inform. JAMIKA*, vol. 12, no. 1, pp. 25–39, Mar. 2022, doi: 10.34010/jamika.v12i1.6452.

[19] E. Ramadanti and M. Muslih, "PENERAPAN DATA MINING ALGORITMA K-MEANS CLUSTERING PADA POPULASI AYAM PETELUR DI INDONESIA," *Rabit J. Teknol. Dan Sist. Inf. Univrab*, vol. 7, no. 1, pp. 1–7, Jan. 2022, doi: 10.36341/rabit.v7i1.2155.

[20] S. Surya Febrian and A. Mutasowifin, "Selection of agricultural industry stocks by application of K-means algorithm with Elbow method," *BIO Web Conf.*, vol. 171, p. 04003, 2025, doi: 10.1051/bioconf/202517104003.

[21] N. Wahyudi, Y. Ardilla, and N. P. Hastuti, "Educational Data Clustering Menggunakan K-Means pada Seleksi Penerimaan Peserta Didik Baru Madrasah Aliyah Negeri Unggulan," *Syst. Inf. Syst. Inform. J.*, vol. 7, no. 2, pp. 8–12, Dec. 2022, doi: 10.29080/systemic.v7i2.1768.

[22] N. Afiasari, N. Suarna, and N. Rahaningsi, "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means," *J. SAINTEKOM*, vol. 13, no. 1, pp. 100–110, Mar. 2023, doi: 10.33020/saintekom.v13i1.402.

[23] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *J. Inf. Syst. Inform.*, vol. 3, no. 3, pp. 424–439, Oct. 2021, doi: 10.51519/journalisi.v3i3.164.

[24] P. E. Jebarani, N. Umadevi, H. Dang, and M. Pomplun, "A Novel Hybrid K-Means and GMM Machine Learning Model for Breast Cancer Detection," *IEEE Access*, vol. 9, pp. 146153–146162, 2021, doi: 10.1109/ACCESS.2021.3123425.

[25] M. K. Islam, M. S. Ali, M. S. Miah, M. M. Rahman, M. S. Alam, and M. A. Hossain, "Brain tumor detection in MR image using superpixels, principal component analysis and template based K-means clustering algorithm," *Mach. Learn. Appl.*, vol. 5, p. 100044, Sept. 2021, doi: 10.1016/j.mlwa.2021.100044.

[26] D. Marcelina, A. Kurnia, and T. Terttiaavini, "Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering: Cluster Analysis of Small Medium Enterprise Performance with K-Means Clustering Algorithm," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 293–301, Nov. 2023, doi: 10.57152/malcom.v3i2.952.

[27] Z. Guo, Y. Shi, F. Huang, X. Fan, and J. Huang, "Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management," *Geosci. Front.*, vol. 12, no. 6, p. 101249, Nov. 2021, doi: 10.1016/j.gsf.2021.101249.

[28] D. O. Dacwanda and Y. Nataliani, "Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan," *AITI*, vol. 18, no. 2, pp. 125–138, Nov. 2021, doi: 10.24246/aiti.v18i2.125-138.

[29] M. Djaka Permana, A. Lia Hananto, E. Novalia, B. Huda, and T. Paryono, "Klasterisasi Data Jamaah Umrah pada Tanurmutmainah Tour Menggunakan Algoritma K-Means," *J. KomtekInfo*, pp. 15–20, Feb. 2023, doi: 10.35134/komtekinfo.v10i1.332.

[30] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *JISKA J. Inform. Sunan Kalijaga*, vol. 7, no. 2, pp. 111–121, May 2022, doi: 10.14421/jiska.2022.7.2.111-121.

[31] L. Zahrotun, Y. Fajri, A. H. S. Jones, and E. Purwaningsih, "Pengelompokkan Mahasiswa Akademik Keperwatan Berdasarkan Asal Sekolah dan Nilai Akademik Menggunakan Metode Clustering K-Means," *Build. Inform. Technol. Sci. BITS*, vol. 3, no. 3, pp. 369–374, Dec. 2021, doi: 10.47065/bits.v3i3.1110.

[32] S. Surohman, L. Fabrianto, F. Riza, and N. M. Faizah, "Korelasi Antara Profil dan Nilai Akademis Siswa dengan Menggunakan Algoritma K-Means," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 8, no. 4, p. 845, July 2021, doi: 10.25126/jtiik.2021843034.