Application of Machine Learning for Clustering Academic Achievement Levels of Elementary School Students

Ahmad Syaukani, Muhammad Syahputra Novelan

Abstract

Academic achievement clustering provides an opportunity to identify student performance patterns and support data-driven educational evaluation. This study aims to apply the K-Means Clustering algorithm to group academic achievement levels among students at Al Muttaqien Private Elementary School. The research employed a quantitative approach using academic records from Grades 1 to 6 covering two semesters. Seven subject scores were selected as numerical features, including Islamic Religious Education, Pancasila, Indonesian Language, English, Mathematics, Physical Education, and Arts and Culture. The dataset was processed through data cleaning, feature selection, and standardization before the clustering algorithm was implemented. A total of 311 student-semester records were included in the analysis, and a three-cluster solution was examined to identify relatively lower, moderate, and higher academic achievement profiles. The results indicated average academic scores of 76.51, 83.32, and 89.29 for the respective clusters, representing 26.4%, 37.9%, and 35.7% of the analyzed records. Cluster evaluation using the Silhouette Score showed that the three-cluster solution achieved a score of 0.387, while the two-cluster solution achieved a higher score of 0.484. These findings indicate that K-Means can provide an interpretable grouping of academic performance, although the selection of the optimal number of clusters requires consideration of both statistical validation and educational objectives. The resulting profiles may support teachers in identifying differences in student achievement and planning differentiated learning strategies. This study demonstrates the potential of unsupervised machine learning as a supporting approach for academic data analysis in elementary education.

***Keywords**:* Machine Learning, K-Means Clustering, Academic Achievement, Student Performance, Educational Data Analysis.

[1]Ahmad Syaukani
Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: kaniharahap@gmail.com

[2]Muhammad Syahputra Novelan
e-mail: putranovelan@dosen.pancabudi.ac.id
2nd International Cofference on the Epicentrum of Economic Global Framework (ICEEGLOF)
Theme: Navigating The Future: Business and Social Paradigms in a Transformative Era.
https://proceeding.pancabudi.ac.id/index.php/ICEEGLOF

Introduction

Education plays a fundamental role in developing students' knowledge, skills, and competencies from an early age. Academic achievement is one of the important indicators used to evaluate students' learning progress and the effectiveness of educational activities. At the elementary school level, students have different learning abilities, levels of subject comprehension, and academic development patterns. These differences require teachers to understand students' achievement characteristics more comprehensively to ensure that learning strategies are aligned with their educational needs. However, academic evaluation is often conducted by examining individual subject scores and calculating average grades without systematically identifying similarities among students. Consequently, the patterns underlying variations in academic achievement may not be fully recognized. Therefore, a systematic analytical approach is needed to organize academic data into meaningful groups that can support educational evaluation and improve the effectiveness of learning management [1] [2].

The increasing availability of academic data in educational institutions creates opportunities to apply machine learning techniques to support data-driven decision-making. Machine learning enables computers to identify patterns and relationships within datasets through computational methods. In educational research, machine learning can be applied to analyze student performance, identify learning patterns, and support the evaluation of academic achievement. One relevant approach is unsupervised learning, which discovers structures within data without requiring predefined class labels. Unlike supervised learning, which depends on previously labeled examples to train a predictive model, unsupervised learning focuses on identifying groups based on the similarities and differences among observations. This characteristic makes unsupervised learning suitable for exploring academic performance patterns when students have not been classified into predetermined achievement categories [3] [4] .

One of the algorithms commonly used for unsupervised learning is K-Means Clustering, which groups data into a specified number of clusters based on the distance between observations and cluster centroids. The algorithm iteratively assigns each observation to the nearest centroid and updates the centroids until the clustering process converges. In the context of elementary school education, K-Means can be applied to numerical academic features, including subject scores, semester grades, and average academic performance. Through this process, students with similar academic characteristics can be grouped into the same cluster, while students with different characteristics are assigned to other clusters. The resulting groups may be interpreted as relatively high, moderate, and low academic achievement levels based on their statistical profiles. However, these interpretations must be established after examining the characteristics of each cluster rather than being assumed as predetermined outputs of the algorithm [5] [6] .

The implementation of K-Means clustering requires appropriate data preparation and evaluation to produce reliable and interpretable results. The process begins with collecting students' academic records, followed by data preprocessing to identify missing values, duplicate records, inconsistent entries, and other data quality issues. The next stage involves selecting relevant numerical features that adequately represent students' academic performance. If the selected variables have substantially different scales, appropriate standardization may be necessary to prevent certain features from disproportionately influencing distance calculations. The number of clusters can be explored using the Elbow Method, while the Silhouette Score can be employed to evaluate the cohesion of observations within clusters and their separation from other clusters. These procedures provide a methodological foundation for assessing the quality of the resulting groups and ensuring that the interpretation of academic achievement is supported by the characteristics of the analyzed data [7] [8] .

Based on these considerations, this study aims to apply the K-Means Clustering algorithm to group academic achievement levels among students at Al Muttaqien Private Elementary School. The research focuses on analyzing academic records, identifying groups of students with similar performance characteristics, and interpreting the resulting clusters to obtain a clearer understanding of academic achievement patterns. The findings are expected to assist teachers and school administrators in identifying variations in student performance, planning differentiated learning strategies, and determining which students may benefit from additional academic support. Furthermore, this study is expected to contribute to the application of machine learning in educational data analysis by demonstrating how clustering techniques can transform numerical academic records into meaningful information for educational evaluation. The results may also provide a foundation for developing data-driven decision-support approaches that support more systematic and informed learning management in elementary education [9] .

Literature Review

2.1 Machine Learning in Educational Data Analysis

Machine learning is a branch of artificial intelligence that enables computer systems to identify patterns, relationships, and structures within datasets to support data analysis and decision-making. In the educational domain, machine learning has been increasingly applied to analyze student performance, identify learning patterns, evaluate academic achievement, and support the development of data-driven educational strategies. Academic data, including subject scores, semester grades, attendance records, and learning assessment results, can provide valuable information about students' learning progress. By applying machine learning techniques, educational institutions can process these data systematically and identify performance characteristics that may be difficult to recognize through conventional evaluation methods. In particular, machine learning can support the grouping of students according to similarities in their academic profiles, allowing teachers to obtain a more structured understanding of differences in learning achievement. The application of machine learning in education is therefore relevant to improving the analytical use of academic records and supporting more informed instructional planning[10] [11].

2.2 K-Means Clustering Algorithm

K-Means is an unsupervised machine learning algorithm widely used to group observations based on similarities in their numerical characteristics. The algorithm operates by assigning data points to clusters according to their proximity to cluster centroids and repeatedly updating the centroids until the clustering process reaches a stable condition. One of the main advantages of K-Means is its relatively straightforward implementation and its ability to process numerical datasets efficiently. In educational research, the algorithm can be applied to academic scores, subject grades, and average student performance to identify groups with similar achievement characteristics. For example, the resulting clusters may represent students with relatively high, moderate, and low academic achievement, depending on the observed statistical profiles of each group. Nevertheless, the quality of clustering depends on data preparation, feature selection, the number of clusters, and the distribution of observations. Consequently, selecting relevant academic features and evaluating the clustering results are essential steps in producing meaningful and interpretable groupings of student performance [12] [13].

2.3 Clustering Academic Achievement in Elementary Education

Clustering academic achievement provides an analytical approach to understanding differences in student performance based on patterns within academic data. At the elementary school level, this approach can help educators identify groups of students who demonstrate

similar achievement characteristics across multiple subjects. Such information can complement conventional assessment practices by providing a broader perspective on academic performance rather than relying exclusively on individual grades or overall averages. In the context of Al Muttaqien Private Elementary School, K-Means clustering can be used to analyze students' academic records and identify groups that reflect different achievement profiles. The resulting information may support teachers in planning enrichment activities for students with stronger academic performance and providing additional learning assistance to students who demonstrate lower achievement levels. To ensure that the groups are meaningful, the interpretation of each cluster should be based on the actual distribution of academic scores, while evaluation methods such as the Silhouette Score may be used to assess clustering quality. Thus, clustering can serve as a supporting analytical tool for educational evaluation and data-driven learning management without replacing teachers' professional judgment [14].

Research Methodology

The research methodology is designed to provide a systematic framework for applying the K-Means Clustering algorithm to group academic achievement levels among students at Al Muttaqien Private Elementary School. This study adopts a quantitative approach by utilizing students' academic records as the primary data source for identifying similarities and differences in academic performance. The research focuses on the application of unsupervised machine learning, in which student groups are formed based on the characteristics of the available academic data rather than predefined achievement labels. The methodological framework consists of several interconnected stages, beginning with data collection and continuing through data preprocessing, feature selection, clustering implementation, cluster evaluation, and interpretation of the results[15] .

The research process begins with collecting academic records containing relevant numerical information, such as subject scores, semester grades, and average academic performance. The collected data are subsequently examined to ensure their completeness, consistency, and suitability for computational analysis. During preprocessing, missing values, duplicate records, and inconsistent data entries are addressed according to appropriate data-cleaning procedures. Relevant academic features are then selected to represent student performance, and standardization is applied when necessary to ensure that differences in measurement scales do not disproportionately influence the clustering process. These preparations are important because the quality and characteristics of the input data directly affect the formation and interpretation of the resulting clusters[16] .

The next stage involves implementing K-Means Clustering to identify groups of students with similar academic performance characteristics. The number of clusters is determined through an appropriate selection procedure, which may include the Elbow Method, while the Silhouette Score can be used to assess the quality of the resulting grouping. After the clustering process, the characteristics of each cluster are examined using descriptive statistics, including average academic scores and differences across subjects. The final results are interpreted to identify academic achievement profiles that may support teachers and school administrators in evaluating student performance and planning appropriate learning strategies. The overall methodology is organized into six stages to ensure that the research process is transparent, reproducible, and aligned with the objective of academic achievement clustering[17] .



Figure 1. Research workflow for K-Means-based academic achievement clustering.

1. Data Collection
   The first stage involves collecting academic data from Al Muttaqien Private Elementary School. The dataset consists of student academic records that represent learning achievement across relevant subjects. Potential variables include mathematics scores, language scores, science scores, semester grades, and average academic performance, depending on data availability. Each record represents an individual student, while the selected numerical variables describe the student's academic characteristics. Data collection must follow applicable school permissions and student data protection requirements. The resulting dataset serves as the primary input for subsequent preprocessing and clustering procedures.
2. Data Preprocessing
   Data preprocessing is performed to improve the quality and consistency of the collected academic records before applying the clustering algorithm. This stage includes checking for missing values, duplicate records, inconsistent score formats, and invalid entries. Missing values are handled using an appropriate method based on their extent and characteristics, while duplicate or inconsistent records are verified against the available source data. Records with unresolved data quality problems are treated carefully to avoid introducing misleading information into the analysis. The output of this stage is a cleaned academic dataset suitable for numerical analysis.
3. Feature Selection and Standardization
   Feature selection determines which academic variables are most relevant to the grouping process. Subject scores and semester grades may be selected to represent different dimensions of student achievement. Variables that do not directly represent academic performance should be excluded unless their inclusion has a clear methodological justification. If the selected variables have substantially different scales or distributions, standardization may be applied to prevent features with larger numerical ranges from dominating the distance calculations. The resulting dataset contains the selected and appropriately prepared features required by the K-Means algorithm.
4. K-Means Clustering
   At this stage, the K-Means Clustering algorithm is applied to group students according to similarities in their academic characteristics. The algorithm initially establishes cluster centroids, assigns each student to the nearest centroid, and iteratively updates the centroids based on the assigned observations. This process continues until the assignments or centroid positions stabilize according to the specified convergence criteria. The number of clusters should be investigated using an appropriate method, such as the Elbow Method, rather than selected solely on assumption. The resulting clusters represent groups of students with similar academic profiles and provide the basis for further evaluation.
5. Cluster Evaluation
   Cluster evaluation is conducted to determine whether the resulting groups provide meaningful separation and cohesion. The Silhouette Score can be used to assess how closely each student fits within their assigned cluster compared with neighboring clusters. The Elbow Method may also help identify a reasonable number of clusters by examining changes

in within-cluster variation across different cluster configurations. In addition to these quantitative measures, cluster sizes and the distribution of academic scores are examined to identify potentially unbalanced or difficult-to-interpret groups. Combining these assessments provides a more comprehensive basis for determining whether the clustering results adequately represent differences in student academic performance.

6. Research workflow for K-Means-based academic achievement clustering
   The final stage involves interpreting the characteristics of each cluster based on its academic score distribution, average performance, and differences across selected subjects. Clusters may be described as relatively high-, moderate-, or low-achievement groups when these descriptions are supported by the observed results. Such labels are assigned after examining the cluster profiles; they are not predefined outputs of K-Means. The findings can then be used to provide information for educational evaluation, including identifying students who may benefit from enrichment activities or additional learning support. The interpretation should remain descriptive and supportive, recognizing that academic scores alone do not fully represent students' abilities, learning circumstances, or overall educational development.

Results

The Results section presents the findings of the K-Means Clustering analysis using the academic recap files for Grades 1–6, covering Semesters 1 and 2. The analysis focuses on seven subjects available across all grade levels: Islamic Religious Education (PAI), Pancasila, Indonesian Language, English, Mathematics, Physical Education (PJOK), and Arts and Culture. The results are organized into three tables and four graphical visualizations to describe the dataset, compare academic performance across subjects, examine the resulting student clusters, and evaluate clustering quality.

The academic records were consolidated into a single dataset. The initial extraction identified 315 student-semester records. 4 records with insufficient valid numerical scores were excluded after zero entries requiring verification were treated as missing, leaving 311 records for clustering. The seven common subject scores were used as numerical features and standardized before K-Means was implemented. Student names and personal identifiers were not included in the analytical output.

The analysis examined a three-cluster solution to describe relatively lower, moderate, and higher academic achievement profiles. The groups were interpreted using their observed mean scores rather than predefined labels. Cluster counts from two to six were also compared using the Silhouette Score. The tables and figures below report descriptive findings, cluster distribution, average performance, and validation results. The three-cluster solution is retained for interpretable achievement profiling, although the silhouette comparison is also reported transparently.

4.1 Dataset Composition

Table 1 shows that the final dataset contains 311 usable records across six grade levels and two semesters. The distribution varies by grade because the number of students and the completeness of academic records differ across files. Combining the records provides a broad view of academic performance across the school, while the grade and semester fields remain available for contextual interpretation. Because the same students may appear in both semesters, the unit of analysis is a student-semester record rather than a unique student.

Table 1 Number of usable student-semester records, based on hours and semesters, after going through data quality filtering

Grade	Semester 1 records	Semester 2 records	Total records

Grade 1	30	30	60
Grade 2	30	30	60
Grade 3	27	27	54
Grade 4	24	23	47
Grade 5	23	23	46
Grade 6	22	22	44
Total	156	155	311

4.2 Descriptive Academic Performance

Table 2 indicates that the mean score across all seven subject features and all retained records is 83.66. The subject-level averages are relatively close, although some differences are visible. Such variation indicates that students do not necessarily demonstrate identical performance across subject areas. These descriptive statistics provide context for the clustering process; however, average scores alone do not identify groups of students with similar multidimensional performance patterns.

Table 2 presents the mean scores for the seven common subjects used in clustering.

Subject	Mean score
Pancasila	84.09
PAI	84.06
B. Indonesia	83.93
Matematika	83.58
Seni Budaya	83.41
B. Inggris	83.29
PJOK	83.23
Overall mean across records	83.66

4.3 K-Means Cluster Profiles

Table 3 shows that the three clusters represent distinct average achievement profiles. The relatively lower-achievement cluster has the lowest mean score, the moderate cluster occupies the middle position, and the relatively higher-achievement cluster has the highest mean. Cluster sizes also differ, indicating that the records are not evenly distributed across achievement profiles. These labels are descriptive summaries of cluster centroids and should not be treated as fixed judgments about individual students or as a substitute for teacher assessment.

Table 3. K-Means Cluster Profiles.

Cluster	Interpretation	Records	Share	Mean score	Record mean range
Cluster 1	Relatively lower achievement	82	26.4%	76.51	71.00–79.86
Cluster 3	Moderate achievement	118	37.9%	83.32	80.00–86.14
Cluster 2	Relatively higher achievement	111	35.7%	89.29	86.43–94.29

Figure 2 compares the number of student-semester records assigned to each cluster. The visualization makes it possible to see which achievement profile contains the largest share of records and whether any cluster is comparatively small. The distribution should be interpreted as a pattern in the analyzed academic records, not as a count of unique students, because a student may contribute a record in both semesters.

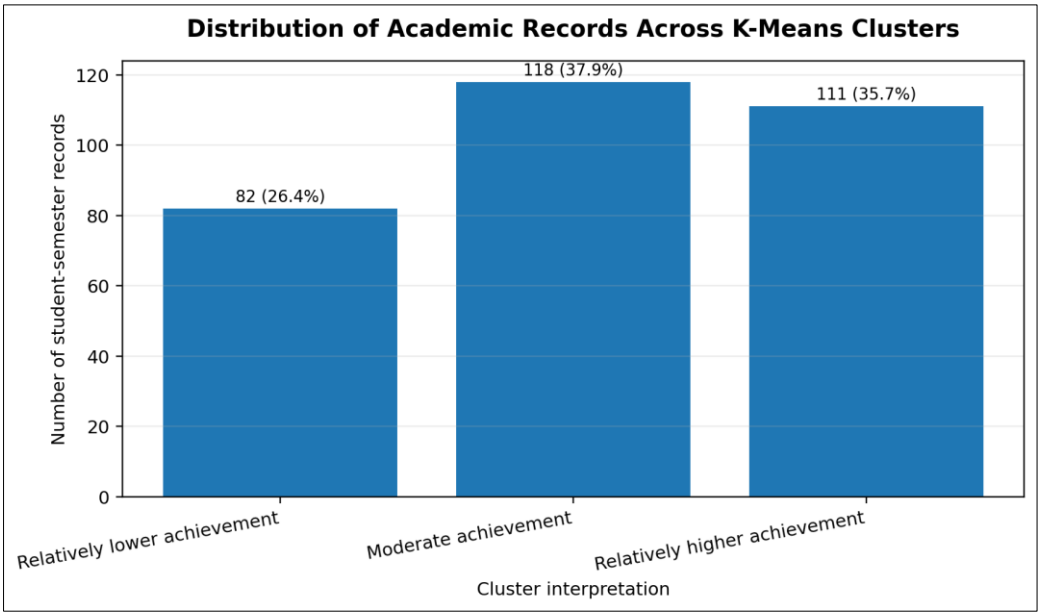


Figure 2. *Distribution of Academic Records Across K-Means Clusters*

Figure 3 compares mean scores across the seven common subjects. The chart helps identify which subjects have relatively higher or lower average scores in the combined dataset. Differences between subjects can inform further investigation by teachers, but the aggregated averages do not establish the causes of those differences and may reflect variations between grade levels and semesters.

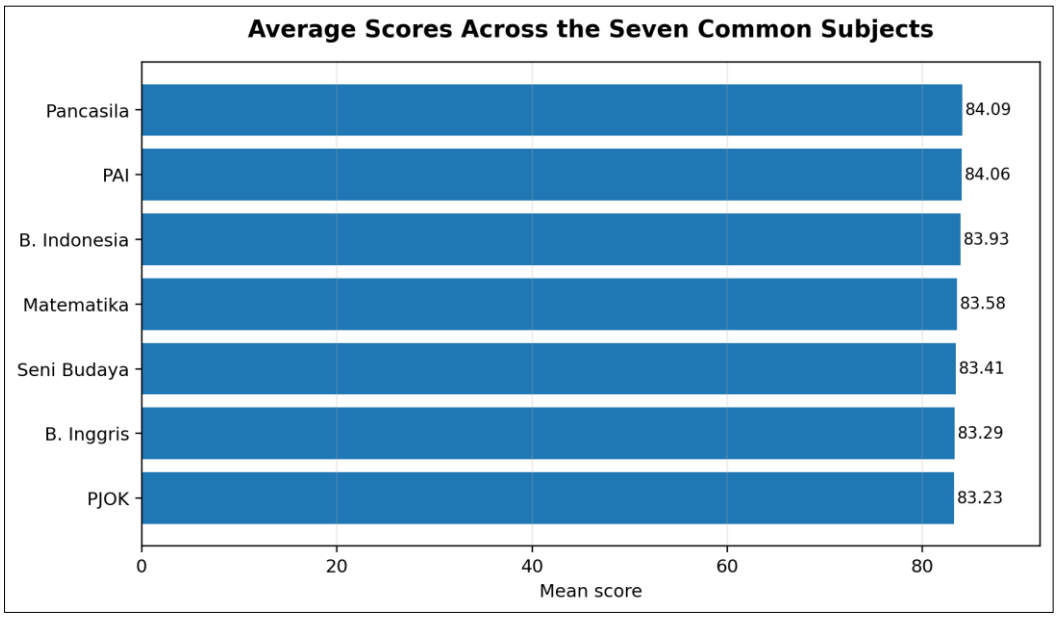


Figure 3. *Average Scores Across the Seven Common Subjects*

Figure 4 illustrates the mean academic score for each cluster, making the relative ordering of the achievement profiles clear. The lower, moderate, and higher labels are based on the observed cluster means. Some overlap in individual record averages may remain, so cluster assignment should be understood as a multidimensional grouping across subject scores rather than a strict threshold classification based only on the average.

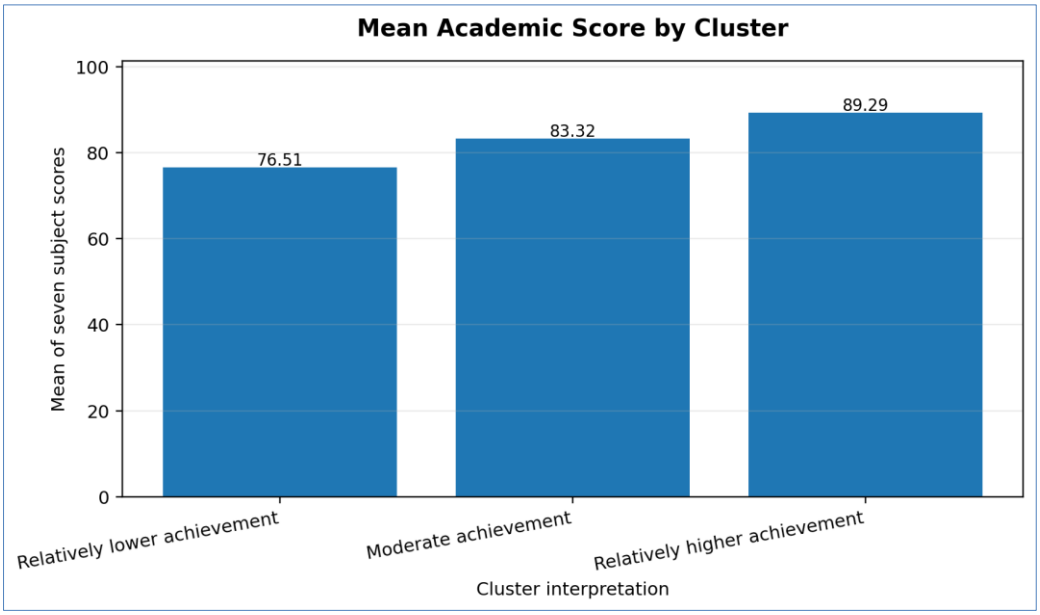


Figure 4. *Mean Academic Score by Cluster*

Figure 5 compares Silhouette Scores for cluster counts from two to six. The two-cluster solution has the highest observed Silhouette Score (0.484), while the three-cluster solution has a score of 0.387. This indicates that two clusters provide stronger separation according to this metric, whereas three clusters offer a more detailed three-level achievement profile. Therefore, the three-cluster solution is useful for descriptive interpretation, but the choice should be justified as a balance between interpretability and validation results rather than claimed as the numerically optimal solution.

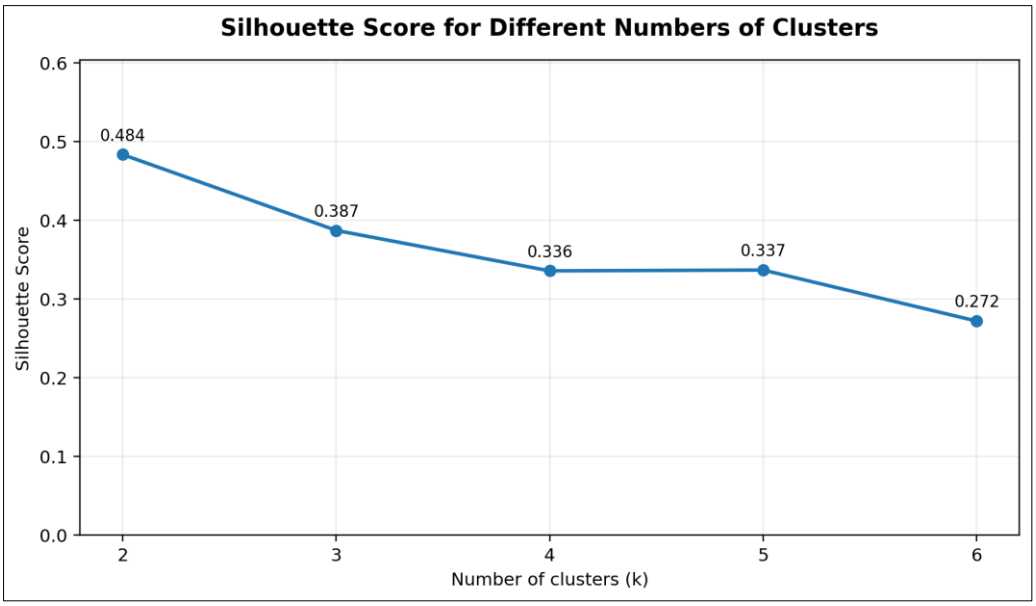


Figure 5. *Silhouette Score for Different Numbers of Clusters*

Discussion

The results demonstrate that the K-Means Clustering algorithm can be applied to group elementary school students according to similarities in their academic performance across seven subject areas. Based on the analysis of 311 student-semester records, the three-cluster solution identified relatively lower, moderate, and higher academic achievement profiles, with mean

scores of 76.51, 83.32, and 89.29, respectively. The moderate-achievement cluster contained the largest proportion of records (37.9%), followed by the relatively higher-achievement cluster (35.7%) and the relatively lower-achievement cluster (26.4%). These findings indicate that academic performance in the analyzed dataset is not homogeneous and that clustering can provide a structured representation of achievement differences. From an educational perspective, this grouping may help teachers identify general learning patterns and develop differentiated instructional strategies. Nevertheless, the cluster labels represent relative academic profiles within the analyzed dataset and should not be interpreted as permanent classifications of individual students.

The clustering evaluation also provides an important consideration regarding model selection. The two-cluster solution achieved a Silhouette Score of 0.484, whereas the three-cluster solution achieved 0.387. This indicates that the two-cluster configuration produced better separation and cohesion according to this evaluation metric, although the three-cluster configuration offers a more detailed distinction among academic achievement profiles. Therefore, selecting three clusters involves balancing quantitative validation with the practical objective of identifying lower, moderate, and higher achievement patterns. These results also highlight the importance of examining academic data across different grade levels and semesters, because differences in curriculum, assessment difficulty, and student development may influence the resulting clusters. In practical application, clustering should be treated as a supporting tool for educational evaluation rather than a replacement for teachers' professional judgment. Further research could compare clustering configurations within individual grade levels and incorporate additional learning indicators to determine whether the identified profiles remain consistent across different educational contexts.

Conclusion

This study applied the K-Means Clustering algorithm to group academic achievement levels among students at Al Muttaqien Private Elementary School using seven academic subject scores. Based on 311 student-semester records, the analysis identified three clusters representing relatively lower, moderate, and higher academic achievement, with average scores of 76.51, 83.32, and 89.29, respectively. The findings demonstrate that K-Means can organize academic records into interpretable groups and provide a structured overview of variations in student performance. These results may assist teachers and school administrators in understanding academic achievement patterns and identifying areas that require further educational attention.

The clustering evaluation revealed that the two-cluster configuration achieved a higher Silhouette Score of 0.484 than the three-cluster configuration, which obtained 0.387. Although the three-cluster solution provides a more detailed interpretation of achievement levels, the results indicate that the number of clusters should be selected by considering both validation metrics and educational objectives. The resulting clusters can support differentiated learning strategies, including enrichment activities and additional academic assistance, but should not be used as the sole basis for evaluating individual students. Future research should examine clustering performance within individual grade levels, incorporate additional learning-related variables, and evaluate the consistency of the identified groups across semesters. These developments could strengthen the practical application of machine learning in educational data analysis and support more informed, data-driven learning management.

References

[1] H. W. Dhany, F. Izhari, and S. Komputer, "Jurnal ICT : Information and Communication Technologies, 15 (2) (2024) 48-54 Prediction analysis condition animal use algorithm (SVM+KNN)," 2024. [Online]. Available: www.ejournal.marqchainstitute.or.id/index.php/JICT

[2] Z. Sitorus, D. Prayogi, M. A. Rizko, A. G. Suteja, and M. R. Harahap, "Implementation of the Insertion Sort Algorithm to Sort Positive Integers in Ascending Order Using Flowgorithm," *Journal of Information Technology, computer science and Electrical Engineering (JITCSE)*, vol. 1, no. 3, pp. 323–328, 2024, doi: 10.30596/jitcse.

[3] K. Khairul, A. H. Nasyuha, A. Ikhwan, M. H. Aly, and A. Ahyanuardi, "Implementation of Multiple Linear Regression to Estimate Profit on Sales of Screen Printing Equipment," *JURNAL INFOTEL*, vol. 15, no. 2, pp. 55–61, Jun. 2023, doi: 10.20895/infotel.v15i2.934.

[4] M. S. Novelan and S. Aryza, "OPTIMIZATION CVRP WITH MACHINE LEARNING FOR IMPROVED CLASSIFICATION OF IMBALANCED DATA FOOD DISTRIBUTION," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 4, pp. 917–925, Jun. 2025, doi: 10.33480/jitk.v10i4.6467.

[5] R. Farta Wijaya, F. Kurniawan, R. R. Putrai, A. Alvin, F. Sains, and D. Teknologi, "OPTIMIZING THE MANAGEMENT OF VILLAGE ACTIVITIES THROUGH INFORMATION SYSTEMS FOR TRANSPARENCY AND ACCOUNTABILITY IN PERTUMBUKAN VILLAGE, WAMPU DISTRICT," 2024.

[6] M. M. Maulana, A. I. Suroso, Y. Nurhadryani, and K. B. Seminar, "The Smart Governance Framework and Enterprise System's Capability for Improving Bio-Business Licensing Services," *Informatics*, vol. 10, no. 2, Jun. 2023, doi: 10.3390/informatics10020053.

[7] A. Putera Utama Siahaan, N. Azizah Harahap, R. Yuni Simanullang, and P. Wanny, "Analysis of Inpatient Data Using Cluster Analysis on Simulation Dataset," *Bulletin of Information Technology (BIT)*, vol. 6, no. 1, pp. 33–39, 2025, doi: 10.47065/bit.v5i2.1830.

[8] M. S. Novelan, S. Efendi, P. Sihombing, and H. Mawengkang, "VEHICLE ROUTING PROBLEM OPTIMIZATION WITH MACHINE LEARNING IN IMBALANCED CLASSIFICATION VEHICLE ROUTE DATA," *Eastern-European Journal of Enterprise Technologies*, vol. 5, no. 3(125), pp. 49–56, 2023, doi: 10.15587/1729-4061.2023.288280.

[9] M. Iqbal and S. Efendi, "Data-Driven Approach for Credit Risk Analysis Using C4.5 Algorithm," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 14, no. 1, pp. 11–20, May 2023, doi: 10.21512/comtech.v14i1.8243.

[10] W. Cetera, W. Gogołek, A. Żołnierski, and D. Jaruga, "Potential for the use of large unstructured data resources by public innovation support institutions," *J. Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00610-6.

[11] G. Rajeshkumar *et al.*, "Smart office automation via faster R-CNN based face recognition and internet of things," *Measurement: Sensors*, vol. 27, Jun. 2023, doi: 10.1016/j.measen.2023.100719.

[12] M. Overton, S. Larson, L. J. Carlson, and S. Kleinschmit, "Public data primacy: the changing landscape of public service delivery as big data gets bigger," *Global Public Policy and Governance*, vol. 2, no. 4, pp. 381–399, Dec. 2022, doi: 10.1007/s43508-022-00052-z.

[13] J. Ilmiah, E. & Erudisi, A. Zahara, M. I. Kabullah, and R. E. Putera, "JIEE: Effectiveness of the OSSRBA (Online Single Submission Risk Based Approach)

System in Business Licensing Services in Payakumbuh City DPMPTSP," vol. 3, pp. 22–40, 2023.

[14] W. Jin, "Challenges and Innovative Countermeasures Faced by Public Administration in the Context of Big Data and Internet of Things," *Math. Probl. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/8949365.

[15] I. Kadek, A. Sugianta, N. Kadek, and W. Patrianingsih, "Evaluasi Naïve Bayes dan K-Nearest Neighbor dalam Klasifikasi Sentimen Ulasan Produk Skincare MSGLOW di Tokopedia," *Jurnal ProTekInfo* |, vol. 12, no. 1, 2025.

[16] N. Wulandari, Y. Cahyana, and H. Hikmayanti Handayani, "Sentiment Analysis on the Relocation of the National Capital (IKN) on Social Media X Using Naive Bayes and K-Nearest Neighbor (KNN) Methods," 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[17] Y. C. Tapidingan and D. Paseru, "Comparative Analysis of Classification Methods of KNN and Naïve Bayes to Determine Stress Level of Junior High School Students," 2020.