Jurnal Media Teknik Elektro dan Komputer
Vol. 02, No.02 (July - December, 2025)
E-ISSN: 3063-2927

Classification of MAN 2 Labuhanbatu Student Achievement Through Learning Achievement Index Components Using K-Means Clustering

Abdul Rasyid[1], Mhd Ikhsan Rifki[2]

[1]Universitas Islam Negeri Sumatera Utara; abdulrasyidtambunan@gmail.com
[2]Universitas Islam Negeri Sumatera Utara; rifki.mhdikhsan@uinsu.ac.id

ABSTRACT

The effective use of academic data is essential for supporting evidence-based decision-making in educational institutions. Educational Data Mining (EDM), particularly through clustering techniques, provides an objective approach to identifying patterns in student academic performance. This study aims to classify students into distinct performance groups as a basis for targeted academic interventions. The K-Means clustering algorithm was applied to examination score data from one subject at MAN 2 Labuhanbatu Utara, consisting of daily assessment, midterm, and final examination scores from 35 eleventh-grade students. The data were preprocessed and normalized using the Min–Max scaling method, while the optimal number of clusters was determined using the Elbow Method based on the Within-Cluster Sum of Squares (WCSS) indicator. The analysis resulted in three clearly defined clusters representing low, moderate, and high academic performance levels. The centroid values and internal validation results indicate that this three-cluster solution adequately reflects the underlying structure of the data. These findings demonstrate the effectiveness of the K-Means algorithm in grouping students without prior labels and provide a data-driven foundation for differentiated instruction, remedial learning, and academic guidance in the school education context.

Keywords: K-Means Clustering, Educational Data Mining, Learning Group, Student Performance

Corresponding Author:
Abdul Rasyid
Universitas Islam Negeri Sumatera Utara; abdulrasyidtambunan@gmail.com

1. INTRODUCTION

Academic grades are the primary indicator used by schools to represent student learning outcomes. Daily grades, midterm exam grades, and final exam grades are routinely collected as a basis for evaluating learning outcomes, determining graduation, and making academic decisions. However, previous studies have indicated that academic grades are commonly used for summative and administrative purposes, while their potential for deeper analytical exploration to uncover patterns and variations in students' learning performance is often underutilized, despite evidence that such data can provide valuable insights when analyzed systematically (Alamgir et al., 2024). In fact, when processed systematically, academic grade data has great potential to reveal the structure of student performance objectively and based on data. This condition indicates the need for an analytical approach that is capable of processing academic grades not only as evaluation results, but also as a source of information to support more accurate and targeted learning decisions.

In the field of research, clustering techniques are one of the most widely used approaches to group students based on their academic characteristics. The K-Means method stands out in particular due to its algorithmic efficiency and its ability to process large amounts of numerical data. K-Means is capable of identifying high-achieving students through grouping based on aspects of attitude, knowledge, and skills (Apriandi et al., 2024). Similar findings also indicate that this method can be used to evaluate learning outcomes by grouping students based on certain academic indicators (Hendrastuty, 2024). Furthermore, research by Habibie and Rakhmat utilized K-Means to classify the influence of additional activities, such as tutoring, on student learning achievement (Habibie & R, 2024). Research conducted by Kurniawan also confirmed that K-Means has significant exploratory capabilities in mapping student academic patterns as a basis for more adaptive learning decision-making (Kurniawan & Ferdiansyah, 2023)

Although these studies prove the success of the K-Means algorithm in educational data analysis, most studies still use multidimensional variables, including non-academic aspects such as attendance, class participation, and interest in learning. This approach provides a broad perspective but does not specifically place test scores as the main indicator of student academic performance(Nurahman et al., 2022). In addition, the studies conducted by Darsono and Andrianti focus more on recommendations for specialization or clustering of educational institutions based on facilities and resources, rather than on mapping student learning groups in the same class or educational unit (Darsono & Andrianti, 2022). These differences in focus, context, and variables indicate a research gap, namely the lack of studies applying K-Means to group students based on exam score patterns in the context of madrasahs, particularly at MAN 2 Labuhanbatu Utara.

A number of studies published between 2023 and 2025 have reported that the K-Means clustering algorithm is effective for analyzing student achievement in educational settings. The use of K-Means to group students based on academic assessment data has been shown to produce meaningful performance classifications when numerical scores are employed as the main input variables (Risal et al., 2024). Similar results were obtained through the application of K-Means to report card data, where the resulting clusters represented distinct levels of academic achievement and supported academic evaluation processes (Adzra et al., 2025).

In contrast, several studies have adopted a multidimensional perspective by integrating non-academic attributes together with academic data. Student clustering based on behavioral indicators such as study habits, attendance, and participation in tutoring activities has been used to analyze learning patterns more comprehensively (Alamgir et al., 2024) n a similar vein, academic performance data have been combined with social factors to obtain broader insights into student group characteristics (Rahma & Ulfah, 2025). Although this approach offers a more holistic understanding of student profiles, it places less emphasis on examination scores as direct measures of academic mastery.

Furthermore, the exploratory capability of K-Means has also been demonstrated in studies utilizing large-scale educational datasets to map general student performance patterns (Alalawi et al., 2023). Despite its effectiveness, such research tends to focus on broad performance grouping and does not specifically prioritize examination scores as the primary basis for clustering within a single subject or institutional context. Collectively, these findings confirm the applicability of K-Means in educational data analysis while simultaneously revealing a research gap concerning the focused use of examination score patterns as pure academic variables, particularly in the context of madrasah education.

This gap is important to fill because daily scores, midterm scores, and final exam scores are objective and direct indicators of student academic achievement without being influenced by other subjective factors. Practically, mapping learning groups based on exam scores can help teachers identify high, medium, and low-performing students, so that differentiated learning strategies can be applied more precisely.

Therefore, this study aims to map student groups based on academic performance patterns derived from English subject assessment scores, specifically daily assessments, midterm examination scores, and final examination scores. These assessment components were selected because they collectively represent students' academic achievement in English as a single subject, allowing performance trends to be identified more accurately and providing meaningful insights to support instructional planning and learning-related decision-making.

2. LITERATURE REVIEW

2.1. Educational Data Mining (EDM)

Educational Data Mining (EDM) is an interdisciplinary field that utilizes data mining and machine learning techniques to extract knowledge from educational data. The development of EDM is driven by the increasing use of digital systems in the learning process and academic administration, which generate large amounts of data, such as evaluation scores, learning activities, and student academic records. Through EDM, educational institutions can understand student learning patterns more objectively and support data-based academic decision-making (Yağcı, 2022). Understand student learning patterns more objectively and support data-based academic decision-making.

In the context of secondary education, EDM plays an important role in analyzing student learning outcomes, particularly in identifying patterns of academic achievement that are not easily recognized through conventional evaluation. The use of exam score data as an object of EDM analysis allows teachers and schools to gain insight into the heterogeneity of student abilities and the potential for different learning needs in each group (Romlah, 2023).

Data mining and visualization techniques have been applied to perform longitudinal analysis of students' academic trajectories, enabling the identification of latent learning patterns through systematic data-driven modeling and analytical processes (Maqsood et al., 2023).

2.2. The Concept of Clustering and Unsupervised Learning

Clustering is a fundamental technique in unsupervised learning that partitions data into several groups based on similarity measures without requiring predefined class labels. This approach is particularly effective when the intrinsic data structure is unknown and must be discovered through computational analysis. In educational datasets, clustering is commonly applied to model variations in students' academic performance and to identify groups with similar learning characteristics. From an algorithmic perspective, distance-based clustering methods particularly the K-Means algorithm exhibit strong exploratory capability by iteratively optimizing centroid positions and minimizing intra-cluster variance, thereby revealing latent structures in academic data and supporting evidence-based decision-making in educational systems (Kurniawan & Ferdiansyah, 2023).

Unlike supervised learning methods, clustering does not require labeled training data, making it more flexible in exploring dynamic academic data. This approach is particularly relevant for exam score analysis because it can reveal student performance structures based on objective, data-driven similarities in academic achievement.

2.3. K-Means Clustering Method

K-Means Clustering is a partition-based clustering algorithm that groups data into K clusters based on the closest distance to the cluster center (centroid). This algorithm works iteratively by minimizing intra-cluster distances until a stable group division is formed. K-Means is widely used because of its simplicity of implementation, computational efficiency, and ability to handle large amounts of numerical data (Hendrastuty, 2024).

In academic data analysis, K-Means is considered appropriate because exam score variables are numerical and structured. Although it has limitations such as sensitivity to the selection of the number of clusters and the presence of outliers, these weaknesses can be minimized through validation techniques such as the Elbow method and Within Cluster Sum of Squares (WCSS) evaluation, so that the clustering results remain relevant and can be interpreted academically.

2.4. The Concept of Academic Value Evaluation

Academic grades are the main indicator in measuring student learning outcomes. Academic evaluation generally includes daily grades, midterm exams, and final exams, which represent the continuous learning process and outcomes. These scores reflect the level of understanding, consistency of learning, and the ability of students to achieve the predetermined learning competencies (Fajar et al., 2024).

In a data-driven approach, academic grade evaluation not only serves as an individual assessment tool but also as a data source for analyzing students' collective learning patterns. Processing academic grades using data mining techniques enables a more objective and systematic evaluation process and supports more adaptive learning decision-making (Prawinugraha et al., 2024).

3. METHODS

3.1. Research Design

This study uses a quantitative approach with an exploratory descriptive design. The quantitative approach was chosen because all data analyzed were numerical values, allowing for objective measurement and structured analysis through unsupervised learning algorithms. An exploratory descriptive design was used to explore natural grouping patterns in student academic data without initial hypotheses, so that the clustering results could reflect student performance characteristics empirically. The selection of this design is in line with the research objective, which is to map learning groups based on similarities in daily scores, midterm scores, and final exam scores within a class. The clustering results are intended to support instructional decisions, particularly in identifying high-, medium-, and low-performing student groups that can be used as a basis for differentiated instructional planning and targeted academic support.

3.2. Data Collection

The data in this study came from official assessment archives managed by English teachers at MAN 2 Labuhanbatu Utara. The data source was a recapitulation of the academic scores of students in class XI IPA 1, which included daily scores, midterm scores, and final exam scores. All data was obtained through direct permission from subject teachers and the school, ensuring its authenticity and administrative validity.

The data collection technique was carried out using the documentation method, which is collecting data on grades that have been determined and used in the learning evaluation process. This method was chosen because it is suitable for obtaining available quantitative data and supports the needs of K-Means algorithm-based analysis. The data obtained was then arranged in tabular format and normalized to ensure compatibility with the clustering analysis procedure.

This research focused on eleventh-grade students at MAN 2 Labuhanbatu Utara during the ongoing academic year. Participants were selected through purposive sampling by including only students from class XI IPA 1 whose academic records were complete, covering daily assessment results, midterm examination scores, and final examination scores. Based on these criteria, 35 students were deemed eligible and subsequently analyzed. Restricting the sample to a single class was intended to preserve uniform learning conditions, allowing the clustering outcomes to represent stable and comparable academic characteristics. The data were sourced from official academic documentation and were collected with formal approval from the relevant subject teachers and the school administration.

This study used three quantitative variables that represent students' academic performance in English. All variables are in the form of numerical data and are treated as input features in the clustering process using the K-Means algorithm. The variables used include:

1. Daily Grades, reflecting students' learning achievements in routine learning activities, including assignments, quizzes, and other formative assessments.
2. Midterm Exam Scores (UTS), which describe students' ability to master the material in the middle of the semester and serve as an indicator of medium-term academic performance.
3. Final Semester Exam (FSE) scores, indicating students' final mastery of the entire semester's material and serving as the main indicator of academic achievement.

The variables were selected as quantitative cognitive indicators and treated as continuous numerical features in the clustering process.

Table 1. Dataset Columns

No	Student Name	Daily scores	Midterm Exam Scores	Final Semester Exam Scores
1.	Ahmad Rayhan Ardani Hrp	81	76	74
2.	Aidil Anwar Situmorang	71	74	70
3.	Ailisya Khairani Pane	96	94	91
…	…	…	…	…
35.	Zahra Fariza	54	48	52

3.3. Data Analysis Techniques

Data processing in this study was carried out using the Python programming language due to its flexibility in data analysis and the availability of libraries that support the implementation of machine learning methods. The entire data processing and analysis process was carried out through the Google Colaboratory (Google Colab) platform, which was chosen because it is stable, easily accessible, and allows Python code to be executed without the need for local software installation.

1. Some of the Python libraries used in this study include:
2. Pandas, used to read, manage, and clean student data in the form of data frames.
3. NumPy, used to support numerical computation operations in the data processing process.
4. Scikit-learn (sklearn), used as the main library in the application of the K-Means Clustering algorithm, the data normalization process using the Min–Max Scaling method, and the calculation of the Within-Cluster Sum of Squares (WCSS) value.
5. Matplotlib, used to generate data visualizations, particularly Elbow Method graphs and the presentation of clustering analysis results

The integration of these platforms and libraries facilitates the data analysis process to be carried out in a more efficient and standardized manner. Moreover, this approach ensures that the procedures can be easily reproduced by other researchers, promoting consistency and reliability in studies that follow the same methodological framework.

3.4. Data Preprocessing

The data preprocessing stage was carried out to ensure that the students' academic data was ready to be used in the clustering process. The data used in this study consisted of daily scores, midterm exam scores, and final exam scores. Because these three variables had the same scale but potentially different value ranges, data normalization was performed to equalize the scale between variables.

Data normalization was performed using the Min–Max Scaling method by converting the values into a range of 0 to 1. This method aims to maintain the proportion of differences between student scores while ensuring that all variables are on the same scale in the Euclidean distance calculation in the K-Means algorithm. In addition, the use of Min–Max Scaling allows the clustering results to be interpreted directly in the context of student academic performance. Mathematically, the Min–Max normalization process is formulated as follows (Yağcı, 2022):

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Description :
x : Original student value
$xmin$: Minimum value for each variable
$xmax$: Maximum value for each variable
x' : Normalized value.

The dataset, after normalization, contains continuous numerical values (floating-point numbers), which is the standard format required for calculating Euclidean distances in the K-Means algorithm. This ensures that the clustering process can proceed effectively. Subsequently, the normalized values serve as the input for performing clustering with the K-Means method.

3.5. Data Normalization

Data normalization is performed to equalize the scale between variable values before the clustering process using the K-Means algorithm. This step is necessary because the K-Means algorithm is distance-based, so differences in the value ranges between variables can affect the clustering results if not handled properly. The daily value, midterm exam value, and final exam value variables have different ranges and distributions, so a transformation is needed so that each variable contributes equally to the analysis process.

The normalization method used in this study is Min–Max Normalization, which transforms the original values into a range of 0 to 1. This normalization is done by subtracting the minimum value and dividing it by the difference between the maximum and minimum values for each variable. With this approach, all numerical data is on the same scale without changing the proportion of the original value distribution. The normalization results are used as input in the K-Means Clustering modeling stage. The application of normalization is expected to improve the accuracy of clustering and ensure that cluster formation more objectively represents students' academic performance patterns, rather than being influenced by the dominance of certain variables.

3.6. K-Means Process

Once the optimal number of clusters has been determined, the K-Means algorithm is employed to organize students according to the similarities in their academic scores. This procedure involves setting the initial centroids, computing the Euclidean distances, assigning data points to clusters, and updating the centroids iteratively until the algorithm converges. At the end of this process, the final centroid for each cluster is established, providing a foundation for analyzing and interpreting patterns in student performance.

3.7. Visualization and Interpretation

The clustering results are visualized using a two-dimensional graph (scatter plot) to display the distribution of students based on the grouping results using the K-Means algorithm. Each cluster is displayed in a different color to facilitate visual observation of the grouping patterns of student scores. The results are interpreted by analyzing the characteristics of each cluster through the centroid values and the distribution of cluster member data. Based on this analysis, the clusters are then interpreted as groups of students with high, medium, and low academic performance levels.

The findings from this analysis serve as a foundation for formulating academic guidance for subject teachers, especially in creating learning strategies that are better aligned with the individual abilities and requirements of each student group.

3.8. Research Flow

Figure 1 illustrates the workflow of the K-Means algorithm applied to academic score clustering. The process starts with the initialization of centroids, followed by the allocation of students into clusters based on similarity, and continues with iterative centroid updates until the algorithm reaches convergence. This visual representation highlights how student data are systematically grouped according to performance patterns. By presenting each step clearly, the diagram helps readers comprehend the underlying logic and sequence of operations in the clustering process, which serves as the foundation for subsequent interpretation and analysis of the results.

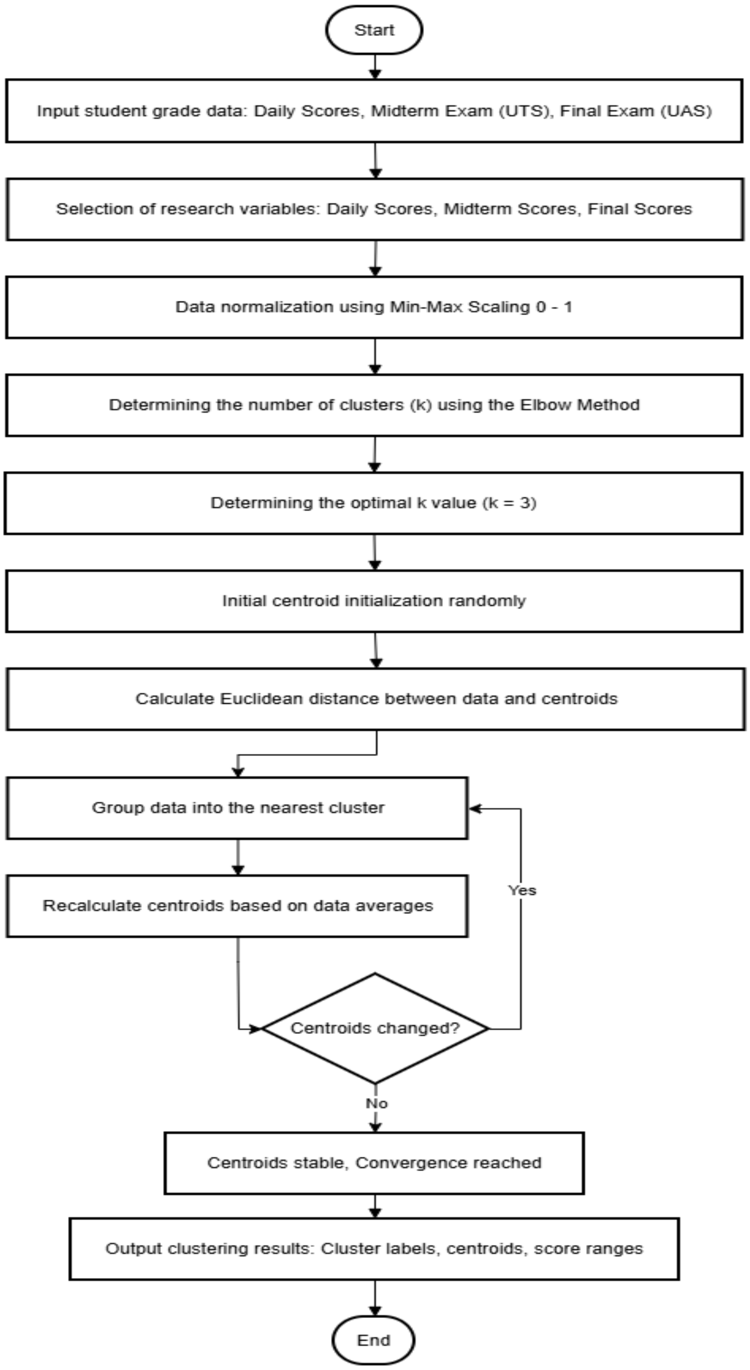

Figure 1. K-Means Algorithm Process Flow in Academic Value Clustering

Based on the flowchart in Figure 3.1, the research began with the collection of student academic score data consisting of daily scores, midterm exam scores, and final exam scores. The data obtained then underwent a preprocessing stage to ensure the completeness and suitability of the data format before further analysis. The next stage is data normalization using the Min–Max Scaling method. This normalization aims to equalize the scale between variables so that there is no domination of one variable in the Euclidean distance calculation in the K-Means algorithm. After the data is on a uniform scale, the number of clusters is determined using the Elbow method by analyzing changes in the Within-Cluster Sum of Squares (WCSS) value against the variation in the number of clusters.

The number of clusters obtained from the Elbow method is then used as a parameter in the clustering process using the K-Means algorithm. The clustering process is carried out through the stages of centroid initialization, calculation of Euclidean distances between data and centroids, grouping of data into the nearest clusters, and iterative centroid updates until convergence is achieved. The final stage of the research flow is the presentation of the results of student clustering into several academic performance groups. The clustering results are then analyzed and interpreted in Chapter IV to obtain an overview of the characteristics of student academic performance based on daily scores, midterm exams, and final exams.

4. RESULTS AND DISCUSSION

4.1. Dataset Overview

The dataset used in this study contains 35 student entries from class XI IPA 1 MAN 2 Labuhanbatu Utara. Each entry represents one student with three main variables relevant to academic evaluation, namely Daily Scores, Midterm Scores, and Final Scores. These three variables were chosen because they represent the formative and summative assessments commonly used in secondary schools to comprehensively describe student academic performance.

Tabel 2. Statistical Summary

Statistics	Daily Scores	Midterm Exam Scores	Final Semester Exam scores
Mean	76.14	75.14	75.66
Median	74	74	75
Minimum	54	48	52
Maksimum	99	100	99
Standard Deviation	12.06	12.64	12.29

Table 2 summarizes the descriptive statistical characteristics of students' academic scores by reporting measures such as the mean, median, minimum, maximum, and standard deviation for each assessment category. The mean value describes the overall average of student scores, whereas the median identifies the central value within the score distribution. The minimum and maximum scores indicate the lowest and highest achievements, respectively, thereby illustrating the range of student performance. In addition, the standard deviation provides information on the extent to which individual scores deviate from the average, reflecting the variability of academic achievement among students.

All statistical measures presented in Table 2 were derived from descriptive statistical analysis applied to the complete dataset comprising 35 student observations. The computations were performed independently for daily assessment scores, midterm examination scores, and final semester examination scores using conventional statistical procedures prior to the clustering stage. This preliminary analysis was conducted to examine the distributional characteristics and variability of the data before applying the K-Means clustering algorithm.

Based on the descriptive statistics in Table 2, the daily, midterm, and final exam scores of the students show relatively balanced averages, namely 76.14, 75.14, and 75.66, respectively. However, the wide range of scores indicated by the difference between the minimum and maximum scores (48-100) and the standard deviation in the range of 12 indicate significant variation in academic ability among students in one class.

This high variation in scores indicates that student academic performance is not homogeneous, requiring an analytical approach capable of identifying natural grouping patterns. Therefore, the next step in this study is to determine the optimal number of clusters using the Elbow method, which aims to identify student grouping structures based on similarities in academic performance patterns.

After analyzing the structure of student score variation and determining the optimal number of clusters, the next step is to normalize the data. Normalization is necessary because the variables used in this study, daily scores, midterm scores, and final exam scores have different ranges and distributions, which could potentially affect the distance calculations in the K-Means algorithm.

In this research, the dataset was standardized through min–max normalization to place all variables on a uniform scale, ensuring that each feature contributed fairly to the clustering process. This procedure was implemented to avoid any single variable from disproportionately influencing the distance computations in the K-Means algorithm, thereby enhancing the consistency of the resulting clusters.

The normalized values were derived using the min–max normalization equation outlined in Equation (1) within the Data Preprocessing section. Each original data point was rescaled to a value ranging from 0 to 1 by applying the respective minimum and maximum values of its variable in the formula.

Tabel 3. Normalisasi data min-max

No	Daily Scores	Midterm Exam Scores	Final Semester Exam scores
0	0.600000	0.538462	0.468085
1	0.377778	0.500000	0.382979
2	0.933333	0.884615	0.829787
3	0.911111	0.942308	1.000000
4	0.644444	0.576923	0.680851

Data normalization was performed using the Min–Max Scaling method to convert the daily, midterm, and final exam scores to the 0–1 interval. This method was chosen because it preserves the proportion of the students' original scores so that the clustering results can still be interpreted in the context of academic performance.

The variables used are continuous quantitative variables, so that methodologically they meet the requirements for standardization and clustering using the K-Means algorithm. In addition, these three types of scores are also pedagogically relevant as cognitive indicators that can represent patterns of student achievement in a subject.

4.2. Descriptive Statistics

Descriptive statistics were used as a preliminary stage to describe the fundamental characteristics of students' academic scores prior to the implementation of the clustering method. This stage plays an important role in examining the distribution of scores, measures of central tendency, and the level of variation among students, which are factors that can affect the clustering results produced by the K-Means algorithm.

The analysis shows that the Daily Scores had a mean value of 76.14, with a median of 74, a minimum score of 54, and a maximum score of 99. The broad range of scores indicates considerable variation in students' formative assessment outcomes, suggesting differences in academic ability within the class.

In the case of Midterm Scores, the average score was 75.14 with a median of 74, while the lowest and highest scores were 48 and 100, respectively. Compared to Daily Scores, the midterm assessment exhibited slightly higher variability, which reflects differences in students' capacity to understand and retain learning materials at the mid-semester stage.

Meanwhile, the Final Exam Scores recorded an average of 75.65 and a median of 75, with scores ranging from 52 to 99. Although the distribution of final scores appears relatively consistent, the observed range still reveals a distinction between students who achieved high performance and those with lower academic outcomes at the end of the semester.

In general, all three assessment components show comparable mean values of approximately 75, accompanied by a moderate to high degree of dispersion. This level of variation indicates that students' academic performance is not uniform and reinforces the relevance of applying the K-Means Clustering method. Through the identification of natural groupings based on score similarities, the clustering process provides a clearer and more systematic representation of students' academic performance levels for subsequent analysis and interpretation.

4.3. Determining the Optimal Number of Clusters

Selecting an appropriate number of clusters is a fundamental step in the implementation of the K-Means algorithm because the chosen value of k determines how well the clustering model captures the inherent structure of the data. In this research, the optimal cluster number was identified using the Elbow Method, a technique that examines variations in the Within-Cluster Sum of Squares (WCSS) as the number of clusters increases. WCSS reflects the degree of cluster cohesion by measuring the sum of squared distances between data points and the centroid of their assigned cluster (Yağcı, 2022):

$$WCSS = \sum_{i=1}^{k} \sum_{x_j \in C_i} ||x_j - \mu_i||^2 \quad (2)$$

where *k* is the number of clusters, *C_i* is the set of data in the i-th cluster, x_j is the j-th data in that cluster, and μ_i is the centroid of the i-th cluster. In this process, WCSS values were calculated for a range of cluster numbers from k = 1 to k = 10 using data that had previously been normalized through the Min–Max scaling approach. The computed WCSS values were subsequently presented in the form of an Elbow graph to observe the trend of WCSS reduction and to identify the point at which further increases in the number of clusters yield diminishing improvements in cluster compactness.

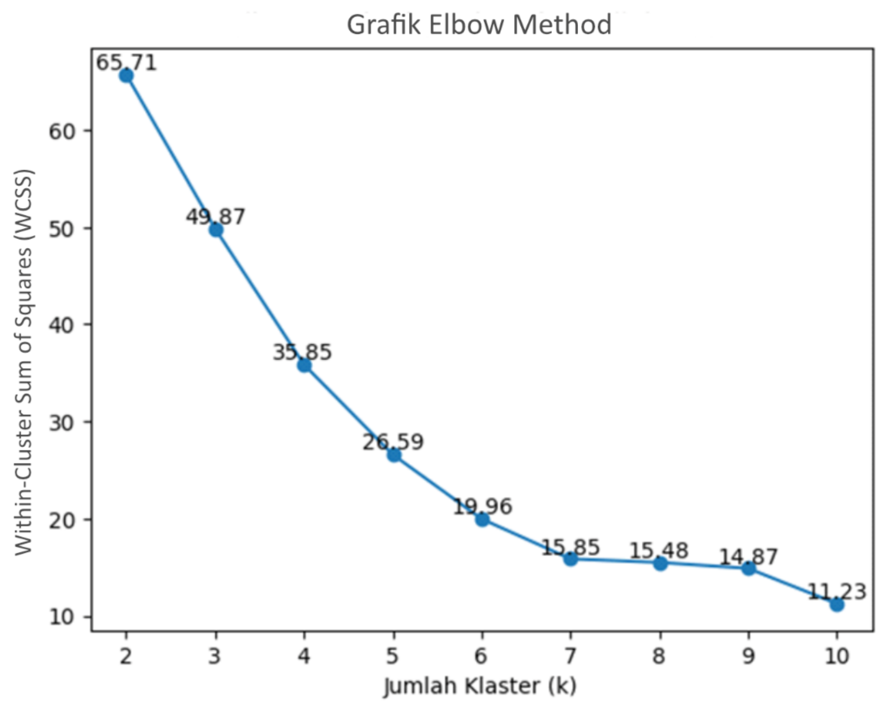

Figure 2. Elbow Method Graph for Determining the Number of Clusters

The Elbow Method visualization shows a sharp decline in WCSS values until $k = 3$. After this point, the rate of decrease slows and the curve tends to level off, suggesting that increasing the number of clusters yields minimal additional improvement in clustering performance. Consequently, three clusters ($k = 3$) are considered optimal in this study, as they achieve an appropriate trade-off between cluster cohesion and ease of interpretation for student academic performance analysis.

4.4. Results and Interpretation of Student Academic Performance Clustering

4.4.1. Formation of Student Academic Performance Clusters

Based on the results of determining the optimal number of clusters using the Elbow Method, the value$k$ = 3 was obtained as the most representative number of clusters in describing the natural structure of student academic score data. Next, the K-Means algorithm was applied to the daily scores, Midterm Exam (UTS) scores, and Final Exam (UAS) scores that had been normalized using the Min– Max Scaling method. This clustering process resulted in three groups of students with different academic performance characteristics, namely high, medium, and low academic performance.

4.4.2. Centroid Value of Each Cluster

Table 4 presents the centroid values of each cluster resulting from K-Means clustering. The centroid values range from 0 to 1 because the data has undergone Min–Max normalization.

Tabel 4. Centroid Value of K-Means Clustering Results (Min–Max)

Cluster Category	Daily Score	Midterm Exam Scores	Final Semester Exam scores
Medium Academic Performance (Cluster 0)	0.561111	0.578526	0.588652
Low Academic Performance (Cluster 1)	0.262500	0.320913	0.269947
High Academic Performance (Cluster 2)	0.898413	0.884615	0.890578

Based on these centroid values, the cluster with the highest centroid value across all variables is categorized as the high academic performance cluster, while the cluster with the lowest centroid value is categorized as the low academic performance cluster. Clusters with centroid values between the two are categorized as the moderate academic performance cluster.

4.4.3. Interpretation of Characteristics of Each Cluster

1. High Academic Performance Cluster

   The high academic performance cluster is distinguished by the largest centroid values across all assessment indicators, with normalized scores of 0.898 for daily evaluations, 0.885 for midterm tests, and 0.891 for final examinations. This pattern suggests that students within this cluster consistently achieve outstanding academic outcomes in all forms of assessment. Further analysis of the original (non-standardized) score data indicates that these students maintain high and stable performance levels, demonstrating strong mastery of learning concepts and sustained academic consistency. As a result, students in this group are suitable candidates for enrichment initiatives or advanced instructional programs designed to further enhance their academic capabilities.

2. Medium Academic Performance Cluster

   The medium academic performance cluster is characterized by centroid values positioned between those of the high and low performance groups, with normalized scores of 0.561 for daily assessments, 0.579 for midterm examinations, and 0.589 for final examinations. This cluster includes students who demonstrate sufficient academic achievement, although their performance tends to vary across different evaluation components. An analysis of the original score data indicates diverse achievement patterns, where students may attain satisfactory results in certain assessments while showing weaker outcomes in others. Consequently, learners in this group would benefit from strengthening core academic concepts and receiving structured instructional support to enhance performance consistency and advance toward higher achievement levels.

3. Low Academic Performance Cluster
   The low academic performance cluster is identified by the smallest centroid values across all evaluated components, with normalized scores of 0.263 for daily assessments, 0.321 for midterm exams, and 0.270 for final exams. This pattern reflects that students in this group generally achieve lower academic outcomes and show less consistency in their learning performance. An examination of the original score distribution further reveals that this cluster occupies a lower score range than the others, indicating challenges in mastering learning materials and meeting expected academic competencies. As the main focus of this research, students in this cluster require prioritized academic support, including remedial learning programs, intensive guidance, and adaptive instructional strategies to narrow achievement gaps and enhance overall academic performance.

4.4.4. Summary of Clustering Results

The use of the K-Means clustering technique resulted in the grouping of students into three clearly differentiated academic performance categories: high, medium, and low. This grouping was derived from the similarity patterns observed in students' daily assessment scores, midterm examination outcomes, and final examination results, offering a systematic representation of variations in academic achievement.

Tabel 5. Number of students per cluster

Cluster	Count
1	**16**
0	**12**
2	**7**

As illustrated in Table 5, the number of students in each cluster varies, reflecting differences in academic performance distribution within the class. In addition, distinctions among the clusters are evident not only from individual assessment components but also from the consistency of students' performance across daily evaluations, midterm examinations, and final examinations, as further demonstrated by the original score ranges presented in Table 6.

Tabel 6. Original Value Range 0-100

Cluster	Daily Score		Midterm Exam Scores		Final Semester Exam scores	
	min	max	min	Max	min	max
0	71	87	73	85	70	85
1	54	74	48	74	52	74
2	90	99	85	100	88	99

An analysis of the centroid values indicates that Cluster 2 comprises students with high academic achievement, Cluster 0 includes students with moderate performance levels, and Cluster 1 consists of students with lower academic achievement. The assignment of individual students to each cluster is detailed in Table 7, which demonstrates the grouping of students based on similarities in their academic performance profiles.

Tabel 7. Student Classification Results Using the K-Means Algorithm

No	Student Name	Cluster Academic	Performance Category
1	Ahmad Rayhan Ardani HRP	0	Medium
2	Aidil Anwar Situmorang	0	Medium
3	Ailisya Khairani Pane	2	High
…	…	…	…
35	Zahra Fariza	1	Low

Of the three clusters identified, particular emphasis is placed on the low academic performance group, as students in this cluster exhibit weaker and less consistent academic results and consequently require prioritized academic support. The high and medium performance clusters are retained as reference groups to facilitate comparison, provide contextual interpretation of performance differences, and support the development of differentiated instructional strategies.

In summary, the results demonstrate that the K-Means algorithm is capable of identifying inherent patterns within student academic data in an objective manner, without the need for predefined class labels. The cluster-based outcomes offer a structured framework for interpreting differences in students' academic performance and provide a foundation for implementing more flexible, data-informed educational approaches. The subsequent section presents the conclusions and recommendations derived from this analysis.

4.4.5. Model Evaluation

Clustering performance was assessed through internal validation measures, specifically the Silhouette Coefficient and the Davies–Bouldin Index, since the K-Means method does not rely on predefined class labels. The evaluation outcomes for the three-cluster solution are summarized in Table X. A Silhouette Coefficient score of 0.35 indicates a moderate degree of intra-cluster similarity and inter-cluster separation, implying that the cluster formation is sufficiently structured, although some overlap between groups remains. Furthermore, the Davies–Bouldin Index value of 0.91 demonstrates an acceptable level of clustering stability. Taken together, these findings indicate that the selected three-cluster configuration is appropriate for identifying patterns in students' academic achievement data.

Tabel 8. Clustering Evaluation Results

Metric	Value
Silhouette Coefficient	0.35
Davies–Bouldin Index	0,91

5. CONCLUSION

This research shows that the K-Means clustering technique is suitable for analyzing student academic achievement at MAN 2 Labuhanbatu Utara by examining patterns derived from daily assessments, midterm examinations, and final examinations. Following data preprocessing and clustering procedures, students were categorized into three distinct achievement groups—high, moderate, and low—with the appropriate number of clusters identified through the Elbow Method.

The clustering outcomes indicate noticeable differences in centroid values across the three groups, reflecting objective variations in students' levels of academic performance. The application of internal clustering validation metrics further confirms that the three-cluster solution adequately represents the underlying structure of the academic data. These results demonstrate the ability of the

K-Means algorithm to uncover inherent groupings within student performance data without the use of predefined labels, aligning with the fundamental principles of unsupervised learning.

Among the resulting clusters, the low academic performance group is highlighted as the central focus of this study, as students within this cluster tend to show lower and less stable achievement across assessment components. Emphasizing this group provides meaningful insights for academic assessment and supports the development of targeted intervention programs, while the high and moderate performance clusters serve as benchmarks for comparison in understanding learning outcome disparities.

In practical terms, the findings of this study offer valuable guidance for educators and schools in designing more adaptive and differentiated instructional approaches, particularly in implementing remedial learning and academic mentoring initiatives for students in the low performance group. From an academic standpoint, this research contributes to the advancement of Educational Data Mining by presenting empirical evidence of clustering-based analysis within the madrasah education context. Future research is encouraged to include additional factors—such as student attendance, learning motivation, socio-economic conditions, and behavioral aspects—and to explore comparisons with alternative clustering algorithms to achieve a more comprehensive depiction of student learning patterns.

REFERENCES

Adzra, S. N., Hasan, F. N., & Kuntoro, A. Y. (2025). Salsa Nabilatul Adzra Penerapan Data Mining dalam Penerapan Data Mining dalam Penilaian Kinerja Akademik Siswa/I SMP YPI Pulogadung dengan Metode K-Means Clustering. *Jurnal Ilmiah informatika* (Vol. 98). http://ejournal.upbatam.ac.id/index.php/jif

Alalawi, S. J., Shaharanee, I. N., & Jamil, J. mohd. (2023). CLUSTERING STUDENT PERFORMANCE DATA USING k-MEANS ALGORITHMS. *Journal of Computational Innovation and Analytics (JCIA)*, *2*(1), 41–55. https://doi.org/10.32890/jcia2023.2.1.3

Alamgir, Z., Akram, H., Karim, S., & Wali, A. (2024a). Enhancing Student Performance Prediction via Educational Data Mining on Academic Data. *Informatics in Education*, *23*(1), 1–24. https://doi.org/10.15388/infedu.2024.04

Alamgir, Z., Akram, H., Karim, S., & Wali, A. (2024). Enhancing Student Performance Prediction via Educational Data Mining on Academic Data. *Informatics in Education*, *23*(1), 1–24. https://doi.org/10.15388/infedu.2024.04

Apriandi, D., Sari, R. M., & Sarif, M. I. (2024). Analisis Clustering Untuk Menentukan Siswa Berprestasi di SMK Swasta TI Panca Dharma Stungkit Menggunakan Metode K-Means. *Jurnal Minfo Polgan*, *13*(1), 1117–1129. https://doi.org/10.33395/jmp.v13i1.13959

Darsono, V., & Andrianti, A. (2022). *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM) Penerapan Data Mining Algoritma K-Means Untuk Rekomendasi Pemilihan Bidang Studi Perguruan Tinggi Pada Siswa SMKN 1 Kota Jambi*. https://ejournal.unama.ac.id/index.php/jakakom

Habibie, A. F., & R, R. K. (2024). Implementasi K-Means Clustering dalam Mengklasifikasi Pengaruh Les Terhadap Prestasi Siswa dengan Metode Elbow. *Journal of Computer System and Informatics (JoSYC)*, *6*(1), 133–147. https://doi.org/10.47065/josyc.v6i1.6201

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, *3*(1), 46–56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Kurniawan, R. P., & Ferdiansyah. (2023). *3 rd Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) 30 Agustus 2023-Jakarta*. *2*(2).

Maqsood, R., Ceravolo, P., Ahmad, M., & Sarfraz, M. S. (2023). Examining students' course trajectories using data mining and visualization approaches. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00423-4

Nurahman, N., Purwanto, A., & Mulyanto, S. (2022). Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(2), 337–350. https://doi.org/10.30812/matrik.v21i2.1411

Rahma, F. A., & Ulfah, S. Z. (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

Risal, A. A. N., Andayani, D. D., Suherman, M. I., & Kaswar, A. B. (2024). *Utilizing the K-Means Clustering Algorithm for Utilizing the K-Means Clustering Algorithm for Analyzing Student Achievement Assessment at SMK Negeri 1 Gowa*.

Romlah, S. (2023). Management of Students Potential Development Using the Data Mining Clustering Method in MAN 2 Malang City. *QALAMUNA: Jurnal Pendidikan, Sosial, Dan Agama*, *15*(1), 95–110. https://doi.org/10.37680/qalamuna.v15i1.2221

Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1). https://doi.org/10.1186/s40561-022-00192-z