EDUMATSAINS
JURNAL PENDIDIKAN MATEMATIKA DAN SAINS
E-ISSN:2527-7235
Volume 11 Issue. 2, October 2026, pp 258-276
http://ejournal.uki.ac.id/index.php/edumatsains

Grouping Students' Learning Outcomes on Questions regarding Exponents and Roots through the K-Means Approach Utilizing Unsupervised Learning

Ruth Helen Simarmata[1], Esterlina[2], Danny Saputra[3]

[1,3]Universitas Sriwijaya
[2]Universits Pendidikan Indonesia

e-mail: danny@ilkom.unsri.ac.id

Article Info
Article history:
Received: September 30th, 2026
Revised: October 25th, 2026
Accepted: October 26th, 2026
Available online: October 31st, 2026

https://doi.org/10.33541/edumatsains.v11i2.8784

Abstract
This study aims to explore patterns of student academic performance specifically about questions on exponents and roots using K-Means clustering and Principal Component Analysis (PCA). An exploratory quantitative method was employed, using data from a final mathematics examination involving 100 ninth-grade students and 10 questions. Twenty students who gave no answers were excluded from the analysis, leaving 80 students for the clustering process. The function of K- means in this research were defined students on the same group with same categorize with similarities responses. PCA was used for dimensionality and visualize for reduce data distribution groups that found. The analysis revealed two distinct groups where 75 students in the primary group scored a mean of 7.36 while 23 students in the secondary group scored a mean of 3.12. The highest silhouette score from these two different groups is 0,3128 that shows different significant patterns that seen. PCA results showed that primary and secondary principal components accounted for 42.70% and 14.89% of the variance with a cumulative explained variance of 57.58%. K-Means grouping effectively shows variations on these findings based on students' responses with exponents and roots question. However, the diverse for the result between grouping that was written down, some overlap shows that the patterns are not entirely separate. These clustering results can serve as a basis for more detailed diagnostic investigations into how students respond to mathematical problem-solving.

Keywords: Student Assessment Analysis, Principal Component Analysis, K-Means Clustering, Exponent and Roots

1. Introduction

Assessment plays a crucial role in mathematics education by providing evidence of student learning progress while identifying errors, misconceptions, and flawed understandings that can

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

guide future instruction (Shimizu & Kang, 2025). As a result, evaluating mathematics should involve more than just verifying right answers. It must investigate how learners understand issues, choose ideas, utilize methods, and assess their own answers. Analyzing students' exam answers can uncover particular phases of problem-solving where mistakes happen this includes cases where learners confuse conceptual misunderstandings with straightforward calculation errors (Pilgrim et al., 2025) . However, results from summative assessments are generally informed as overall scores. Although a general overview of performance is shown in these scores provided. They can disguise important differences in how students perform on specific questions. Students with similar total scores may demonstrate unique error patterns, in varied level difficulties, and different types of procedural difficulties. Therefore, examining each problem individually is essential for identifying more specific achievement patterns and uncovered learning obstacles that might otherwise be overlooked when looking only for overall scores.

The material of exponents and roots needs this kind of analysis because of the skill proficiency of exponents and roots topic, the relationship between exponents and roots, and excellent algebraic operations skills. Students' written responses were analyzed to identify errors **classified into** the stages of comprehension transformation process execution and final answer presentation (Kenney & Ntow, 2024). Students need to interpret the verbal information from the question into mathematical expression, carry out necessary calculations, and interpret the original context of the resulting solutions all which constitute students' difficulties that might be more pronounced in contextual situations (Wisenöcker et al., 2024). Specifically, regarding exponents and roots, student misconceptions often arise from an inadequate understanding of fundamental concepts, an inability to link exponential and radical forms, and the incorrect generalization of algebraic rules (Şenay, 2024). Related evidence from research on mathematical word problems also indicates that students may grasp parts of a problem yet still make errors during the transformation, procedural, computational, or answer-writing stages of the solution process (Ilonga & Chirimbana, 2024)

These findings indicate that student learning outcomes regarding exponents and roots vary and cannot always be adequately represented by a single total score category. Previous research has thoroughly investigated student misunderstandings and challenges in learning associated with these subjects. Nevertheless, the majority of the evaluations have concentrated on established achievement classifications or particular categories of mistakes. These methods might not adequately reflect scenarios where learners with similar overall scores show significantly diverse learning results. For example, one learner could be strong in symbolic operations but find contextual problems difficult, while another learner, despite having the same total score, may struggle more with processes that include roots. Recognizing these patterns is essential since each one may indicate unique learning requirements, thus necessitating different types of educational support. Progress in educational data mining and machine learning presents chances to analyze patterns in student learning outcomes in a more organized way. Educational data mining combines computational techniques to assess, forecast, visualize, and detect patterns in educational data, utilizing methods such as classification, regression, and clustering (Feng, 2024)

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

Unlike supervised learning which creates models by utilizing predefined classification labels, as shown in a study by (Christian, Friansa, et al., 2024). Unsupervised learning discovers naturally occurring patterns by considering the similarities between various observations. These unsupervised methods identify data groupings, similar to those used by (Christian, Pradipta, et al., 2024) Within the framework of evaluation analysis, this method helps group students based on their response patterns across multiple test items simultaneously, without first defining achievement categories. Delineating learning profiles through an analysis based on students' total scores provides empirical evidence for cluster groupings in educational data. This is in line with research by (Ke Ting et al., 2024) that distinguishes three learner profiles based on each dimension of students' contribution in this research using a method of grouping clustering of K-means. These grouping clustering can classify students with several patterns of particular characteristics as opposed to categorizing students' based on a single overall score.

Another study by (Medina-Gual et al., 2025) shows that students' learning outcomes are detected using grouping cluster, shown that these students' profiles can support the analysis of students' diagnosis needs and the strong development of appropriate educational interventions. Therefore, the function of grouping students' answers is not only to divide students into particular groups but also to identify a combination of characteristics that distinguish students' profile outcome with other students. Therefore, the results of grouping need to be completed with qualitative analysis through students' written answers to show mathematical characteristics from every identified profile.

Accordingly, this research aims to integrate assessment scores on scores level, unsupervised learning and an analysis student written answer. Item-level scores represent student learning outcomes multidimensionally, while clustering identifies student groups based on similarities in their learning outcome patterns. An analysis of written responses is then used to interpret the characteristics of each resulting profile. This approach differs from conventional grouping based on total scores because student profiles are determined in a data-driven manner, based on response patterns across all assessment items. The unsupervised learning analysis was implemented using Python.

Based on this framework, the study aims to analyze student learning outcome profiles regarding exponents and roots using item-level assessment scores and an unsupervised learning approach implemented in Python and to analyze written responses to identify characteristics of learning outcomes, errors, and mathematical difficulties associated with each profile. These findings are expected to provide an analytical overview of the summative assessment results, serve as a basis for designing more appropriate instructional follow up, and assist teachers in understanding the diversity of student learning outcomes.

2. Methods

The methodology is using quantitative approaches comprises the scores by students on each item using unsupervised learning. Unsupervised learning in this research using K-means algorithm for quantitative analysis. Q1 until Q10 was used for variables on grouping process, its process for

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

grouping student based on students' written answer patterns on every question in contrast to using total scores. This research was conducted by all students, namely one hundred students, finished assessment scores with topic exponents and roots. Ten questions are used in this study though summative assessment. The topic of the assessment is exponents and roots. Each item was assigned a maximum score according to the assessment scoring rubric. The maximum score of each mathematical assessment item that has been implemented is presented in table 1

Table 1. Maximum score of each item

Items	Maximum score
Q1	11
Q2	12
Q3	10
Q4	12
Q5	12
Q6	7
Q7	12
Q8	10
Q9	7
Q10	7

The item scores represented students' outcomes on each assessment item, while the total scores represented the sum of the scores obtained across the ten items. These ten item level scores (Q1-Q10), each items score represented students' performance on the corresponding assessment item. Furthermore, the total assessment score was calculated as the sum of ten item scores. For the K-means the ten item scores were used as input variables to characterize students' assessment performance profiles. The total score was used to describe overall assessment performance, instead of as an additional clustering variable. The research procedure consists of four main stages, comprising data collection, data preprocessing, K-means grouping, and cluster evaluation. These stages were systematically conducted to prepare the students' assessment data, identify student performance profiles based on item-level scores, and evaluate the quality and stability of the resulting clusters. The overall research workflow is illustrated in figure 1.

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

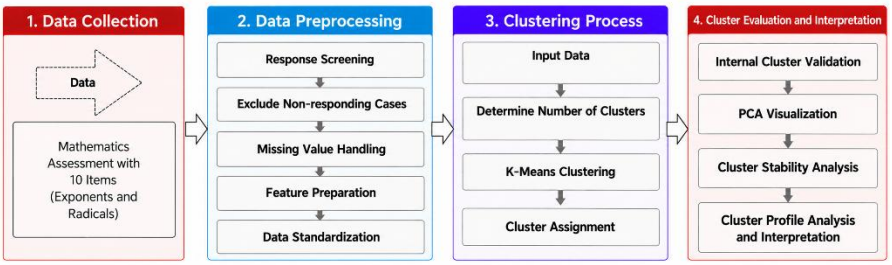


Figure 1. Framework for Data Processing in This Research

Based on figure 1, the data preprocessing stage begins with checking responses completeness and handling missing values. The initial dataset consisted of 100 students where 78 students had complete responses, and two students had partial responses containing five missing item scores while the remaining 20 students did not answer any assessment items. There are 20 students with unanswered cases that were excluded because no observed assessment scores were available to construct students' profiles for distance-based clustering, in contrast to because of presumed low mathematical ability. The two partial response cases were retained to preserve the available information. Missing item scores were handled using median imputation, in which each missing value was replaced with the median of its corresponding assessment item. Therefore, 80 students were included in the analysis. Although median imputation enabled data retention, it may affect score variablitiy and inter-student distances, while excluding fully unanswered cases may introduce selection bias. This preprocess stage is an important part in this analysis machine learning, since the quality of data and statistic characteristic of a dataset can impact the result model of reliability analysis from the outcomes (Uddin & Lu, 2024).

Following the inspection and handling missing data, the scores for ten items were standardized using standard scaler. This process involves transforming the scores for each variable based on its mean and standard deviation, ensuring that all items are on a comparable scale for distance-based measurements. This step is crucial because the k-means relies on distance metrics to determine the proximity between observations. Therefore, variations in the score ranges across different items could otherwise skew the grouping results. Meanwhile, standardization does not guarantee that every item contributes equally to the formation of clusters, it helps mitigate any dominance that might arise solely from differences in scale.

This research using quantitative methods using unsupervised learning with K-Means Algorithm. Ten test items (Q1-Q10) served as input variables for the students' grouping process. Each variable represents a student's score on the corresponding test item, while the total score calculated as the sum of scores across all items is used to reflect overall assessment outcomes in contrast to serving as an additional variable in the grouping process. Therefore, the clustering is based on the patterns of students' scores across individual test items, thereby capturing the similarities in assessment profiles among students'

The number of clusters was determined by evaluating configurations ranging from K=2 to K=10, using both the Elbow Method and the Silhouette Score. In the Elbow Method, the inertia

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

value was observed to assess changes in the within-cluster variation as the number of clusters increased. Meanwhile, the Silhouette Score was employed to evaluate the quality of cluster separation and compactness based on the distances between observations. In this study, the Silhouette Score served as the primary internal validation criterion, while the Elbow Method provided supplementary diagnostic information. The evaluation results indicated that a two-cluster configuration showed the highest Silhouette Score is 0.3128 outperforming all other tested configurations. Therefore, a two-cluster model was selected for further analysis. Nevertheless, this Silhouette Score suggests that the separation between clusters remains limited to moderate. Therefore, the result of groupings should not be interpreted as representing entirely distinct or clear-cut segments.

The K-Means algorithm was applied to the standardized score data, with the number of clusters set to two (K = 2). The *random_state* parameter was set to 42 to ensure reproducibility, while *n_init* was set to 10 to allow the algorithm to run with ten different initializations and select the solution with the lowest inertia. After clustering, cluster stability was evaluated to examine the consistency of the results across different initializations and data samples. The resulting clusters were then visualized in two dimensions using Principal Component Analysis (PCA), providing an overview of how students were distributed based on similarities in their scores. To further examine students' problem-solving strategies and the types of errors they made, written responses were selected purposively from each cluster for further analysis. The findings from these responses were then integrated with the clustering results to provide a more detailed understanding of the problem-solving approaches and error patterns within each group.

The clustering process revealed two groups, namely Cluster 1 representing 71.25% of the sample with 57 students and Cluster 2 representing 28.75% of the sample with 23 students. To identify the characteristics of each cluster, a profile analysis was conducted by calculating the mean scores for each individual test item (Q1–Q10) within each cluster. Comparing these mean scores helped identify items that exhibited relatively larger or smaller score gaps between the clusters. Based on this comparison, the first cluster displayed a profile of relatively higher scores, while the second cluster showed a profile of relatively lower scores. The largest difference in mean scores was observed for Q8 (5.87 points), followed by Q2 (5.48 points) and Q1 (5.05 points). Conversely, the smallest difference in mean scores was found for Q9 (2.38 points). These findings were used to describe the assessment performance patterns within each cluster. Labeling clusters as groups with relatively higher or lower score profiles is descriptive in nature, rather than an absolute classification of students' mathematical ability.

3. Result and Discussion

The number of clusters being used for testing various k values range from k = 2 to k =10. The configuration was determined by observing the inertia values using the Elbow Method and evaluating the quality of cluster separation using the Silhouette Score. The analysis was performed on ten standardized assessment item scores, ranging from Q1 to Q10. The results indicate that inertia decreases from 549.80 at k = 2 to 241,10 at k = 10. This decline follows the expected pattern,

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

as increasing the number of clusters allows for reduced variation within each individual cluster. However, since inertia tends to decrease as the number of clusters grows, this metric alone cannot be used to determine the most appropriate number of clusters.

According to Harahap et al., (2023) the K-Means can be used to group academic achievements based on similarities across several assessment indicators. This finding is relevant because the present study also avoids relying solely on a single total score, as opposed to analyzing the patterns of scores from ten individual test items. Therefore, the resulting clusters illustrate the varying achievement tendencies among students. However, the similarity in approach does not imply that the number or characteristics of the clusters must be identical, as the variables employed differ. The result of elbow method is illustrated in figure 2.

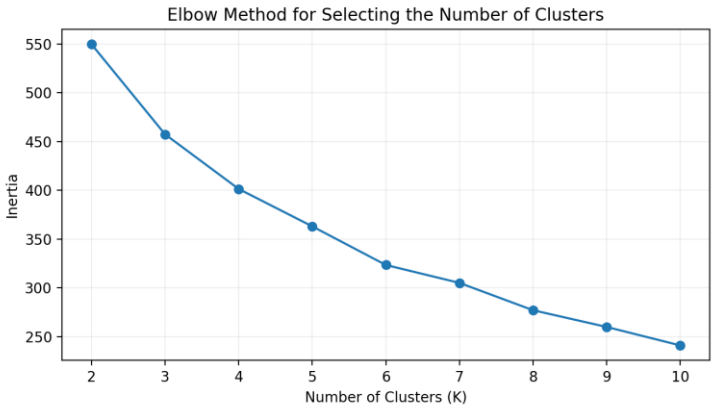


Figure 2. Result of Elbow Method

According to Figure 1, the inertia value decreases significantly with the initial addition of a few clusters, after which it decreases more gradually. This graph provides insight into the trade-off between the number of clusters and the within cluster variation. However, it does not reveal an elbow point that can be definitively identified as the sole optimal choice. Therefore, the elbow method is treated as a piece of supporting diagnostic information.

In the Elbow Method, the inertia value measures the sum of the squared distances of student scores from their respective cluster centers. As the number of clusters increases, it becomes easier for the algorithm to assign students to centers that are closer to them. Therefore, inertia generally decreases even if adding further segments is not necessarily useful for educational purposes. The graph in this study shows a decline from K=2 to K=10, yet it lacks a sufficiently distinct change in slope to definitively identify a single optimal K value. Therefore, the selection cannot be based solely on the magnitude of the inertia. This interpretation is crucial, while increasing the number of clusters may yield more specific profiles, it also carries the risk of splitting groups that share similar score patterns.

Figure 2 shows the highest score from silhouette scores found in k = 2 with figure of 0,3128. The values for other configurations were obtained on range between 0,213 and 0,254. As it has the highest score among the configurations tested, k=2 is selected as the final chosen as best solution for analysis of this research.

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

According to Hilyawan et al., (2025) show that student performance segmentation, emphasize that determining the number of clusters requires considering both validity metrics and the meaningfulness of the resulting groups. The Elbow Method illustrates the extent to which score variation within clusters decreases as k = 2 increases, while the Silhouette Score indicates how well each group is separated from the others. These two metrics complement each other because a continuously decreasing inertia is not sufficient justification for adding more clusters if the separation between them does not improve. Another study by Azmi et al., (2026) also employed the Elbow Method alongside validity indices to evaluate the clustering of student learning outcomes. The relevance of both studies lies in the principle of evaluating clusters using multiple criteria, rather than in a requirement to arrive at the same number of clusters. In the dataset for this study, the Elbow Method did not yield a clear elbow point, whereas the highest Silhouette Score was obtained at k = 2. Therefore, k =2 was selected as the most justifiable configuration among the alternatives tested, although the strength of the separation must still be assessed with caution. The result of Silhouette score is illustrated in figure 3.

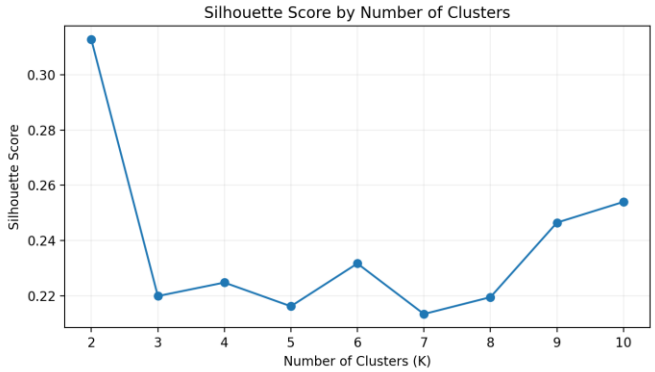


Figure 3. Result of Silhouette Score

Although k=2 showed the best score among the options evaluated, the value of 0.3128 does not indicate strong cluster separation. These results are more accurately interpreted as reflecting limited to moderate separation, with the possibility that students in different clusters may exhibit similar score profiles. Therefore, the two clusters are selected as an exploratory representation of score patterns within the dataset, rather than as evidence that students naturally fall into two distinct ability categories.

A Silhouette Score of 0.3128 suggests that the separation between the two groups is not yet robust, even though it represents the best value among the range of k=2 to k=10. This implies that some students across different clusters may still share relatively similar score patterns. According to Azmi et al., (2026) demonstrated that K-Means analysis of learning outcomes can yield varying numbers of clusters and validity scores, depending on the structure of the data employed. Therefore, discrepancies in findings across different studies should not be used as a definitive benchmark to claim that one model is superior to another. In this study, the two clusters are

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

understood as a summary of score patterns useful for exploratory purposes, rather than an absolute classification of student ability. The distribution of students in each cluster that has been implemented is presented in table 2.

Another study by (Mahdy et al., 2025) compared several clustering algorithms indicates that the choice of method and the number of clusters should be determined based on an evaluation of how well the results fit the data. In that study, a hierarchical clustering approach revealed a two-cluster solution that was deemed the most appropriate according to the validity indices employed outperforming K-Means. This comparison underscores that the selection of two clusters in the present study is not a universal rule applicable to every dataset. Therefore, K=2 was retained because it indicated the highest Silhouette Score among the K-Means configurations tested, while acknowledging that the quality of the separation remains limited.

Table 2. Distribution of Students in Each Cluster

	Number of students	Percentage	Average	Control
Cluster 1	57	71,25%	7,36	High scoring group
Cluster 2	23	28,75%	3,12	Low scoring group
Total	80	100%		

The K-Means model produced two clusters. Cluster 1 consists of 57 students in percentage 71.25% and Cluster 2 consists of 23 students (28.75%). Based on a comparison of scores for each item, Cluster 1 exhibits a relatively higher score profile, whereas Cluster 2 shows a relatively lower score profile. These labels are used to facilitate the presentation of results and are descriptive in nature, rather than representing absolute categories of mathematical ability. The average profiles reported in the initial analysis are 7.36 for Cluster 1 and 3.12 for Cluster 2. While these figures indicate a difference in score levels between the groups, they must be interpreted in conjunction with the score patterns for each individual item. Cluster membership reflects relative similarity across ten score variables, as an alternative to a comprehensive assessment of student ability.

The majority of students, totaling 57 students or 71.25%, were classified into the profile with relatively higher scores, while 23 students or 28.75% were classified into the profile with relatively lower scores. This proportion is not equivalent to a mastery rate, as the grouping is based on score similarity rather than a passing threshold. Consistent with the findings of Sabrina Rochadi & Ridwan Yudhanegara., (2024) revealed that clustering mathematics test results can help reveal a diversity of ability profiles that cannot be adequately captured by a single summary measure. This is relevant to the current study, which utilizes multiple scores to examine achievement patterns among students. However, the two profiles generated here describe only the characteristics of scores on items Q1–Q10 and cannot be directly used to establish general categories of mathematical ability.

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

Another research by (Rahma, 2025) show that grouping various achievement indicators can yield distinct student profiles, aiding in the understanding of diverse learning outcomes. also employed K-Means clustering to categorize academic performance. Both studies support the use of clustering as an initial overview of achievement variation. The distinction in the present study is its focus on ten specific items within a single subject area. Therefore, the interpretation is confined to response patterns regarding exponents and roots, without accounting for social factors or the underlying causes of the observed learning outcomes.

As reported by Farha et al., (2024) also demonstrates that similarities in academic outcomes can be used to form achievement groups via the K-Means method. While their study identified three groups, the present study identified two, based on the score patterns across ten items. Such differences are expected, given that the datasets and the degree of inter-individual variability differ between the two studies. The primary relevance lies in the interpretation of the clustering results. Cluster labels are assigned only after examining the score characteristics of each group, instead of pre-defined categories of ability established prior to the analysis.

It should be noted that the analysis includes 80 students comprising 78 with complete responses and two with partial responses, for which missing values were imputed using the median value for each item. Twenty students who did not provide answers to any of the items were excluded, as they lacked the observed scores necessary to construct a distance-based profile. This exclusion does not imply that they have low ability, and it may introduce selection bias. Therefore, the cluster percentages apply only to the data analyzed.

The study conducted by (Alvarez-Garcia et al., 2024) revealed that cluster analysis can help identify the diversity of student profiles when multiple indicators are examined simultaneously. This finding is relevant to the present study because aggregate scores alone may mask differences in performance on individual items. However, the resulting clusters are inherently dependent on the specific indicators being analyzed. Therefore, this study limits its interpretation to patterns in assessment scores, rather than addressing student characteristics that were not measured.

In order to help readers visualize the distribution of students, the clustering results are represented in two dimensions using Principal Component Analysis (PCA). PCA is applied after the clustering process and serves solely for visualization purposes. The K-Means model itself is trained using all ten standardized item scores. The result K-means clustering is illustrated in figure 4.

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

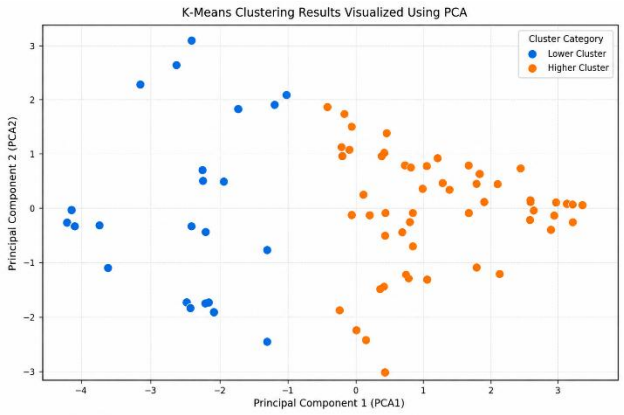


Figure 4. Result of K- Means Clustering Using PCA

The first principal component (PC1) accounts for 42.70% of the data variance, while the second principal component (PC2) accounts for 14.89%. Based on these rounded figures, the two components together represent approximately 57.59% of the data variance. This percentage reflects the visual representation capacity of PCA rather than the quality of cluster separation.

Figure 3 illustrates the distribution tendencies of the two profiles in distinct areas. However, it also demonstrates that a two-dimensional representation cannot capture the full information contained within the ten variables. The distances and proximity between points on the PCA plot are projections. Therefore, the original analytical space does not perfectly correspond to the distances within While this visualization facilitates the exploration identification of patterns, it does not replace the results of numerical validation.

PCA condenses ten score variables into two linear combinations that capture the greatest variance in the data. PC1 explains 42.70% of the variance, and PC2 explains 14.89%, with total percentage is 57.59% (based on rounded values). This implies that approximately 42.41% of the variance remains outside the two-dimensional plane shown in Figure 4. Therefore, points that appear close together in the plot are not necessarily highly similar when all ten scores are considered. Previous research by Rahma., (2025) employed PCA and t-SNE to visualize student clusters derived from academic and social data, demonstrating the use of projection at communicate profiles. In the present study. However, K-Means clustering was explicitly performed within the ten-dimensional space of standardized variables, with the results subsequently projected solely for visualization purposes. Given this approach, the observed distribution in the plot should not be used to replace or refine the Silhouette Score results.

Similar research by Wu, (2026) indicates that K-Means utilized clustering results as supplementary information for modeling student performance, illustrating that cluster structures can offer further analytical utility. However, this study does not construct a predictive model. PCA is used solely to clarify the visualization of the distribution, while K-Means continues to operate based on the ten item scores. Therefore, the separation of points on the PCA plot should not be interpreted as evidence of predictive accuracy or as a validation of the resulting clusters.

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

The subsequent analysis compares the mean scores for each individual test item to understand where the gaps between clusters appear larger or smaller. This comparison provides a more detailed breakdown than summary scores alone. The average score for each item was calculated based on the student results within each cluster. Cluster 1 comprised 57 students with relatively higher profile scores, whereas Cluster 2 consisted of 23 students with relatively lower profile scores. The score difference was determined by subtracting the average score of Cluster 2 from that of Cluster 1 for each test item. The results of this comparison are presented in Table 3.

Table 3. Comparison of Average Scores Per Items

Items	Relatively higher profile	Relatively higher profile	Difference
Q1	10,14	5,09	5,05
Q2	9,74	4,26	5,48
Q3	8,53	5,04	3,49
Q4	8,96	5,00	3,96
Q5	5,68	1,22	4,46
Q6	6,19	1,52	4,67
Q7	4,19	0,61	4,02
Q8	10	2,74	5,87
Q9	6,21	3,83	2,38
Q10	4,88	1,87	3,01

The largest difference is observed in Q8, with a margin of 5.87 points, followed by Q2 with 5.48 points and Q1 with 5.05 points. Meanwhile, the smallest difference is found in Q9, at 2.38 points. These results indicate that the achievement gap between clusters varies across individual test items. Furthermore, Q7 has the lowest average score in both clusters 4.63 and 0.61, respectively making it an item that requires further review to better understand student achievement patterns. The result of Mean item score is illustrated in figure 5.

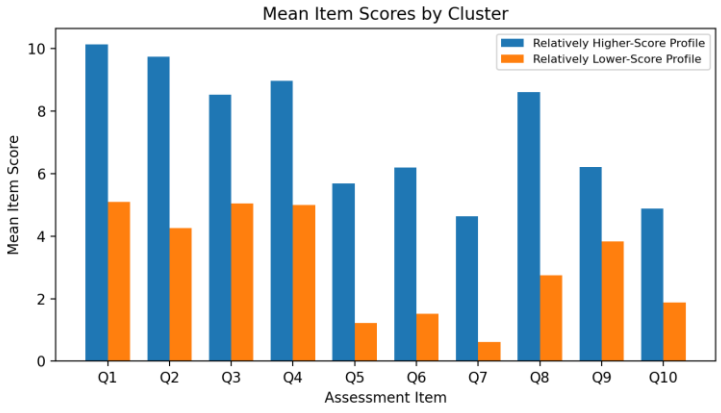


Figure 5. Result of Mean Item Score

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

Based on figure 5, the largest mean difference was found for Q8 (5.87 points), followed by Q2 (5.48) and Q1 (5.05). Conversely, the smallest differences were observed for Q9 (2.38), followed by Q10 (3.01). These results indicate that the score gap between the two profiles varies by item and is not uniform across the entire assessment. For the relatively higher profile, the highest means were found for Q1 (10.14), Q2 (9.74), Q4 (8.96), and Q8 (8.61), while lower means were observed for Q7 (4.63) and Q10 (4.88). For the relatively lower profile, the highest means were found for Q1 (5.09), Q3 (5.04), and Q4 (5.00), whereas the lowest scores were observed for Q7 (0.61), Q5 (1.22), Q6 (1.52), and Q10 (1.87).

Comparing individual items provides insights that cannot be gleaned solely from the cluster averages of 7.36 and 3.12. For instance, Q8 exhibits the largest gap between the two clusters (5.87 points). This indicates a notable disparity in performance between the groups on this item. However, without examining the mathematical demands of the question and its maximum possible score, it remains unclear whether this item is the most difficult or simply the most effective at discriminating between students' ability levels. Conversely, Q7 shows low averages for both clusters 4.63 and 0.61 which requires the teacher's attention, as performance on this item is relatively poor across both profiles. Nevertheless, determining the appropriate follow-up action requires an examination of the students' actual work, rather than relying solely on a comparison of averages.

The findings from a study by (Octaria et al., 2026) provide a closer point of comparison within the context of mathematics education. They applied the K-Means algorithm to mathematics competition results using standardized scores and evaluating the clusters via the Elbow Method and Silhouette Score ultimately identifying three distinct achievement groups. In the present study, these same evaluation tools were employed, yet the optimal results pointed toward two score profiles. This comparison reinforces the rationale that the number of clusters should be determined by the patterns observed in the data. Furthermore, while mapping based on mathematics scores can reveal the diversity of student achievement, it does not yet account for the underlying thought processes that lead to students' answers.

The low mean scores for Q7, at 4.63 and 0.61, suggest a different pattern from the substantial difference seen in Q8. Q7 requires further examination due to the consistently low performance across both student profiles, while Q8 requires attention because of the considerable variation in mean scores between the two groups. These findings provide a basis for further analysis, particularly in finding the specific requirements of Q7 that students have not adequately fulfilled and examining the differences in responses between higher- and lower-scoring students on Q8. Nevertheless, without standardized maximum scores for each item and a comprehensive analysis of the instrument's content, comparisons of raw scores alone cannot provide a reliable basis for figuring out the relative difficulty of individual questions.

Differences in ability across clusters for each test item indicate that students' mastery of the material on exponents and roots is not entirely uniform. In the context of problems involving exponents. A study conducted by Liawati et al., (2026) found that variations in students' thought processes based on the APOS theory stages action, process, object, and schema. In their study, students with higher ability levels demonstrated more comprehensive connections between these

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

stages, whereas students with moderate or low ability levels faced limitations at more complex stages of understanding. These findings offer a valuable perspective for interpreting the score variations observed in the current study. In particular, the low performance observed on Q5, Q6, Q7, and Q10 requires further investigation through written responses or interviews to understand how students conceptualize the material and execute the solution steps. However, since this study only categorizes scores from Q1–Q10 without analyzing the APOS stages, the differences between clusters cannot be directly interpreted as evidence of distinct thought processes or specific misconceptions. Therefore, the K-Means results serve as an initial mapping to identify which parts of the assessment require more in-depth examination.

The findings reported by Sabrina Rochadi & Ridwan Yudhanegara., (2024) shows that grouping students based on mathematics test results can help reveal differences in their profiles regarding conceptual understanding and problem-solving. They identified three clusters, noting that the relationship between these two abilities varies across the clusters. The relevance of this to the present study lies in the importance of examining internal patterns in contrast to merely counting the number of students once they have been grouped. For instance, in the data for items Q1–Q10, Cluster 1 shows high averages for Q1 and Q2, but significantly lower averages for Q7 and Q10. This suggests that even a group with a generally higher overall profile may still contain specific items that require further review. However, because the current study does not categorize each test item into specific indicators of conceptual understanding or problem-solving, these numerical values cannot yet be interpreted as representing the same cognitive abilities. The result of Item level score profiles are illustrated in figure 6.

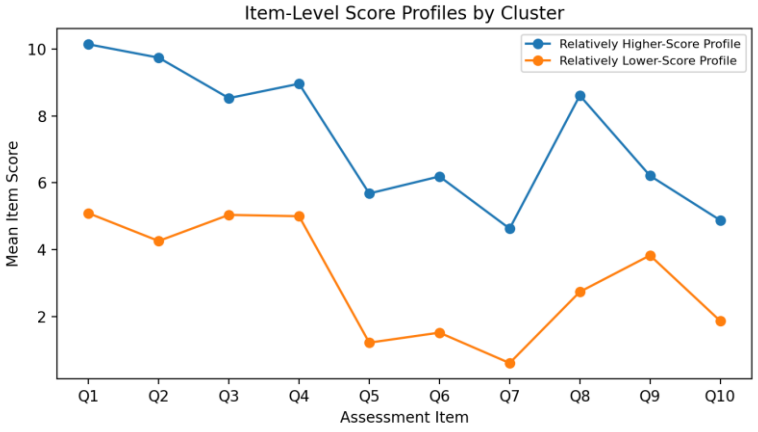


Figure 6. Results of Item Level Score Profiles

These differences can help educators show items that requires further review. However, a gap in the mean score does not prove that a particular item is more effective, more difficult, or measures a specific ability. Drawing such conclusions requires an examination of the item content, the scoring rubric, the competency indicators, and the psychometric characteristics of the item.

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

Similarly, a lower score is not direct evidence of a specific misconception, procedural weakness, or calculation error.

The findings show that grouping students based on their scores across ten items can reveal variations in assessment performance profiles that might otherwise go unnoticed if attention were focused solely on the total score. Students with similar total scores may show different distributions of scores across individual items, while students with relatively similar scoring patterns may fall into the same cluster. Thus, clustering provides an exploratory method for seeing similarities in achievement patterns, rather than serving as a replacement for the individual evaluation of student learning.

Previous research by (Yuting Tan & Tao Yang, 2026) shows that clustering in the study of mathematical problem-solving can reveal diverse student profiles, even when indicators of the problem-solving process are included in the analysis. In the context of the present study, this implies that students do not necessarily show the same achievement patterns across all aspects of a task. However, this study relies solely on scores from items Q1-Q10, rather than the sequence of steps taken or the students' underlying thought processes. Therefore, while the identified clusters highlight differences in performance on specific items, they do not yet account for variations in strategies, misconceptions, or types of errors.

From an instructional perspective, teachers might prioritize items Q8, Q2, and Q1 to examine what distinguishes the responses of the two groups, while items Q7, Q5, Q6, and Q10 requires further investigation, given that the average performance for one or both profiles is relatively low. Such an examination could involve reviewing rubrics, analysing examples of correct and incorrect responses, scrutinizing calculation steps, and assessing how well each item aligns with specific learning goals. This approach is not an automatic outcome of the clustering algorithm itself, but rather a pedagogical follow-up that can be pursued once numerical patterns have been found. The results obtained by Rahma, 2025; Sujana et al., (2025) indicated that clusters serve as a first indicator of achievement heterogeneity such as instructional decisions. However, still require more context. In this way, the clustering results do not merely serve as labels for distinct groups but rather provide a foundation for asking more targeted questions about students' needs.

The research conducted by (Alaoui et al., 2026) suggest that the clustering of student performance can serve as a basis for tailoring learning support to specific achievement profiles. This is relevant to the present study, as score-based groupings can help determine the focus of assessment reviews. Nevertheless, K-Means clustering results alone are insufficient for designing effective interventions. Teachers must still examine the content of assessment items, rubrics, and student responses before making instructional decisions.

Research conducted by (Sulaiman et al., 2025) demonstrates the use of K-Means to group students based on indicators of interest in learning mathematics, with observable differences in exam performance across the groups. While these findings highlight the potential of cluster analysis as a tool for pedagogical reflection, the variables employed in that study differ from those used here. Since the grouping in the present study is based solely on scores from items Q1–Q10, it is not possible to draw conclusions regarding students' interests or motivation. A righter application

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

would be to use these score profiles as a first guide for selecting specific items for further examination, while supplementing the results with evidence from the learning process.

In practice, score profiles can provide supporting information for planning assessment reviews, helping teacher discussions, or further investigating student needs. Students falling into a relatively low-scoring profile should not be at once labelled or assigned to a specific intervention based solely on their cluster. Instead, teachers should integrate this information with other forms of evidence such as results from different assessments, classroom observations, and dialogues with students. This recommendation stands for a potential implication of the quantitative mapping, rather than an intervention outcome that has been empirically evaluated in this study.

Overall, this study demonstrates the utility of the K-Means Algorithm in simplifying the variation in scores across ten items into two more interpretable, descriptive profiles. Its primary advantage lies in its ability to reveal patterns of achievement at the item level and provide a focal point for assessment evaluation. However, the modest Silhouette scores, limitations in data completeness, and the absence of an analysis of the problem-solving process preclude drawing causal conclusions or establishing permanent ability categories. Future research could examine the stability of these clusters across different samples, compare various clustering methods, and link scores to an analysis of errors in student responses. Therefore, the primary value of the current findings lies in their role as an exploratory mapping tool helping educators understand variations in student achievement instead of as a substitute for an in-depth assessment of each student's individual learning process.

4. Conclusion

This study aims to cluster the learning outcomes of ninth-grade students on problems involving exponents and roots, using the K-Means method within an unsupervised learning framework. Based on an analysis of data from 80 students comprising 78 complete responses and 2 partial responses with missing values imputed using the median for each item testing for the number of clusters (from K=2 to K=10) revealed that K=2 is the optimal configuration, yielding the highest Silhouette Score of 0.3128. The clustering process resulted in two groups where Cluster 1 contained 57 students or 71.25% of the sample with a relatively higher score profile and Cluster 2 contained 23 students or 28.75% of the sample with a relatively lower score profile. Principal Component Analysis (PCA) is used on this research to illustrate the distribution of these two clusters, with the first two principal components accounting for approximately 57.59% of the data variance (rounded figures).

An analysis of the mean scores for each item indicates that the largest gap between the clusters occurs in Q8 (5.87 points), followed by Q2 (5.48 points) and Q1 (5.05 points), while the smallest gap is found in Q9 (2.38 points). These findings suggest that K-Means can serve as an exploratory approach for mapping patterns of similarity in student assessment results and identifying specific items that require further investigation. Nevertheless, the Silhouette Score indicates that the cluster separation is limited to moderate. These findings show that K-Means can be used as an exploratory approach to map the similarity of student assessment patterns, revealing

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

the diversity of achievements on each item, and identifying items that need further review. Besides generating student groupings, this study provides an illustration that differences in learning outcomes are not only visible from overall scores but also from variations in achievement on each question item. This information adds value because it can help teachers get a more detailed mapping of assessment results and serve as a basis for evaluating the attainment of learning in exponent and root topics. Moreover, this research offers a chance to develop learning outcome analysis more efficiently and systematically by using the assessment data that's already available. So, this study contributes by applying unsupervised learning to explore student achievement patterns while also providing initial information that can be used in developing data-based math learning evaluations.

5. References

Alaoui, G. M., Zouhair, A., & Khabbachi, I. (2026). Students' performance clustering for future personalized in learning virtual reality. *International Journal of Electrical and Computer Engineering (IJECE)*, *16*(1), 297. https://doi.org/10.11591/ijece.v16i1.pp297-310

Alvarez-Garcia, M., Arenas-Parra, M., & Ibar-Alonso, R. (2024). Uncovering student profiles. An explainable cluster analysis approach to PISA 2022. *Computers and Education*, *223*. https://doi.org/10.1016/j.compedu.2024.105166

Azmi, M. F., Abidin, D. Z., & Jasmir, J. (2026). K-Means Clustering with Elbow Method and Validity Indices for Classifying Student Academic Achievement Based on Knowledge Scores at SDN 48 Kota Jambi. *Jurnal Teknik Informatika (Jutif)*, *7*(1), 571–586. https://doi.org/10.52436/1.jutif.2026.7.1.5349

Christian, H., Friansa, K., Pradipta, J., Haq, I. N., & Leksono, E. (2024). Comparasion of HVAC Energy Consumption Prediction in an Academic Building using LSTM and DNN. *Jurnal Ecotipe (Electronic, Control, Telecommunication, Information, and Power Engineering)*, *11*(1), 77–87. https://doi.org/10.33019/jurnalecotipe.v11i1.4488

Christian, H., Pradipta, J., Haq, I. N., Hanadi, Friansa, K., Leksono, E., & Esterlina. (2024). Reliability Evaluation of Advanced Metering Infrastructure Based on Isolation Forest Anomalies Detection. *International Conference on Artificial Intelligence and Mechatronics System, AIMS 2024*. https://doi.org/10.1109/AIMS61812.2024.10512537

Farha, I., Nola Febriana Saputri, Melati Elvita, Andri Azmul Fauzi, & Fidia Deny Tisna Amijaya. (2024). Klasterisasi Potensi Mahasiswa Program Studi Matematika Universitas Mulawarman Menggunakan Metode K-Means Untuk Menyaring Mahasiswa Berprestasi. *MILANG Journal of Mathematics and Its Applications*, *20*(2), 89–99. https://doi.org/10.29244/milang.20.2.89-99

Feng, G. , & F. M. (2024). Research on learning behavior patterns from the perspective of educational data mining: Evaluation, prediction and visualization. *Elsevier*, *237*.

Harahap, A. A., Raihan, M., Amani, N., & Andini, R. (2023). *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Comparison of Unsupervised Learning Techniques for Clustering Data on the Number of Villages in Indonesia Perbandingan Teknik Unsupervised*

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

Learning untuk Pengelompokan Data Jumlah Desa Di Indonesia. https://journal.irpi.or.id/index.php/sentimas

Hilyawan, F. R., Syafrullah, M., & Sudija, I. (2025). A Systematic Literature Review on the Optimization of K-Means and Agglomerative Clustering for Student Performance Segmentation: A Comparative Analysis of Elbow and Silhouette Methods. *Edcomtech: Jurnal Kajian Teknologi Pendidikan*, *10*(1), 73–88. https://doi.org/10.17977/um039v10i12025p73-88

Ilonga, H. K., & Chirimbana, M. (2024). *Iilonga, H. K., & Chirimbana, M. (2024)*.

Ke Ting, C., Ibrahim, N., Huspi, S. H., & Wan Kadir, W. M. N. (2024). Multidimensional Context Clustering to Analyse Student Engagement in Online Learning Environment. *International Journal of Innovative Computing*, *14*(2), 89–96. https://doi.org/10.11113/ijic.v14n2.487

Kenney, S., & Ntow, F. D. (2024). Unveiling the Errors Learners Make When Solving Word Problems Involving Algebraic Task. *SAGE Open*, *14*(4). https://doi.org/10.1177/21582440241299245

Liawati, Huda, N., & Marzal, J. (2026). *Analyzing Students' Thinking Processes in Solving Exponential Problems Based on APOS Theory: A Qualitative Study at SMA Negeri 7 Tanjung Jabung Timur*. *10*(4), 69–81. https://doi.org/10.33541/edumats

Mahdy, I. F., Haqqi, M. W. N., & Oktiardi, R. N. (2025). Comparative Study Of Clustering Algorithm To Determine Student Profiles Based On Academic Ability. *VARIANCE: Journal of Statistics and Its Applications*, *7*(2), 231–240. https://doi.org/10.30598/variancevol7iss2page231-240

Medina-Gual, L., Medina-Velázquez, L., & Parejo, J.-L. (2025). A tridimensional model of AI literacy: An empirical analysis of student performance and demographic patterns in higher education. In *Australasian Journal of Educational Technology* (Vol. 2025, Number 5).

Octaria, E. A., Ridwan, R., & Saputra, H. (2026). Pemetaan Kemampuan Matematis Peserta International Kangaroo Mathematics Contest Menggunakan K-Means Clustering. *JPG: Jurnal Pendidikan Guru Journal*, *6*(1), 64–75. https://ejournal.uika-bogor.ac.id/index.php/jpg/index

Pilgrim, M. E., Burks, L. C., Ryals, M., & Hill-Lindsay, S. (2025). College algebra students' perceptions of exam errors and the problem-solving process. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1359713

Rahma, F. (2025). Clustering Students Based on Academic Performance and Social Factors: An Unsupervised Learning Approach to Identify Student Patterns. *International Journal for Applied Information Management*, *5*(3), 139–154. https://doi.org/10.47738/ijaim.v5i3.109

Sabrina Rochadi, Z., & Ridwan Yudhanegara, M. (2024). *Jurnal Didactical Mathematics Analisis Cluster untuk Hubungan antara Kemampuan Pemahaman Konsep dan Kemampuan Pemecahan Masalah Matematis Siswa Menggunakan K-Means Clustering* (Vol. 6, Number 1). https://ejournal.unma.ac.id/index.php/dm

Şenay, Ş. C. (2024). Analysis of Misconceptions and Errors Regarding Exponential and Radical Expressions Through the Theory of Reducing Abstraction. *Research on Education and Psychology*, *8*(2), 281–295. https://doi.org/10.54535/rep.1520588

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

Shimizu, Y., & Kang, H. (2025). Research on classroom practice and students' errors in mathematics education: a scoping review of recent developments for 2018-2023. *ZDM - Mathematics Education*, *57*(4), 695–710. https://doi.org/10.1007/s11858-025-01704-0

Sujana, T., Astuti, R., Prihartono, W., & Hamonangan, R. (2025). *Journal of Artificial Intelligence and Engineering Applications Implementation of Data Mining Using the K-Means Algorithm to Group Students Based on Academic Performance* (Vol. 4, Number 2). https://ioinformatic.org/

Sulaiman, H., Saepul Rohman, R., Wintana, D., & Abdul Ghani, M. (2025). Analyzing Students' Interest in Mathematics Through the Implementation of the K-Means Clustering Algorithm. *Journal of Information and Technology) Accredited Sinta*, *4*.

Uddin, S., & Lu, H. (2024). Dataset meta-level and statistical features affect machine learning performance. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-51825-x

Wu, M. (2026). *ICCK Transactions on Educational Data Mining K-Means Clustering-Based Feature Generation for Student Performance Prediction*. *2*(1). https://doi.org/10.62762/TEDM.2026.716076

Yuting Tan, & Tao Yang. (2026). Cognitive process grouping on students' mathematical creative problem solving: Evidence from clustering analysis, lag sequence analysis, and hidden Markov model. *Elsevier*, *252*.

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia