Jurnal Esensi Infokom Vol 10 No. 2 Oktober 2026
e-ISSN: 2828-6707
https://esensijournal.com/index.php/infokom
DOI: 10.55886/infokom.v10i2.553

Student Participation Clustering in Extracurricular Activities Using the K-Means Algorithm: A Case Study at Nurul Hidayah Islamic Elementary School

Anis Mirza[1], Melani Dewi Lusita[2], Yudi Irawan Chandra[3]
[1]*Universitas Pamulang*
Jl. Surya Kencana No.1, Pamulang, Kota Tangerang Selatan, Banten, Indonesia 15417
[2,3]*STMIK Jakarta STI&K*
Jalan BRI No.17 Radio Dalam, Kebayoran Baru, Jakarta Selatan, Indonesia 12140
[1]dosen00289@unpam.ac.id, [2]melanilusita@gmail.com, [3]yirawanc@gmail.com

***Abstract*— Extracurricular activities play an important role in supporting students' academic and non-academic development. However, differences in student involvement levels often make it difficult for schools to identify participation patterns and provide appropriate evaluation strategies. This study aims to analyze and classify student participation patterns in extracurricular activities using the K-Means clustering algorithm at Nurul Hidayah Islamic Elementary School. The research applies a quantitative approach by utilizing student participation data collected from extracurricular activity records. The data were processed through several stages, including data preparation, normalization, determination of the optimal number of clusters, and implementation of the K-Means algorithm to group students based on their participation characteristics. The results of the clustering process provide several student participation groups with different involvement levels, allowing schools to identify active, moderate, and low participation patterns. The findings indicate that the K-Means algorithm can be applied as an analytical approach to support data-driven evaluation of extracurricular activities. This research is expected to assist schools in developing more targeted strategies to increase student engagement and optimize extracurricular program management.**
Keywords— Student Participation, Extracurricular Activities, K-Means Clustering, Data Mining, Elementary School.

Intisari— **Kegiatan ekstrakurikuler memiliki peran penting dalam mendukung perkembangan akademik maupun nonakademik siswa. Namun, perbedaan tingkat keterlibatan siswa sering kali menyebabkan sekolah mengalami kesulitan dalam mengidentifikasi pola partisipasi dan menentukan strategi evaluasi yang tepat. Penelitian ini bertujuan untuk menganalisis dan mengelompokkan pola partisipasi siswa dalam kegiatan ekstrakurikuler menggunakan algoritma K-Means clustering pada SD Islam Nurul Hidayah. Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan data partisipasi siswa yang diperoleh dari catatan kegiatan ekstrakurikuler. Tahapan penelitian meliputi proses persiapan data, normalisasi data, penentuan jumlah klaster optimal, serta penerapan algoritma K-Means untuk mengelompokkan siswa berdasarkan karakteristik partisipasinya. Hasil proses klasterisasi menghasilkan beberapa kelompok tingkat partisipasi siswa dengan karakteristik keterlibatan yang berbeda, sehingga sekolah dapat mengidentifikasi pola siswa dengan tingkat partisipasi tinggi, sedang, dan rendah. Hasil penelitian menunjukkan bahwa algoritma K-Means dapat diterapkan sebagai pendekatan analisis untuk mendukung evaluasi kegiatan ekstrakurikuler berbasis data. Penelitian ini diharapkan dapat membantu pihak sekolah dalam menyusun strategi yang lebih tepat untuk meningkatkan keterlibatan siswa serta mengoptimalkan pengelolaan program ekstrakurikuler.**
Kata kunci— Partisipasi Siswa, Kegiatan Ekstrakurikuler, K-Means Clustering, Data Mining, Sekolah Dasar.

I. INTRODUCTION

Extracurricular activities are an integral component of the educational process that supports students' holistic development beyond academic achievement. Through extracurricular programs, students have opportunities to develop their talents, interests, social skills, discipline, teamwork, leadership, and sense of responsibility [1], [2], [3]. At the elementary school level, these activities contribute significantly to character development because students are introduced to various learning experiences that complement classroom instruction. Therefore, effective management and evaluation of extracurricular activities are necessary to ensure that these programs provide optimal benefits for students.

Student involvement is one of the important factors in determining the success of extracurricular activities. A high level of involvement reflects students' interest, commitment, and willingness to participate in school programs [4], [5]. However, the level of student engagement in extracurricular activities is not always the same. Some students actively participate, consistently attend activities, demonstrate discipline, and contribute positively during the process. In contrast, other students may show limited involvement due to low interest, irregular attendance, or lack of initiative. These differences indicate the need for an objective evaluation approach to identify student involvement patterns [6], [7].

SD Islam Nurul Hidayah, as an elementary education institution, provides various extracurricular activities to support students' academic and personal development. However, the

e-ISSN: 2828-6707
https://esensijournal.com/index.php/infokom
DOI: 10.55886/infokom.v10i2.553

evaluation of student involvement in these activities remains a challenge. The assessment process is generally based on manual observations conducted by extracurricular instructors, which may require considerable time and potentially lead to subjective judgments. In addition, the available student activity records have not been fully utilized as a source of information for supporting school decision-making. Without an appropriate analysis process, schools may experience difficulties in identifying students who require additional guidance or developing strategies to improve extracurricular engagement.

The development of information technology has enabled educational institutions to utilize data analysis approaches for extracting meaningful information from available data. One of the approaches that can be applied is data mining, which is a technique used to discover patterns and knowledge from large datasets. Data mining methods have been widely implemented in education to analyze student characteristics, identify learning patterns, and support decision-making processes [8], [9], [10]. Through data-driven analysis, schools can obtain more objective information compared to conventional evaluation methods.

One data mining technique that is suitable for analyzing student involvement patterns is clustering [11], [12], [13]. Clustering is an unsupervised learning method that groups data objects based on similarities in their characteristics. Among various clustering algorithms, K-Means Clustering is one of the most commonly used methods due to its simplicity, computational efficiency, and ability to divide datasets into specific groups. The algorithm determines clusters based on the distance between data points and cluster centers, allowing similar characteristics to be placed within the same group. Previous studies have demonstrated that K-Means can be applied in educational contexts, such as student classification, learning behavior analysis, and academic performance grouping.

In this study, the K-Means Clustering algorithm is applied to analyze student involvement in extracurricular activities at SD Islam Nurul Hidayah. The analysis is conducted using five main variables that represent student involvement characteristics: attendance, activeness, discipline, initiative, and contribution. These variables are selected because they represent behavioral aspects that reflect students' commitment and engagement during extracurricular activities. The clustering process is expected to produce groups of students with different levels of involvement, namely high, medium, and low engagement categories.

The results of this clustering analysis can provide valuable information for schools in evaluating extracurricular activities and designing more targeted improvement strategies. Students with lower involvement levels can receive appropriate support and guidance, while students with higher involvement can be encouraged to maintain and develop their abilities. Furthermore, the implementation of K-Means Clustering provides an objective approach for grouping students based on measurable indicators rather than relying solely on subjective observations.

Based on these considerations, this study aims to apply the K-Means Clustering algorithm to analyze and classify student involvement patterns in extracurricular activities at SD Islam Nurul Hidayah. The results of this research are expected to support data-driven evaluation and assist schools in improving the effectiveness of extracurricular program management.

II. BACKGROUND / LITERATURE REVIEW

Several previous studies have demonstrated the effectiveness of the K-Means Clustering algorithm in grouping data based on specific characteristics. Mita Amilia et al. applied K-Means Clustering to classify regencies/cities in Indonesia based on poverty levels. The study utilized the Knowledge Discovery in Database (KDD) approach and successfully identified regional clusters with different socioeconomic characteristics. The results indicated that K-Means could support data-driven decision-making by identifying priority areas for policy intervention [14].

Furthermore, Della Tri Cahaya et al. implemented the K-Means Clustering algorithm to classify rice production potential in Pagar Alam City. The study applied the CRISP-DM framework and generated three clusters representing low, medium, and high agricultural potential areas. The findings confirmed that clustering methods can effectively categorize objects based on similarity patterns and provide useful information for evaluation and planning processes [15].

In the educational field, Aditia Yudistira et al. applied K-Means Clustering to group students based on academic achievement, attitude, and discipline indicators. The study demonstrated that clustering techniques can be used to identify student characteristics and support educational evaluation. The optimal number of clusters was determined using the Elbow Method and evaluated using the Silhouette Coefficient, indicating that K-Means was capable of producing meaningful student groupings [16].

Although previous studies have shown the potential of K-Means Clustering in various domains, most educational studies have focused on academic performance indicators, such as grades, discipline, and learning outcomes. Meanwhile, student involvement in extracurricular activities has received less attention, even though extracurricular participation represents an important aspect of students' character development, social skills, responsibility, and talent development. In addition, the evaluation of extracurricular involvement is often still conducted through subjective observations, making it difficult for schools to identify student groups objectively.

Therefore, this study proposes the application of K-Means Clustering to analyze student involvement patterns in extracurricular activities at SD Islam Nurul Hidayah. Unlike previous studies that primarily classify students based on academic performance, this research focuses on non-academic involvement indicators consisting of attendance, activeness, discipline, initiative, and contribution. The novelty of this research lies in applying a data mining approach to categorize

students based on extracurricular involvement levels into high, medium, and low clusters, providing an objective basis for schools to evaluate programs and develop appropriate strategies to improve student engagement

The novelty of this research lies in the application of the K-Means Clustering algorithm to analyze student involvement patterns in extracurricular activities based on non-academic indicators, including attendance, activeness, discipline, initiative, and contribution. Unlike previous studies that primarily focused on clustering students based on academic-related aspects, this study emphasizes the identification of student engagement levels in extracurricular activities. This approach provides a new perspective on utilizing data mining techniques to support the evaluation and improvement of extracurricular program management in elementary schools.

III. RESEARCH METHODOLOGY

This study employs a quantitative research approach combined with a system development method to analyze student involvement patterns in extracurricular activities using the K-Means Clustering algorithm. The research was conducted at SD Islam Nurul Hidayah, focusing on fifth-grade students participating in extracurricular activities during the 2025/2026 academic year. The main objective of this research is to develop a clustering-based system that can classify students according to their level of involvement in extracurricular activities. The research methodology consists of several stages, including data collection, data preprocessing, clustering analysis, system development, and evaluation of the results.

A. Data Collection and Research Variables

The data used in this study were obtained from student extracurricular activity records and assessments conducted by extracurricular supervisors. The collected data represent student involvement characteristics in extracurricular activities. Five main variables were used as clustering attributes, namely attendance, activeness, discipline, initiative, and contribution.

Attendance represents the frequency of student presence during extracurricular activities. Activeness describes students' participation and involvement during activity sessions. Discipline reflects students' compliance with activity rules and responsibilities. Initiative indicates students' willingness to participate independently and provide ideas during activities. Contribution represents students' involvement in supporting and assisting the implementation of extracurricular programs.

The assessment data for activeness, discipline, initiative, and contribution were obtained using a questionnaire based on a Likert scale ranging from 1 to 5. Each variable consisted of several assessment items that generated numerical scores representing student involvement characteristics. These variables were selected because they provide measurable indicators for evaluating student engagement beyond academic performance.

B. Data Preprocessing

Before applying the clustering algorithm, the collected data underwent preprocessing to ensure that the data were suitable for analysis. The preprocessing stage included data cleaning, data transformation, and data normalization. Data cleaning was conducted to identify incomplete or inconsistent data records, while data transformation was performed to convert assessment results into numerical formats that could be processed by the algorithm.

Normalization was applied because the variables used in the clustering process had different measurement ranges. By normalizing the data, each variable could contribute proportionally to the clustering process and prevent certain attributes from dominating the calculation. The processed dataset was then used as input for the K-Means Clustering algorithm.

C. K-Means Clustering Analysis

The clustering process was performed using the K-Means Clustering algorithm. K-Means is an unsupervised learning algorithm that groups data objects based on similarities among their characteristics. The algorithm works by determining initial cluster centers (centroids), calculating the distance between each data point and the centroid, assigning data points to the nearest cluster, and updating the centroid values iteratively until the cluster distribution reaches a stable condition [17], [18], [19].

In this study, students were grouped into three clusters representing different levels of involvement in extracurricular activities: high, medium, and low involvement. The determination of three clusters was based on the school's evaluation requirements to simplify the interpretation of clustering results and support the student guidance process.

The quality of the clustering results was evaluated using clustering evaluation methods to determine whether the generated clusters were appropriate. The evaluation process aimed to identify the optimal clustering configuration and ensure that the resulting groups represented meaningful differences among students.

D. System Development Using Fountain Model

The web-based clustering system developed in this research follows the Fountain Model approach, as presented in Figure 1. The Fountain Model is a software development approach that allows overlapping activities between development stages. Unlike sequential development models, this model provides flexibility because developers can return to previous stages whenever improvements or modifications are required [20].

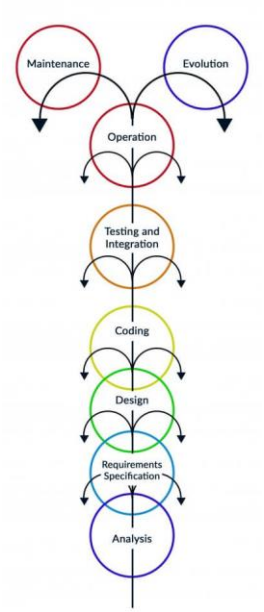


Figure 1. System Development Process Using the Fountain Model

The implementation of the Fountain Model in this study consists of several activities:

A. The first stage is requirement analysis, which focuses on identifying system requirements, user needs, and data requirements. At this stage, the process of evaluating extracurricular involvement at SD Islam Nurul Hidayah was analyzed to determine the necessary system features.
B. The second stage is system design, which includes designing the system architecture, database structure, interface layout, and data processing flow. The design focuses on developing a system capable of managing student data, performing clustering calculations, and displaying clustering results.
C. The third stage is implementation, where the designed system is developed into a web-based application. The system is implemented using Python as the main programming language and integrates the K-Means Clustering algorithm for data analysis. The developed system provides features for data input, preprocessing, clustering execution, and visualization of student groups.
D. The fourth stage is testing and evaluation, which aims to verify system functionality and validate the clustering results. Functional testing is conducted to ensure that each system feature operates according to the requirements. The clustering results are also analyzed to determine the characteristics of each student group.
E. The final stage is refinement and maintenance, where improvements are made based on testing results and user feedback. The flexible nature of the Fountain Model allows continuous adjustments to improve system performance and usability.

Through the combination of the Fountain Model and K-Means Clustering algorithm, this research develops an analytical system capable of identifying student involvement patterns objectively. The resulting clusters are expected to provide useful information for SD Islam Nurul Hidayah in evaluating extracurricular activities and developing strategies to increase student engagement.

IV. RESULTS AND DISCUSSION

The data used in this study consist of student involvement data from 48 fifth-grade students at SD Islam Nurul Hidayah who participated in extracurricular activities during the period of January to April 2026, as presented in Table 1. To maintain student confidentiality and comply with research ethics principles, all personal identities were anonymized by assigning unique codes or IDs (e.g., Student 1, Student 2, and so forth). The data were obtained from two main sources:

A. Attendance Records

Attendance records were used as the data source for the attendance variable. The attendance data were collected from extracurricular attendance documentation recorded by activity supervisors during each meeting. The attendance score was calculated based on the percentage of student attendance compared to the total number of extracurricular sessions conducted during the January–April 2026 period.

B. Extracurricular Supervisor Questionnaire

The questionnaire completed by extracurricular supervisors was used as the data source for the activeness, discipline, initiative, and contribution variables. The assessment was conducted by each supervisor responsible for the extracurricular activities followed by the students. Each supervisor provided evaluations based on direct observation and monitoring of student behavior during extracurricular sessions. Each variable consisted of four statement items measured using a Likert scale to obtain numerical scores representing students' involvement characteristics presented in Table I.

TABLE I. DATA SOURCES

No	Variables	Description	Data Sources	Range of Values
1	Attendance	Percentage of student attendance in extracurricular activities	Attendance Register	0 - 100
2	Active Participation	Level of active student participation during activities	Advisor Questionnaire	4 - 20
3	Discipline	Level of student discipline in following activity rules	Advisor Questionnaire	4 - 20
4	Initiative	Level of student initiative in extracurricular activities	Advisor Questionnaire	4 - 20

5	Contribution	Level of student contribution to extracurricular activities	Advisor Questionnaire	4 - 20

The dataset used in this study consists of 48 student records. Each record represents one student participating in extracurricular activities and contains five attributes used as variables in the clustering process. These variables were selected because they are considered capable of comprehensively representing students' involvement levels during extracurricular activities. The description of the variables used in this study is presented in Table II.

TABLE II. EXAMPLE RESEARCH DATASET (AFTER ANONYMIZATION)

Student	Attendance	Active Participation	Discipline	Initiatives	Contributions
Student 1	100	20	20	19	20
Student 2	87,5	20	20	19	18
Student 3	87,5	12	12	14	12
Student 4	100	16	16	16	16
…	…	…	…	…	…
Student 48	87,5	17	17	15	17

C. System Design Interfaces and Workflow

The system consists of 4 (four) main interface modules that are interconnected to support the execution flow of the K-Means algorithm, as shown in Figure 2.

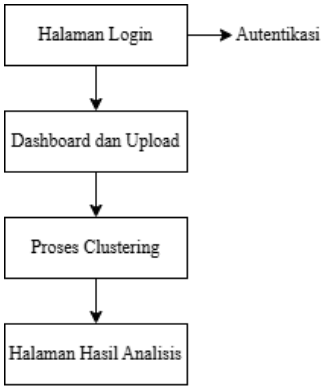


Figure 2. Flowchart of the Interface and System Design Process

D. K-Means Analysis Processing Module

Function: Executes the steps of the K-Means algorithm based on Python libraries in a structured manner:

Data Normalization: Converts the assessment variables to the range [0, 1] using the Min-Max Scaler with the formula:

$$X=(X-Xmin)/(Xmax-Xmin) \quad (1)$$

Calculate Distances and Iterate: Calculates the Euclidean distance from each student's data point to the centroid in each iteration until convergence:

$$X = \frac{X - Xmin}{Xmax - Xmin}$$

$$d\,(x,c) = \sqrt{\sum_{i=1}^{n}(x_i - C_i)^2} \quad (2)$$

Cluster Categories: Automatic grouping of participation levels (High, Medium, Low) based on the average of the centroid values.

DBI Evaluation: Calculation of the Davies-Bouldin Index (DBI) score to determine the quality of cluster separation.

E. Calculation of Proximity Distance

The degree of similarity or distance between each student sample (Student 1 through Student 4) and each centroid (C1, C2, C3) is calculated using the Euclidean distance formula for a 5-dimensional space:

$$d = \sqrt{\sum_{i=1}^{5}(x_i - C_i)^2}$$

$$d = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + (x_3 - c_3)^2 + (x_4 - c_4)^2 + (x_5 - c_5)^2} \quad (3)$$

After the data clustering using the K-Means algorithm (k=3) was completed, the next step was to measure the performance and quality of the resulting clusters. This evaluation aims to measure the degree of cohesion within a cluster as well as the degree of separation between different clusters. In this study, the specific evaluation method applied is the Davies-Bouldin Index (DBI).

F. Davies-Bouldin Index (DBI).

The DBI is calculated based on the ratio of the sum of intra-cluster dispersion to the inter-cluster centroid distance. A DBI value closer to zero (0) indicates that the resulting cluster structure has an optimal level of separation and better clustering performance. The DBI calculation process involves the following five mathematical steps:

Intra-Cluster Dispersion Formula (S_i):

Calculates the average distance of all cluster members from their own cluster centroid to measure the cluster's internal density.

$$S_i = \frac{1}{T_i}\sum_{j=1}^{T_i} d(x_j, c_i) \quad (4)$$

Notes:

1. S_i is the dispersion value or internal density of the i-th cluster
2. T_i is the number of members in the cluster
3. i & d(x_(j,) c_i) is the Euclidean distance of object x_j to centroid c_i.

Inter-Cluster Centroid Distance (M_ij):
Calculates the Euclidean distance between the centroids of two different clusters to measure inter-cluster separation.

$$M_{ij} = d(c_i, c_j) = \sqrt{\sum_{k=1}^{n} (c_{ik} - c_{jk})^2} \quad (5)$$

Notes:
1. M_ij is the Euclidean distance between the centroid of cluster i (c_i) and the centroid of cluster j (c_j).
2. c_ik is the coordinate value of the centroid of cluster i on the kth criterion.
3. c_jk is the coordinate value of the jth cluster's centroid on the kth criterion indicator.
4. n is the total number of criteria/features used.

Inter-Cluster Similarity Ratio (R_ij):
This formula is used to calculate the ratio of the sum of the dispersion of two clusters to the separation distance between their centroids.

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \quad (6)$$

Explanation:
1. R_ij is the similarity ratio between cluster i and cluster j.
2. S_i is the intra-cluster dispersion of cluster i.
3. S_j is the intra-cluster variance of cluster j.
4. M_ij is the distance between the centroids of cluster i and cluster j.

Maximum Similarity Ratio (D_i):
This formula is used to select the highest (worst) similarity ratio value for cluster i when compared to other distinct clusters.

$$D_i = \max_{j \neq i}(R_{ij}) \quad (7)$$

G. Development Environment
The implementation of the student participation clustering analysis system was built using the Python programming language with the Flask framework to develop a web-based application. Development was carried out in the Visual Studio Code Integrated Development Environment (IDE) using several supporting libraries listed in Table III below:

TABLE III. REFERENCES USED IN THE IMPLEMENTATION

No	Library	Function
1	Flask	Web framework for building application interfaces
2	Flask-Login	User authentication management (login/logout)
3	Pandas	Manipulation and processing of tabular data
4	NumPy	Numerical computation and matrix operations
5	Scikit-Learn	Implementation of the K-Means algorithm and DBI evaluation
6	Matplotlib	Graphical visualization (Elbow Method, DBI, comparison)
7	Seaborn	More interactive statistical visualization
8	openpyxl	Reading and writing Excel files

The application flowchart illustrates the sequence of interfaces and processes that users go through to operate the web-based student participation cluster analysis system.

A. Login Form
The Login Form page is the main entry point to the system interface and serves as a security mechanism (authentication) for user access rights. On this page, users are required to enter their registered username and password in the provided fields, as shown in Figure 3.

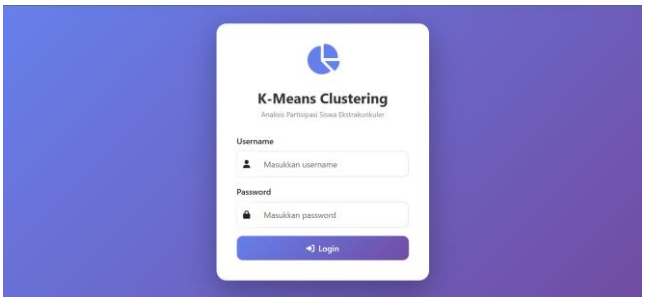


Figure 3. Login Form Page

B. Dashboard
This page serves as a control center for uploading student participation datasets and processing K-means clustering, as shown in Figure 4.

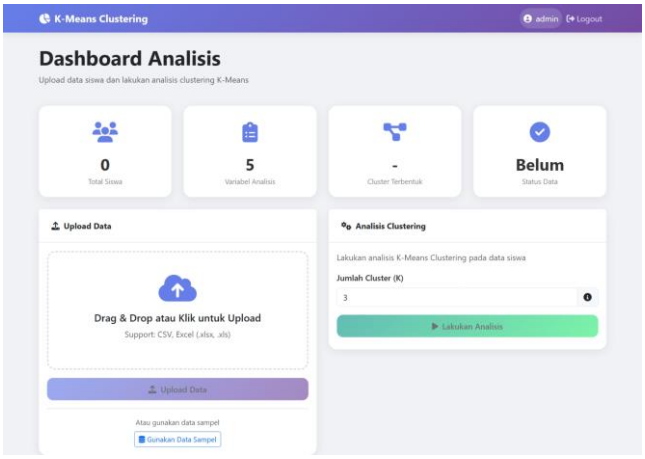


Figure 4. Dashboard Page View

Figure 5 shows the initial screen of the web-based Python application after logging in. The implementation process for this system begins by selecting "Drag & Drop or Click to Upload."

Jurnal Esensi Infokom Vol 10 No. 2 Oktober 2026
e-ISSN: 2828-6707
https://esensijournal.com/index.php/infokom
DOI: 10.55886/infokom.v10i2.553

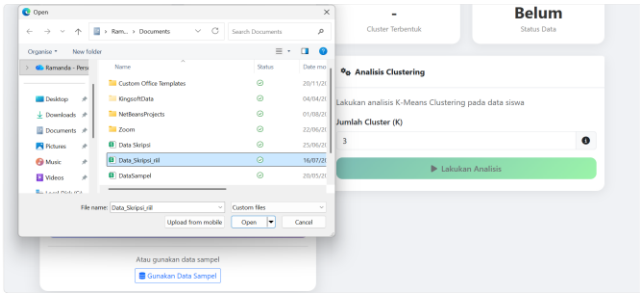


Figure 5. Data import process

Figure 6 illustrates the data import process. This process is carried out by selecting data from the local computer ("My Computer"). The file used is "Data_Skripsi_riil.xlsx," which contains the complete participation data for 48 students in extracurricular activities at SDI Nurul Hidayah. Once uploaded, the application displays all of this data.

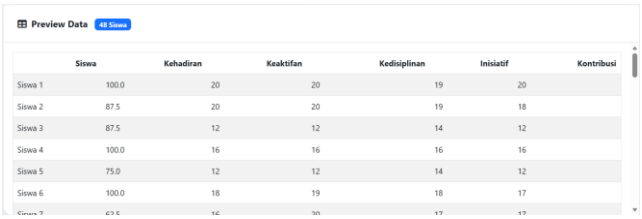


Figure 6. Data display after upload.

C. Clustering Analysis Execution Process

In this section, the user inputs the parameter for the number of clusters (k)—determined during the initial research planning stage—as 3 clusters (k = 3). These represent categories of student participation levels (e.g., High Participation, Moderate Participation, and Low Participation), as shown in Figure 7.

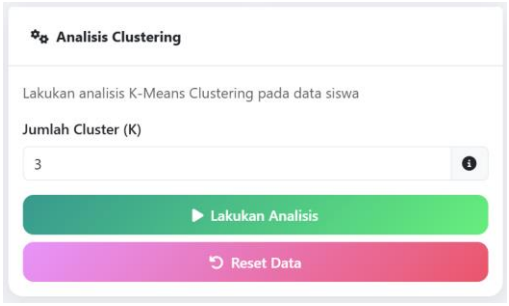


Figure 7. Clustering Analysis Execution Process

D. Visualization of Student Participation Clusters

Figure 8 presents a scatter plot visualization of data from 48 students, grouped into three clusters based on the results of the K-Means clustering algorithm. Although the clustering process utilized all five assessment variables (Attendance, Engagement, Discipline, Initiative, and Contribution), the visualization is projected onto a two-dimensional plane using two primary representative variables—Attendance (%) on the X-axis and Engagement on the Y-axis—to facilitate the interpretation of the data distribution.

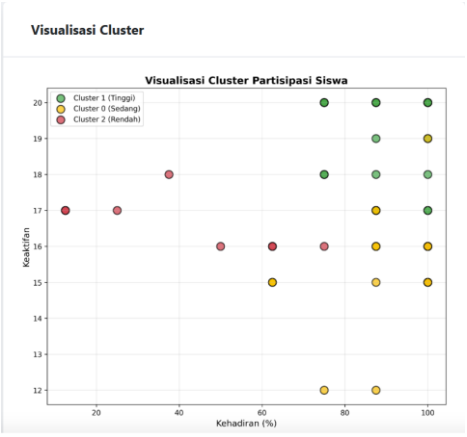


Figure 8. Visualization of Student Participation Clusters

E. Chart Comparing Variable Averages by Cluster

Figure 9 presents a grouped bar chart showing the comparison of mean values for the five assessment variables (Attendance, Participation, Discipline, Initiative, and Contribution) across each of the formed clusters.

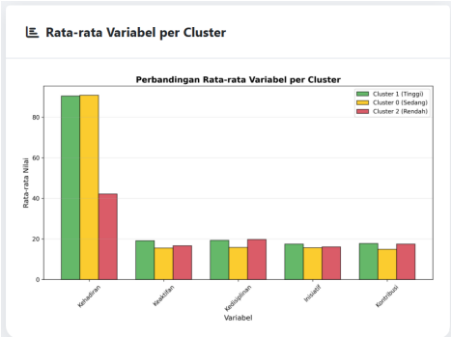


Figure 9. Graph comparing variable averages by cluster.

F. Student Distribution

Figure 10 presents a pie chart illustrating the percentage proportions and distribution of 48 students across three categories of extracurricular activity participation levels, based on the results of K-Means clustering.

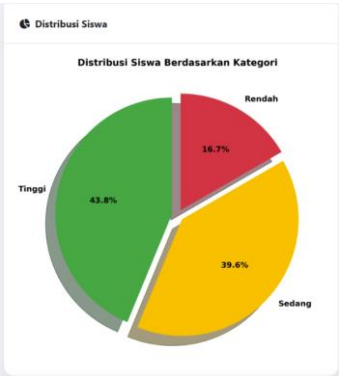


Figure 10. Pie chart of student distribution by category.

Jurnal Esensi Infokom Vol 10 No. 2 Oktober 2026
e-ISSN: 2828-6707
https://esensijournal.com/index.php/infokom
DOI: 10.55886/infokom.v10i2.553

G. Elbow Method

Figure 11 presents the Elbow Method evaluation graph, which measures the decrease in the Sum of Squared Errors (SSE) value against varying numbers of clusters (K), ranging from K = 2 to K = 6. This analysis aims to identify the optimal number of clusters based on the inflection point (the "elbow").

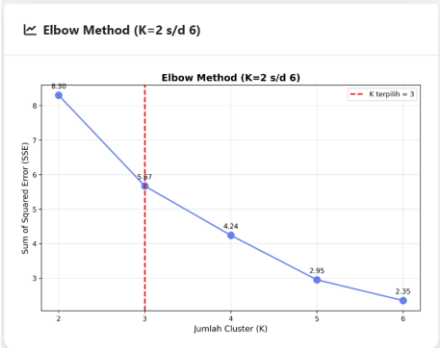


Figure 11. Elbow Method Graph

H. Davies-Bouldin Index

Figure 12 presents a graph showing the cluster quality assessment using the Davies-Bouldin Index (DBI) metric for cluster counts ranging from K = 2 to K = 6. The DBI metric measures intra-cluster density (cohesion) and inter-cluster separation; the lower the DBI value (closer to 0), the better the quality of the resulting cluster separation.

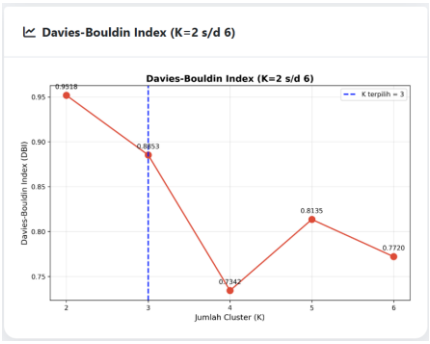


Figure 12. Davies-Bouldin Index Graph

I. Comparison of the Elbow Method and Davies-Bouldin Index

Figure 13 presents a dual-axis plot comparing the trends of the Sum of Squared Error (SSE) values from the Elbow Method and the Davies-Bouldin Index (DBI) scores simultaneously, ranging from K = 2 to K = 6.

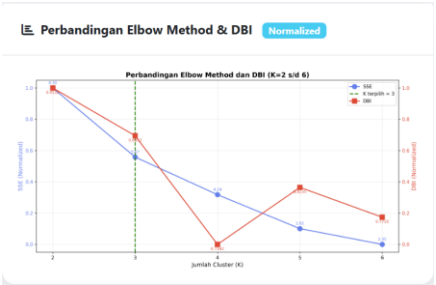


Figure 13. Comparison graph of the Elbow Method and DBI

J. K-Means Iteration Table

Figure 14 displays the iteration process of the K-Means algorithm, which calculates the Euclidean distance from each student's data to each cluster center point (centroid). At each iteration stage, the system calculates the distance between variables and automatically groups students into the cluster with the smallest distance value. This iteration process continues until convergence occurs in the third iteration, where the centroid positions and group membership of the 48 student data are stable and no longer change.

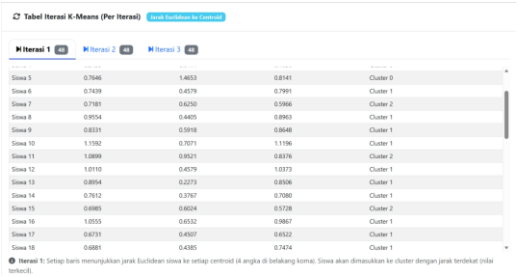


Figure 14. K-Means Iteration Table

K. Final Centroids

Figure 15 presents a table of the final cluster centers (centroids) reverted to their original value scales—prior to normalization. The figure displays a representative profile for each group: Cluster 1 maps to the "High" category, showing the highest average values across nearly all variables (Attendance: 90.48, Engagement: 19.10, Discipline: 19.33, Initiative: 17.48, Contribution: 17.71); Cluster 0 represents the "Medium" category, characterized by high attendance (90.79) but moderate levels for other attributes (Engagement: 15.53 and Contribution: 14.89); and Cluster 2 corresponds to the "Low" category, as it exhibits the lowest average attendance value (42.19).

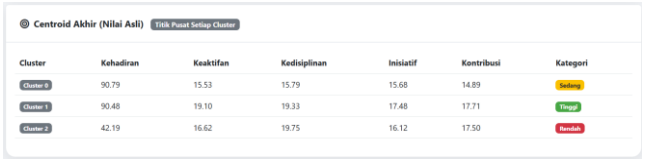


Figure 15. Final Centroids (Original Values)

L. Summary by Cluster

Figure 16 presents a summary of the characteristics and ranking of student participation groups based on the average centroid values for each variable:



Figure 16. Summary by Cluster

M. Student List per Cluster

Figure 17 displays a tabular interface displaying the student list grouped into their respective clusters using the K-Means algorithm. This display features tabbed navigation to filter members per cluster: Cluster 1 (21 students - High Category), Cluster 0 (19 students - Medium Category), and Cluster 2 (8 students - Low Category).

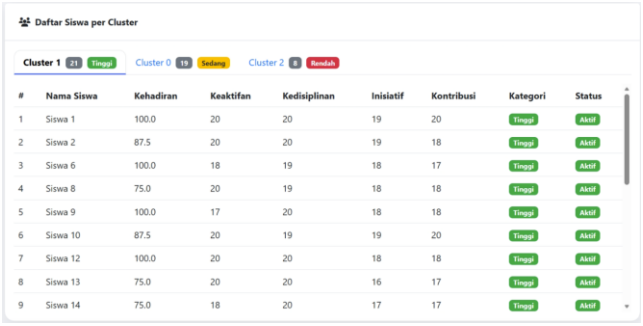


Figure 17. Student List per Cluster

Based on the determination of k = 3 clusters, cluster center (centroid) calculations were performed to understand the characteristics of each student participation group in extracurricular activities at SD Islam Nurul Hidayah. The centroid values resulting from data processing using the K-Means algorithm are presented in Table IV below:

TABLE IV. CLUSTER CENTROIDS FOR K=3 (ORIGINAL SCALE)

Attribute	**Cluster 0 (Medium)**	**Cluster 1 (High)**	**Cluster 2 (Low)**
Attendance (%)	90.79	90.48	42.19
Active Participation	15.53	19.10	16.62
Discipline	15.79	19.33	19.75
Initiative	15.68	17.48	16.12
Contribution	14.89	17.71	17.50
Number of Students	19 Students	21 Students	8 Students

N. Conclusion on School Feedback

Based on the resulting groupings, the system's effectiveness was confirmed and validated with the extracurricular coordinator and the management of SD Islam Nurul Hidayah. The school responded very positively to these grouping results. This K-Means-based mapping of student participation provides a transparent and objective overview of student involvement. The strategic implications of this grouping for SD Islam Nurul Hidayah are as follows:

A. High Cluster (Cluster 1): Students receive recognition (rewards) and are entrusted with leadership roles (e.g., group leaders) in extracurricular activities to mentor their peers.
B. Medium Cluster (Cluster 0): Students are provided with motivational programs—such as more interactive training methods—to encourage them to take initiative and contribute actively.
C. Low Cluster (Cluster 2): Extracurricular supervisors and homeroom teachers employ a personalized approach and conduct attendance evaluations to identify the obstacles affecting student attendance.

V. CONCLUSION

This study successfully implemented the K-Means Clustering algorithm to analyze and classify student involvement patterns in extracurricular activities at SD Islam Nurul Hidayah. The clustering process was conducted using five main variables, namely attendance, activeness, discipline, initiative, and contribution, which represent students' involvement characteristics in extracurricular activities.

The implementation of K-Means Clustering produced three groups of students with different levels of involvement, consisting of high, medium, and low involvement clusters. These clustering results provide an objective overview of student characteristics based on their extracurricular activity involvement. The application of the clustering approach demonstrates that student activity data can be utilized as a source of information to support school evaluation and decision-making processes.

The developed web-based clustering system using the Fountain Model also supports the process of data management, clustering analysis, and visualization of results. By utilizing this system, schools can identify student involvement patterns more efficiently compared to conventional evaluation methods that rely primarily on manual observation. The results of this research indicate that the K-Means Clustering algorithm can be applied as an analytical approach to support the evaluation and improvement of extracurricular program management.

For future research, several improvements can be considered. First, the number of datasets can be expanded by involving students from different grade levels or multiple academic periods to obtain more representative clustering results. Second, future studies may compare the performance of K-Means with other clustering algorithms, such as Hierarchical Clustering, DBSCAN, or Fuzzy C-Means, to determine the most suitable method for analyzing student involvement patterns.

Furthermore, additional evaluation methods, such as Silhouette Coefficient or Davies-Bouldin Index, can be applied to measure cluster quality more comprehensively. Future system development may also include additional features, such as student involvement monitoring dashboards, recommendation features for student development strategies, and integration with school information systems. These improvements are expected to enhance the usefulness of the system as a decision-support tool for managing extracurricular activities.

REFERENCES

[1] D. Sopiandy, S. Hariati, S. Hariyadi, R. R. A. Asis, and S. N. Umar, "Upaya Meningkatkan Partisipasi Aktif Murid melalui Metode Diskusi dan Presentasi pada Mata Pelajaran PPKn di MAS Nurul Yaqin Dawi-Dawi Kabupaten Kolaka," *J. Humanit. Katalisator Perubahan Dan Inov. Pendidik.*, vol. 12, no. 2, pp. 899–914, Jun. 2026, doi: 10.29408/jhm.v12i2.34381.

[2] L. Lerianto and A. N. Chamidah, "Penerapan Model Pembelajaran Kooperatif dengan Pendekatan Jigsaw untuk Meningkatkan Partisipasi Murid Berkesulitan Belajar di dalam Kelas Inklusif," *J. Ragam Pengabdi.*, vol. 3, no. 3, pp. 9440–9451, Sep. 2026, doi: 10.62710/2rysnf37.

[3] G. Y. Prasetiyo *et al.*, "Upaya Meningkatkan Partisipasi Murid Dalam Kegiatan Diskusi Kelompok Melalui Lkpd Terstruktur Berbasis Tutor Sebaya Pada Pembelajaran IPAS Kelas IV," *Pendas J. Ilm. Pendidik. Dasar*, vol. 11, no. 2, pp. 953–967, Jul. 2026, doi: 10.23969/jp.v11i2.56594.

[4] S. Kartika *et al.*, "Gender-Bias Dalam Interaksi Pembelajaran: Analisis Penyebab Rendahnya Partisipasi Belajar Murid Laki-Laki Di Kelas Di SMA NEGERI 1 Panyabungan," *Pendas J. Ilm. Pendidik. Dasar*, vol. 11, no. 01, pp. 34–45, Mar. 2026, doi: 10.23969/jp.v11i01.42975.

[5] R. N. Rochma, MAHMUD, and F. Z. A. Farizy, "PENGARUH IKLIM Belajar Dan Interaksi Guru Terhadap Partisipasi Murid Pada Pembelajaran SKI," *Dar El-Falah*, vol. 5, no. 2, pp. 133–148, Jul. 2026, doi: 10.67409/6q288p26.

[6] N. Nuraini and S. Supriyanto, "Analisis Partisipasi Murid SMA Dalam Pembelajaran Daring: Studi Pada Murid Kelas IPS XI SMAS AL-IZHAR Jakarta," *J. Anal. Pendidik. Sos.*, vol. 1, no. 1, pp. 1–7, Nov. 2023.

[7] H. Jafar, L. Hakim, and D. Putra, "Manajemen Hubungan Masyarakat Dalam Meningkatkan Partisipasi Wali Murid di SMP Negeri 27 Kerinci," *J. Sustain.*, vol. 6, no. 2, pp. 720–728, Dec. 2023, doi: 10.32923/kjmp.v6i2.4083.

[8] X. Huang and Y. Tong, "A Parameter Assessment of Teaching Quality Indicators Based on Data Class Mining Fuzzy K-Mean Type Clustering," *Scalable Comput. Pract. Exp.*, vol. 25, no. 3, pp. 1578–1587, Apr. 2024, doi: 10.12694/scpe.v25i3.2768.

[9] M. Kalra, N. Lal, and S. Qamar, "K-Mean Clustering Algorithm Approach for Data Mining of Heterogeneous Data," in *Information and Communication Technology for Sustainable Development*, D. K. Mishra, M. K. Nayak, and A. Joshi, Eds., Singapore: Springer, 2018, pp. 61–70. doi: 10.1007/978-981-10-3920-1_7.

[10] Y. Guo, B. Wu, and Y. Liu, "Multidimensional data mining using a K-mean algorithm based on the forest management inventory of Fujian Province, China," *TELKOMNIKA Indones. J. Electr. Eng.*, vol. 11, no. 12, pp. 7290–7294, Dec. 2013.

[11] S. R. Durugkar, R. Raja, K. K. Nagwanshi, and S. Kumar, "Introduction to Data Mining," in *Data Mining and Machine Learning Applications*, John Wiley & Sons, Ltd, 2022, pp. 1–19. doi: 10.1002/9781119792529.ch1.

[12] M. K. Gupta and P. Chandra, "A comprehensive survey of data mining," *Int. J. Inf. Technol.*, vol. 12, no. 4, pp. 1243–1257, Dec. 2020, doi: 10.1007/s41870-020-00427-7.

[13] D. L. Olson and G. Lauhoff, "Descriptive Data Mining," in *Descriptive Data Mining*, D. L. Olson and G. Lauhoff, Eds., Singapore: Springer, 2019, pp. 129–130. doi: 10.1007/978-981-13-7181-3_8.

[14] M. Amelia, A. Faqih, and A. R. Rinaldi, "Penerapan Metode K-Means Clustering Dalam Pemetaan Kemiskinan Kabupaten/Kota Di Indonesia Untuk Perencanaan Kebijakan Yang TepaT," *J. Inform. Dan Tek. Elektro Terap.*, vol. 13, no. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6231.

[15] D. T. Cahaya, D. Puspita, and R. Syahri, "Penerapan Metode K-Means Clustering Untuk Pengelompokan Potensi Padi Di Kota Pagar Alam," *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 2187–2193, Apr. 2024, doi: 10.36040/jati.v8i2.9432.

[16] A. Yudistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *J. Artif. Intell. Technol. Inf. JAITI*, vol. 1, no. 1, pp. 20–28, Mar. 2023, doi: 10.58602/jaiti.v1i1.22.

[17] T. P. Van Song, Y. Qiao, M. Vicic, and A. Hadasch, "A New Method for Skin Color Classification Based on Global CIELAB Data and k - Mean Clustering," *Color Res. Appl.*, vol. 51, no. 1, p. e70012, 2026, doi: 10.1002/col.70012.

[18] I. K. Khan *et al.*, "Determining the optimal number of clusters by Enhanced Gap Statistic in K-mean algorithm," *Egypt. Inform. J.*, vol. 27, p. 100504, Sep. 2024, doi: 10.1016/j.eij.2024.100504.

[19] A. C. Pise and K. J. Karande, "K-Mean Energy Efficient Optimal Cluster Based Routing Protocol in Vehicular Ad-Hoc Networks," in *Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough: Volume 4*, V. K. Gunjan, J. M. Zurada, and N. Singh, Eds., Cham: Springer International Publishing, 2024, pp. 305–313. doi: 10.1007/978-3-031-43009-1_25.

[20] R. M. K. Sandi, Carudin, and I. Maulana, "Implementasi Sistem E-Learning Berbasis Web Menggunakan SDLC Model Fountain (Studi Kasus: SMA Islam Cikuya Bungbulang)," *J. Sci. Res. Dev.*, vol. 8, no. 1, pp. 1077–1090, Sep. 2026, doi: 10.56670/jsrd.v8i1.1674.