Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2059−2068
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8585

Student Class Grouping in Junior High Schools Based on Academic Performance Using the Fuzzy C-Means Method

Tommy Bustomi[1,*], Jundro Daud Hasiholan[2], Kusno Harianto[2]

[1] Program Studi Teknik Informatika Multimedia, Politeknik Negeri Samarinda, Samarinda, Indonesia
[2] Program Studi Sistem Informasi, STMIK Widya Cipta Dharma, Samarinda, Indonesia
Email: [1,*]tbustomi@gmail.com, [2]daudjundro@gmail.com, [3]kusno@wicida.ac.id
Corresponding Author Email: tbustomi@gmail.com
Submitted: **22/10/2025**; Accepted: **31/12/2025**; Published**: 31/12/2025**

Abstrak−Differences in academic abilities among junior high school students often pose a challenge for schools in conducting class groupings objectively and efficiently. Many educational institutions, including SMP Negeri Y, still rely on manual grouping methods that are subjective and do not accurately reflect the actual conditions of students. Inaccurate grouping may lead to imbalanced learning processes, where students with high and low academic abilities are placed in the same group without considering their performance variations. Therefore, a data-driven approach is needed to represent student characteristics comprehensively and flexibly. This study aims to apply the Fuzzy C-Means (FCM) method to cluster students of SMP Negeri Y based on four main attributes: Academic Average, Attitude Score, Activeness Score, and Attendance. The FCM method was chosen for its ability to handle data uncertainty and assign multiple membership degrees to each student across different clusters. Prior to clustering, the data underwent a preprocessing stage involving data cleaning, normalization using StandardScaler, and scale adjustment across attributes to improve the accuracy of Euclidean distance calculations. The analysis results revealed the formation of two main clusters representing student academic performance levels. Cluster 0 has an average academic score of 78.37 with moderate attitude and activeness levels, while Cluster 1 shows a higher academic average of 82.18 accompanied by better attitude, activeness, and attendance scores. Based on the highest membership degree, 38 students were assigned to Cluster 0 and 26 students to Cluster 1. Model evaluation using Fuzzy Partition Coefficient (FPC), Modified Partition Coefficient (MPC), and Silhouette Score indicated the optimal configuration at a fuzziness level of m = 2, yielding FPC = 0.680, MPC = 0.359, and Silhouette Score = 0.334. These findings demonstrate that FCM is effective in representing variations in student abilities more realistically, while also providing an objective foundation for schools to design adaptive learning strategies and implement data-driven academic policies.

Keywords: Fuzzy C-Means; Clustering; Academic Performance; Student Grouping; Data Mining

1. INTRODUCTION

The rapid development of information technology over the past decade has had a significant impact on various aspects of human life, including the field of education. Advances in data mining have enabled the extraction of valuable insights from the vast volumes of data generated daily by educational institutions. Data that previously served merely as archives can now be processed into strategic sources of evidence-based decision-making. Through data analysis techniques, schools and educational institutions can identify patterns of student achievement, detect learning difficulties at an early stage, and evaluate the effectiveness of implemented teaching processes [1]. In this context, academic data analysis plays a crucial role as a tool for grouping students based on their abilities, interests, and academic performance levels.

One commonly used data analysis technique is clustering, which refers to the process of grouping data based on shared characteristics. Among various clustering methods, Fuzzy C-Means (FCM) holds a prominent position due to its ability to handle uncertain data and overlapping cluster boundaries. Unlike methods such as K-Means, which require each data point to belong exclusively to one cluster, FCM allows each object to possess a membership degree across multiple clusters simultaneously [2]. This approach aligns well with the complex nature of educational data, where students often exhibit varying levels of ability that cannot be rigidly classified.

In the educational context, the FCM method has proven effective in several recent studies. For instance, [3] demonstrated that implementing FCM optimized with Particle Swarm Optimization (PSO) increased graduation prediction accuracy by up to 86%. Similarly, [4] found that combining Self-Organizing Maps (SOM) and FCM effectively clustered teacher data to support managerial decision-making in educational institutions. These findings reinforce evidence that FCM provides more flexible and accurate clustering results in educational data contexts..

One of the major challenges faced by many junior high schools, including SMP Negeri Y, is the difficulty of objectively and efficiently grouping students based on academic performance. According to the Ministry of Education and Culture, approximately 35% of junior high school students in Indonesia have not yet achieved the national minimum standard average score. This performance disparity may stem from internal factors, such as differences in learning abilities and motivation, as well as external factors, such as socioeconomic background and learning environment support [5]. Such complexity makes educational data heterogeneous, nonlinear, and uncertain, posing challenges for conventional classification methods [6].

Therefore, a data analysis approach capable of representing actual conditions more comprehensively is required [7]. In this regard, the Fuzzy C-Means method emerges as a highly promising solution. By assigning membership degrees to each student, this method allows individuals to belong to multiple groups proportionally based on the

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2059−2068
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8585

similarity of their academic scores. This is highly relevant, as in reality, student ability cannot be dichotomized into "smart" and "less capable" categories but rather exists along a continuous spectrum of competence [8].

Previous research has shown that FCM is effective in clustering educational data based on academic performance indicators. For example, [9] applied FCM to analyze e-commerce data across 34 Indonesian provinces, demonstrating its effectiveness in handling large-scale data variations and producing well-defined clusters. Meanwhile, [10] developed a new FCM-based educational segmentation framework optimized using the Differential Evolution algorithm, which generated eight accurate clusters for student recruitment strategies. This shows that FCM is not only valuable for academic analysis but also applicable in broader educational management strategies. Furthermore, [11] successfully grouped Indonesian primary and secondary schools based on national education standards using FCM, resulting in four well-validated clusters.

In the context of SMP Negeri Y, the application of FCM is expected to assist the school in grouping students based on their academic excellence. With accurate clustering results, the school can design more adaptive and targeted learning strategies. For instance, students with similar competency levels can be taught through appropriate approaches, while those on the boundary between clusters may receive special interventions to enhance their performance. Such a data-driven approach marks an important step toward implementing adaptive and evidence-based education, a growing trend in the global digital transformation of education [12].

Moreover, the advantage of the FCM method lies in its ability to handle overlapping data [9]. In student datasets, it is common to find individuals with average scores that fall between two categories, such as "fairly good" and "good." Conventional methods like K-Means would assign such students to only one cluster, thus overlooking data nuances. However, FCM enables a more realistic grouping by accounting for membership probabilities across multiple clusters. This approach aligns with the inherently dynamic and multidimensional nature of human characteristics [13].

Methodologically, FCM operates by randomly initializing cluster centers and then iteratively updating them until reaching a minimal error threshold. This iterative process enhances accuracy, as cluster centers dynamically adjust to the data distribution. Consequently, this procedure yields more optimal data partitions compared to non-fuzzy methods [4]. Therefore, FCM not only delivers more accurate clustering outcomes but also supports objective evaluation and educational policy planning [11].

In summary, the aim of this study is to apply the Fuzzy C-Means (FCM) method to cluster students at SMP Negeri Y based on their academic achievements. This research is expected to make a tangible contribution to the application of data mining in education, particularly in helping schools form homogeneous and efficient learning groups. Furthermore, the study seeks to demonstrate that FCM can serve as a decision-support tool capable of representing students' abilities more equitably and comprehensively [8].

2. RESEARCH METHODOLOGY

2.1 Research Stages

The research stages present the methodological framework for classifying junior high school students based on academic achievement using the Fuzzy C-Means method. This diagram illustrates a systematic sequence beginning with problem identification and ending with the generation of clustering results ready for interpretation. Each stage is organized sequentially to ensure that the analytical process is conducted in a structured, objective, and replicable manner. Through the flow shown in Figure 1, it can be understood how student grade data are processed step by step to produce class group information that more accurately represents students' academic proficiency levels.

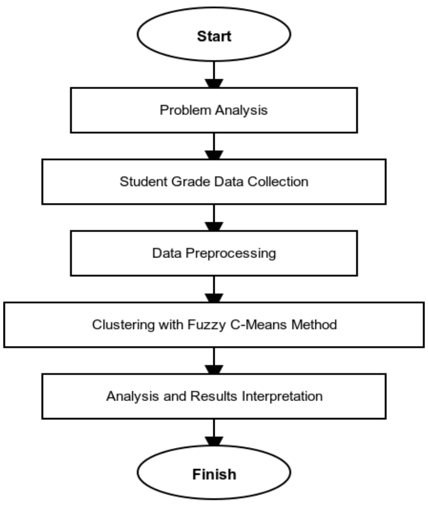

Figure 1. Research Stages

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2059−2068
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8585

Based on Figure 1, the research process begins with problem analysis, which involves identifying the school's need to create fair and data-driven class groupings. The next stage is student grade data collection, focusing on gathering student grade data as the main variable in the analysis. The collected data then undergo data preprocessing to ensure data quality, including data cleaning, normalization, and handling of missing values. Once the data are ready, the clustering with Fuzzy C-Means method stage is performed as the core process aimed at grouping students based on similarities in academic performance, with fuzzy membership degrees. The final stage is analysis and results interpretation, where the clusters are analyzed to extract academic meaning and provide a basis for decision-making in student class placement. This flow demonstrates that each stage is logically interconnected to produce an objective and data-driven class grouping.

2.2 Data Mining and Clustering

Data mining is a field of study focused on extracting information or hidden patterns from large datasets using techniques from statistics, artificial intelligence, and machine learning[14]. In the educational context, this approach is known as Educational Data Mining (EDM), which aims to discover new insights from student-related data such as academic grades, attendance, and learning activities[15]. EDM plays a crucial role in helping educators understand students' learning behaviors and design more effective interventions based on data-driven evidence. With the growing volume of educational data stored digitally, the application of data mining can improve the quality of academic decision-making, such as identifying at-risk students and providing personalized learning recommendations. Therefore, data mining serves a strategic role in realizing an analytics-based education system that adapts to students' needs[16]. One of the primary techniques in data mining widely applied in the education domain is clustering, or data grouping[17]. Clustering is used to group students into several clusters based on similarities in characteristics such as exam scores, learning activities, and class participation levels[18]. Through clustering, educational institutions can automatically identify groups of high-, medium-, and low-performing students without manual intervention[19]. This technique helps schools understand general patterns in academic data and develop targeted policies, such as special tutoring or remedial programs[20]. Furthermore, clustering also serves as an essential preprocessing stage in machine learning–based academic performance prediction systems[14] [21]. In general, clustering methods are divided into two main approaches: hard clustering and fuzzy clustering[22]. In hard clustering, such as the K-Means algorithm, each object can only belong to one cluster closest to its centroid[18]. In contrast, the Fuzzy C-Means (FCM) algorithm provides greater flexibility by allowing each data point to belong to more than one cluster with varying membership degrees[17]. This method is particularly useful in educational contexts where differences in student performance are not always clear-cut and often gradual[20].

2.3 Fuzzy C-Means (FCM)

The Fuzzy C-Means (FCM) method is a clustering algorithm that allows each data object to have a degree of membership to every cluster rather than belonging exclusively to a single one[21]. In FCM, the objective function minimized is the total of each membership degree raised to the power m, multiplied by the squared distance to the cluster center. The algorithm iteratively updates the cluster centers and membership values until convergence is achieved. The advantages of this method lie in its flexibility to handle data with ambiguous cluster boundaries and its ability to assign partial memberships that more realistically represent real-world cases [23]. However, FCM also has limitations, such as its sensitivity to the initial selection of cluster centers and the tendency to become trapped in local minima if not properly configured[17]. The main objective of FCM is to minimize the cost function J_m , as defined in Equation (1),

$$J_m = \sum_{i=1}^{C} \sum_{j=1}^{N} \mu_{ij}^{m} \parallel x_i - c_j \parallel^2 \tag{1}$$

Where N denotes the number of data points, C the number of clusters, the x_i i^{th} data point, c_j the center of cluster j, μ_{ij} the membership degree of x_i in cluster j, and m (usually $1{,}5 \leq m \leq 2{,}5$) the fuzziness coefficient controlling cluster overlap. The membership value for each data point is then updated according to Equation (2),

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\| x_i - c_j \|}{\| x_i - c_k \|} \right)^{\frac{2}{m-1}}} \tag{2}$$

subject to the constraint that the total membership for every data point equals one, as expressed in Equation (3),

$$\sum_{j=1}^{C} \mu_{ij} = 1 \tag{3}$$

After updating the membership degrees, the cluster centers are recalculated using Equation (4),

$$c_j = \frac{\sum_{i=1}^{N} \mu_{ij}^{m} x_i}{\sum_{i=1}^{N} \mu_{ij}^{m}} \tag{4}$$

This iterative process continues until convergence is reached, which occurs when the difference between cluster centers in consecutive iterations falls below a defined threshold ε as given in Equation (5),

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2059−2068
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8585

$$\left\|C^{(t+1)} - C^{t}\right\| < \varepsilon \tag{5}$$

Through this iterative optimization process, the FCM algorithm dynamically adjusts cluster centers based on data distribution, resulting in more accurate and flexible clustering outcomes compared to conventional hard clustering methods such as K-Means[16][14].

3. RESULTS AND DISCUSSION

3.1 Overview of the Dataset

The data used in this study consist of the compiled academic scores of students from SMP Negeri Y for the 2024/2025 academic year. The dataset was obtained from the school's academic division and includes four main attributes: Academic Average, Attitude Score, Activeness Score, and Attendance, representing the cognitive, affective, and psychomotor aspects of student performance comprehensively. This data served as the input for the clustering process using the Fuzzy C-Means (FCM) algorithm. This method was selected because it effectively handles data uncertainty and produces more flexible clustering results compared to hard clustering methods such as K-Means. Through FCM, each student possesses a degree of membership in more than one cluster according to the similarity of their characteristics, resulting in a more realistic analysis of student performance.

Before the clustering process, the data underwent preprocessing, which included checking for missing values, removing duplicate entries, and adjusting the scale across attributes to ensure balanced variable weights. The attributes Attitude Score, Activeness Score, and Attendance were also converted from percentage values to decimal form to standardize numerical formats with other attributes. This step ensured that the dataset was in optimal condition and ready for clustering using the Fuzzy C-Means algorithm. The cleaned and normalized final dataset is presented in Table 1.

Table 1. Student Scores of SMP Negeri Y After Attribute Scale Adjustment

No	Student Name	Academic Average	Attitude Score	Activeness Score	Attendance
1	Davin Alvaro Pratama	85.3	0.19	0.2	100
2	Gilang Aditya Purnomo	78.3	0.15	0.16	97
3	Keysha Aurelia Putri	80.8	0.15	0.17	94
4	Naila Zahira Ramadhani	76.8	0.15	0.16	96
5	Farhan Rizqy Saputra	78.5	0.15	0.16	92
...	...	...	...	...	...
63	Anindya Cahya Lestari	84.1	0.18	0.2	95
64	Reyhan Althaf Nasution	78.4	0.15	0.17	93

In Table 1, aside from the removal of irrelevant sub-attributes, the values of Attitude Score, Activeness Score, and Attendance were converted into decimal form (originally in percentages) to align the scale of each attribute. This adjustment ensures that all variables contribute equally in the Euclidean distance calculation used in the Fuzzy C-Means algorithm. Consequently, the resulting clusters are more accurate and objectively represent the students' academic performance. The data preprocessing process ensured that the dataset was in optimal condition before applying the Fuzzy C-Means method. The preprocessing stages included checking for missing values, selecting numerical features, and normalizing the data using the StandardScaler function from the scikit-learn library. The examination showed that all attributes were complete, and no imputation was required. Feature selection ensured that only numerical attributes were used for clustering, while normalization using StandardScaler standardized the variable scales to prevent bias during distance calculations. The StandardScaler converts each attribute into a distribution with a mean of 0 and a standard deviation of 1.

	No	Rata-rata Akademik	Nilai Sikap	Nilai Keaktifan	Kehadiran
0	-1.705196	-1.422926	-0.770344	-1.079743	-0.182631
1	-1.651062	-0.729227	-0.770344	-1.757229	-0.516586
2	-1.596929	0.080089	-0.770344	-0.402257	0.485278
3	-1.542796	1.159176	-0.045314	0.952714	0.151323
4	-1.488663	-0.189683	-0.770344	-1.757229	-2.186358
...	...	...	...	...	...
59	1.488663	0.504016	1.404744	1.630200	-0.182631
60	1.542796	0.311322	-0.045314	0.952714	0.151323
61	1.596929	-1.384387	-0.045314	-0.402257	1.153186
62	1.651062	0.195705	0.679715	0.952714	1.153186
63	1.705196	-0.497994	0.679715	-0.402257	0.485278

64 rows × 5 columns

Figure 1. Normalization Results

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2059−2068
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8585

Based on the normalization results in Figure 1, all attributes Academic Average, Attitude Score, Activeness Score, and Attendance are now on a uniform scale. The data values are centered around 0, with variations generally ranging from −2 to +2. This indicates that normalization successfully standardized the scale of each feature, ensuring that no attribute dominates the clustering process due to differences in units or magnitudes.

This adjustment is crucial, as prior to normalization, each attribute had different value ranges. With the standardized scales, the data became ready for the clustering process using the Fuzzy C-Means method. Normalization also helped the algorithm compute distances more accurately and accelerated the convergence of the clustering process. The initial membership degree values are shown in Figure 2, illustrating that each data point has two membership degrees corresponding to Cluster 0 and Cluster 1.

	Data Point	Cluster 0 Init	Cluster 1 Init
0	1	0.210163	0.789837
1	2	0.596538	0.403462
2	3	0.162658	0.837342
3	4	0.303624	0.696376
4	5	0.443536	0.556464
...	...	...	...
59	60	0.824669	0.175331
60	61	0.584455	0.415545
61	62	0.339562	0.660438
62	63	0.213461	0.786539
63	64	0.980178	0.019822

64 rows × 3 columns

Figure 2. Initial Membership Degrees

As shown in Figure 2, the table presents the initial membership degree values generated randomly at the start of the Fuzzy C-Means (FCM) algorithm. Each data point has two membership degree values, one for Cluster 0 and another for Cluster 1, representing the degree of closeness or likelihood that a data point belongs to each cluster. The initial membership degrees were generated randomly, as illustrated in Figure 3. At this stage, each data point obtains two membership degree values corresponding to its closeness to each cluster. According to the fundamental concept of Fuzzy C-Means, the sum of these two membership degrees for each data point always equals 1.

> Centroid Akhir (Scaled):

	No	Rata-rata Akademik	Nilai Sikap	Nilai Keaktifan	Kehadiran
Cluster 0	-0.192910	-0.585240	-0.600952	-0.659778	-0.496668
Cluster 1	0.144661	0.882483	0.905013	0.862482	0.589432

Figure 3. Final Centroids (Normalized Data)

However, since the centroid values are still in the normalized scale, a reverse transformation (inverse transform) is required to convert them back to their original scale. This transformation facilitates interpretation, allowing the obtained centroids to be directly compared with the original data to better describe each cluster's characteristics. The adjusted centroids based on the original data are shown in Figure 4.

> Centroid Akhir (Asli):

	No	Rata-rata Akademik	Nilai Sikap	Nilai Keaktifan	Kehadiran
Cluster 0	28.936385	78.373615	0.152336	0.166199	95.059643
Cluster 1	35.172312	82.182040	0.173107	0.188668	98.311881

Figure 4. Final Centroids (Original Data)

Based on Figure 4, the final centroid values (in original data scale) show clear distinctions between the two formed clusters. Cluster 0 has an average academic score of 78.373615, an attitude score of 0.152336, an activeness score of 0.166199, and an attendance score of 95.059643. This indicates that students in this cluster demonstrate good academic performance and attendance, but moderate attitude and activeness levels. Meanwhile, Cluster 1 exhibits higher centroid values, with an average academic score of 82.182040, an attitude score of 0.173107, an activeness score of 0.188668, and an attendance score of 98.311881. Thus, this cluster represents students with overall higher performance in academic achievement, attitude, activeness, and attendance.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2059−2068
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8585

Overall, the two clusters can be distinguished based on student performance, where Cluster 1 represents high-performing students and Cluster 0 represents students with moderate achievements. Based on the proximity of each student's attribute values to the corresponding cluster centroid, 38 students belong to Cluster 0, while 26 students are grouped into Cluster 1, as shown in Figure 5.

Cluster	
0	38
1	26

Figure 5. Number of Data Points per Cluster

As displayed in Figure 5, 38 students belong to Cluster 0 and 26 students belong to Cluster 1. The grouping was determined based on the highest membership degree for each student toward a given cluster. Consequently, students in Cluster 0 slightly outnumber those in Cluster 1. The cluster assignment was made by evaluating the proximity of each data point to the cluster centroids, ensuring that each student is placed in the group that best represents their characteristics.

	Cluster 0	Cluster 1
0	0.828560	0.171440
1	0.831783	0.168217
2	0.699745	0.300255
3	0.329531	0.670469
4	0.781543	0.218457
...	...	...
59	0.201903	0.798097
60	0.321404	0.678596
61	0.591991	0.408009
62	0.221266	0.778734
63	0.488647	0.511353

64 rows × 2 columns

Figure 6. Final Membership Degrees

As shown in Figure 6, each row of data represents a student's degree of membership to Cluster 0 and Cluster 1. Clustering was performed by assigning each student to the cluster with the higher membership degree. For example, in the first row, the student's membership degree to Cluster 0 is 0.828560, while the degree to Cluster 1 is 0.171440. Since the value for Cluster 0 is higher, this student is categorized under Cluster 0. The same principle applies to all other data points, where the cluster assignment is based on the maximum membership degree.

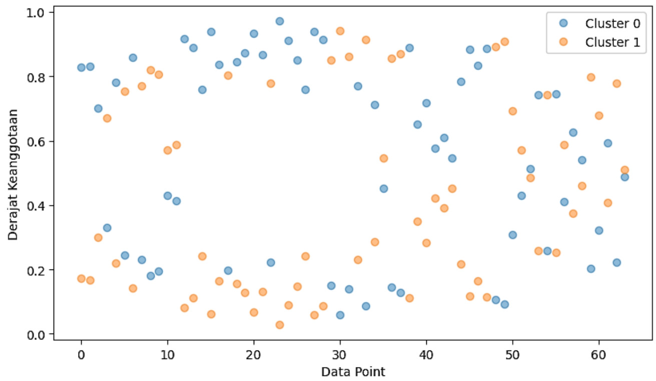

Figure 7. Membership Degree Distribution

In Figure 7, each data point represents a student's membership degree to the two clusters resulting from the Fuzzy C-Means clustering process. Membership degree values range from 0 to 1, where a higher value indicates stronger similarity to the cluster's characteristics. The visualization shows that most students have distinct membership degrees closer to 0 or 1 allowing the clusters to be interpreted clearly. Students are assigned to a particular cluster when their membership degree to that cluster exceeds 0.5. This demonstrates that the FCM model successfully differentiates student characteristics based on the selected variables.

Visually, the chart reveals that most students are clearly associated with a particular cluster, though some have intermediate membership values (around 0.4–0.6), indicating mixed characteristics between the two clusters. This is typical in fuzzy clustering, as each data point can belong partially to multiple clusters. Such cases can be further analyzed to identify factors contributing to their shared characteristics.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2059−2068
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8585

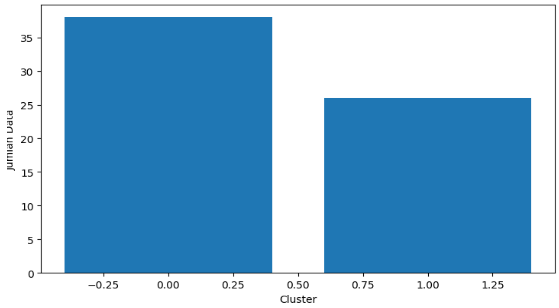

Figure 8. Clustering Result Distribution

This visualization displays the distribution of data points in each cluster. The X-axis shows the two clusters (Cluster 0 and Cluster 1), while the Y-axis represents the number of data points in each cluster. The height of each bar clearly illustrates the comparison between cluster sizes. The chart indicates that Cluster 0 contains approximately 38 data points, while Cluster 1 consists of about 26. This suggests that most data points share characteristics closer to those of Cluster 0.

The effect of the fuzziness parameter (m) was analyzed for cluster numbers ranging from 2 to 6, evaluated using three validation metrics: Fuzzy Partition Coefficient (FPC), Modified Partition Coefficient (MPC), and Silhouette Score, calculated using cluster.cmeans from skfuzzy and silhouette_score from sklearn.metrics. The clustering performance evaluation results are shown in Figure 9.

```
========================================
EVALUASI JUMLAH CLUSTER OPTIMAL
========================================
n_clusters = 2: FPC = 0.680, MPC = 0.359, Silhouette = 0.334
n_clusters = 3: FPC = 0.592, MPC = 0.388, Silhouette = 0.333
n_clusters = 4: FPC = 0.466, MPC = 0.289, Silhouette = 0.281
n_clusters = 5: FPC = 0.414, MPC = 0.267, Silhouette = 0.216
n_clusters = 6: FPC = 0.398, MPC = 0.277, Silhouette = 0.206
```

Figure 9. Cluster Evaluation for Fuzziness = 2

Based on Figure 9, the optimal number of clusters is 2, indicated by the highest validation metric values: FPC = 0.680, MPC = 0.359, and Silhouette Score = 0.334. As the number of clusters increases, the FPC, MPC, and Silhouette values decrease, suggesting that the separation quality between clusters diminishes. This confirms that forming more than two clusters does not improve cluster quality significantly.

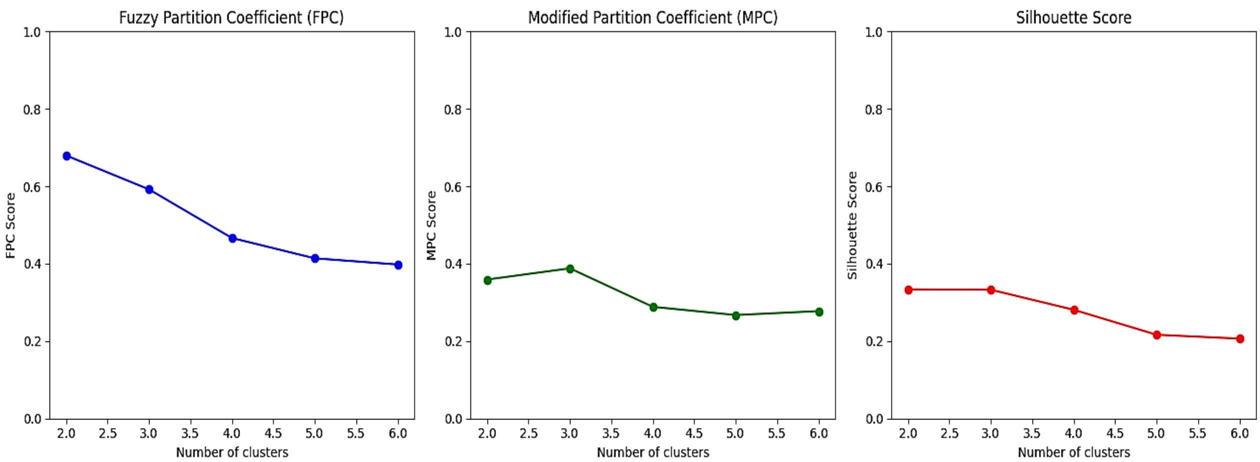

Figure 10. Cluster Evaluation Graph for Fuzziness = 2

As shown in Figure 10, the three metrics (FPC, MPC, and Silhouette Score) all reach their highest values when the number of clusters is 2. Specifically, FPC peaks at 0.680, MPC at 0.359, and Silhouette Score at 0.334. These results indicate that the optimal clustering configuration occurs at two clusters, as it provides the best separation and internal cohesion..

```
n_clusters = 2: FPC = 0.553, MPC = 0.106, Silhouette = 0.324
n_clusters = 3: FPC = 0.410, MPC = 0.115, Silhouette = 0.333
n_clusters = 4: FPC = 0.303, MPC = 0.071, Silhouette = 0.258
n_clusters = 5: FPC = 0.242, MPC = 0.053, Silhouette = 0.187
n_clusters = 6: FPC = 0.219, MPC = 0.063, Silhouette = 0.148
```

Figure 11. Cluster Evaluation for Fuzziness = 3

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2059−2068
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8585

Based on the results, the FPC value is highest for two clusters (0.553) and gradually decreases to 0.219 for six clusters. The MPC shows a similar pattern, decreasing from 0.106 (two clusters) to 0.063 (six clusters). The Silhouette Score peaks slightly at 0.333 for three clusters but remains close to the value for two clusters (0.324).

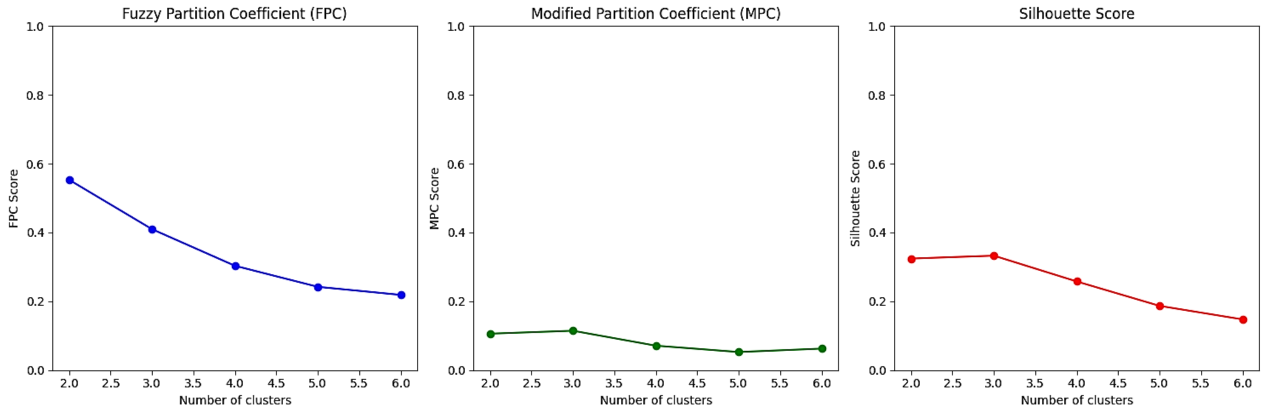

Figure 12. Cluster Evaluation Graph for Fuzziness = 3

The results again show that the highest metric values are achieved when the number of clusters (k) = 2, confirming the stability of the clustering structure.

```
n_clusters = 2: FPC = 0.519, MPC = 0.039, Silhouette = 0.321
n_clusters = 3: FPC = 0.363, MPC = 0.044, Silhouette = 0.337
n_clusters = 4: FPC = 0.271, MPC = 0.028, Silhouette = 0.247
n_clusters = 5: FPC = 0.215, MPC = 0.019, Silhouette = 0.095
n_clusters = 6: FPC = 0.180, MPC = 0.015, Silhouette = 0.152
```

Figure 13. Cluster Evaluation for Fuzziness = 4

Similarly, Figure 13 shows that the configuration with two clusters performs best, with the highest FPC (0.519), MPC (0.039), and Silhouette Score (0.321) among all tested configurations. As the number of clusters increases from 3 to 6, all metrics generally decline, indicating reduced clustering quality.

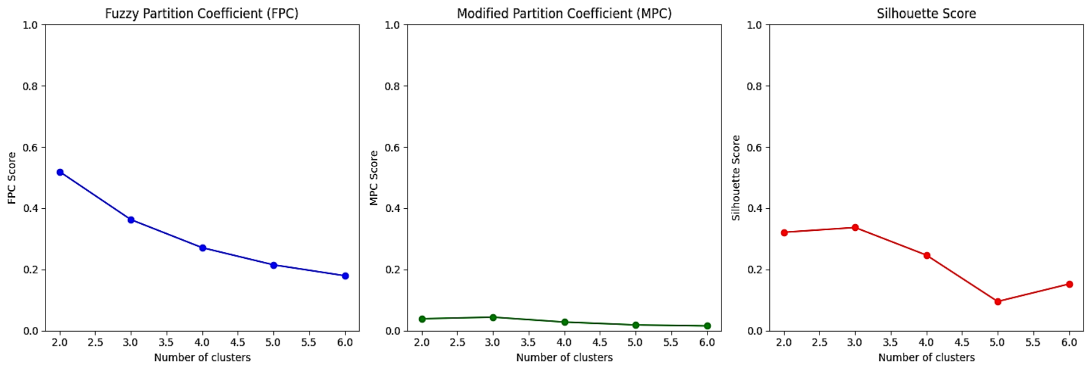

Figure 14. Cluster Evaluation Graph for Fuzziness = 4

The graph shows consistent trends across all three metrics: FPC, MPC, and Silhouette Score reach their maximum values when k = 2 and decrease with higher k values. This suggests that the clearest and most stable cluster structure is achieved when the dataset is divided into two groups.

Table 2. Comparison of Fuzzy C-Means Tests Based on Fuzziness Levels

Evaluation Metric	Fuzziness = 2	Fuzziness = 3	Fuzziness = 4
Fuzzy Partition Coefficient (FPC)	0.680	0.553	0.519
Modified Partition Coefficient (MPC)	0.359	0.106	0.039
Silhouette Score	0.334	0.324	0.321

Based on Table 2, all three metrics FPC, MPC, and Silhouette Score show a decreasing trend as the fuzziness parameter (m) increases from 2 to 4. The highest values are obtained at m = 2, indicating that at this level, the cluster structure is clearest and most well-defined. For the FPC, the value at m = 2 is 0.680, decreasing to 0.553 at m = 3 and 0.519 at m = 4, suggesting that cluster distinctness decreases with increasing fuzziness. Similarly, the MPC drops sharply from 0.359 at m = 2 to 0.106 at m = 3 and 0.039 at m = 4, indicating that inter-cluster boundaries become increasingly blurred. The Silhouette Score remains relatively stable across tests, with values of 0.334, 0.324, and 0.321

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2059−2068
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8585

for m = 2, 3, and 4, respectively. Although the difference is small, the highest score at m = 2 confirms that this configuration achieves the best balance between internal cohesion and inter-cluster separation.

Overall, the results consistently demonstrate that the optimal fuzziness value is m = 2, as it yields the highest evaluation metrics (FPC, MPC, and Silhouette Score). Therefore, it can be concluded that m = 2 produces the clearest, most stable, and most optimal cluster structure for the Fuzzy C-Means analysis in this study.

3.2 Discussion

The implementation of the Fuzzy C-Means (FCM) method on the academic data of students at SMP Negeri Y revealed that the dataset could be divided into two main clusters with distinct characteristics. Cluster 0 had an average academic score of 78.37, attitude score of 0.152, activeness score of 0.166, and attendance rate of 95.05, whereas Cluster 1 showed an average academic score of 82.18, attitude score of 0.173, activeness score of 0.189, and attendance rate of 98.31. Based on these centroid differences, Cluster 1 represents high-achieving students, while Cluster 0 reflects students with moderate performance.

The composition of students in each cluster shows that 38 students (59.4%) belong to Cluster 0, and 26 students (40.6%) belong to Cluster 1. This distribution indicates that the majority of students remain at a moderate achievement level, suggesting that the school should provide additional support or remedial programs for this group. Conversely, the high-performing cluster could be given enrichment programs to further develop their potential.

These findings are consistent with [3], who demonstrated that FCM optimized with Particle Swarm Optimization (PSO) increased graduation prediction accuracy up to 86% by effectively handling data with overlapping cluster boundaries. Similarly, Rahman (2022) found that combining Self-Organizing Maps (SOM) with FCM produced more flexible clustering results for teacher data. The current study also aligns with [11], who successfully classified Indonesian schools into four clusters based on national education standards with valid outcomes. Hence, the application of FCM in the context of SMP Negeri Y reinforces empirical evidence that this method is highly relevant for clustering complex and gradual educational data.

Furthermore, the Fuzzy Partition Coefficient (FPC) value of 0.680, Modified Partition Coefficient (MPC) of 0.359, and Silhouette Score of 0.334 indicate that the two formed clusters have a good separation quality. These evaluation results are consistent with findings from [9] and [10], who also reported that configurations of two to four clusters yield the most optimal structure for medium-sized datasets with high variability.

The FCM approach proves to be more adaptive than hard clustering methods such as K-Means because it allows each student to possess a degree of membership in more than one cluster. This flexibility reflects real educational conditions, where students' abilities cannot be rigidly categorized. For instance, a student with a high academic score but moderate activeness may have nearly equal membership in two clusters. This finding supports the view of Zhang & Han (2021) that FCM more accurately represents the multidimensional nature of human characteristics.

From a practical standpoint, the findings of this study provide valuable insights for schools in designing differentiated instruction and implementing data-driven decision-making. Teachers can use the clustering results to tailor teaching strategies according to the ability levels of each student group. Additionally, students whose membership degrees approach the threshold (0.4–0.6) should be prioritized for early intervention, as they exhibit potential for either academic improvement or decline.

In conclusion, this study strengthens empirical evidence that the Fuzzy C-Means method is an effective and flexible approach for analyzing academic data in secondary schools. The findings not only support previous studies but also offer practical contributions to the development of adaptive and analytics-based education systems within the school environment.

4. CONCLUSION

Based on the results of the study on class grouping based on students' academic performance at SMP Negeri Y using the Fuzzy C-Means (FCM) method, several conclusions can be drawn. The FCM method has proven effective in clustering students into two main groups based on four variables: Academic Average, Attitude Score, Activeness Score, and Attendance. The first cluster represents students with high academic performance, while the second cluster corresponds to those with moderate achievement levels. The parameter testing results indicate that a fuzziness value of m = 2 produces the most optimal cluster structure compared to m = 3 and m = 4. This is evidenced by the highest validation metric values at m = 2, namely a Fuzzy Partition Coefficient (FPC) of 0.680, a Modified Partition Coefficient (MPC) of 0.359, and a Silhouette Score of 0.334. The decrease in all three metrics at m = 3 and m = 4 suggests that increasing the fuzziness value reduces the clarity and quality of the cluster formation. The findings also emphasize that data preprocessing and normalization play a crucial role in achieving accurate clustering outcomes. These stages ensure that all attributes contribute equally to the analysis, preventing any single variable from dominating the clustering process. The clustering results reveal that students in Cluster 1 exhibit higher academic averages, attitudes, activeness, and attendance levels compared to those in Cluster 0. Therefore, the clustering results can serve as a valuable reference for the school in designing more effective learning strategies, such as forming balanced study groups, providing targeted academic guidance, or recognizing high-achieving students. Overall, this study demonstrates that the Fuzzy C-Means method produces clear, flexible, and representative clustering results for academic data. A fuzziness value of m = 2 is recommended as the optimal parameter to generate stable and easily interpretable cluster structures.

REFERENCES

[1] I. Auliya, F. Fitri, N. Amalita, and O. Mukhti, "Comparison of K-Means and Fuzzy C-Means Algorithms for Clustering Based on Happiness Index Components Across Provinces in Indonesia," *UNP J. Stat. Data Sci.*, vol. 2, no. 1, pp. 114–121, 2024, doi: 10.24036/ujsds/vol2-iss1/150.
[2] G. C. Messakh, M. N. Hayati, S. Sifriyani, and G. C. Messakh, "Comparison K-Means and Fuzzy C-Means in Regencies / Cities Grouping Based on Educational Indicators," *J. Varian*, vol. 7, no. 1, pp. 99–114, 2023, doi: 10.30812/varian.v7i1.2879.
[3] H. Jamhur, "Pemodelan Prediksi Predikat Kelulusan Mahasiswa Menggunakan Fuzzy C-Means Berbasis Particle Swarm Optimization," *TeknoIS J. Ilm. Teknol. Inf. dan Sains*, vol. 10, no. 1, pp. 13–24, 2020, doi: 10.36350/jbs.v10i1.79.
[4] A. Rahman, "Analysis of Self-Organizing Maps and Fuzzy C-Means methods in Clustering Teacher Data for Nominations of Candidates for Education Unit Supervisors," *Kontigensi J. Ilm. Manaj.*, vol. 10, no. 2, pp. 431–438, 2022, doi: 10.56457/jimk.v10i2.356.
[5] M. W. Warolemba, Resmawan, and D. R. Isa, "Analisis Cluster Fuzzy C-Means dan Diskriminan untuk Pengelompokan Data Kesejahteraan Rakyat," *SAINSMAT J. Ilm. Ilmu Pengetah. Alam*, vol. 12, no. 2, pp. 141–152, 2023, [Online]. Available: https://journal.unm.ac.id/index.php/sainsmat/article/view/7181
[6] A. Fadlil, I. Riadi, and Y. Mulyana, "Integration of fuzzy C-Means and SAW methods on education fee assistance recipients," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 8, no. 2, 2023, doi: 10.22219/kinetik.v8i2.1636.
[7] Y. Xu, L. Zhang, P. Yu, and C. Zhao, "Using fuzzy C-means Clustering to Identify Heavy Metal Polluted Soil in a Certain Area of Shanghai Using fuzzy C-means Clustering to Identify Heavy Metal Polluted Soil in a Certain Area of Shanghai," in *IOP Conference Series: Earth and Environmental Science*, 2021. doi: 10.1088/1755-1315/660/1/012092.
[8] L. R. Ilmi, Haryanto, and M. Sumunaringtyas, "Clustering of High School Quality Using Fuzzy C-Means in the Special Region of Yogyakarta Province," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 10, no. 2, 2025, doi: 10.22219/kinetik.v10i2.2187.
[9] B. Oktaviandi, T. O. Mukhti, Y. Kurniawati, and Z. Martha, "Implementation of the Fuzzy C-Means Clustering Method in Grouping Provinces in Indonesia based on the Types of Goods Sold in E-commerce Businesses in 2022," *UNP J. Stat. Data Sci.*, vol. 2, no. 3, pp. 360–365, 2024, doi: 10.24036/ujsds/vol2-iss3/210.
[10] R. Bakri, B. Sobirov, N. P. Astuti, A. S. Ahmar, and P. K. Singh, "A New Framework for Dynamic Educational Marketing Segmentation in Student Recruitment: Optimizing Fuzzy C-Means with Metaheuristic Techniques," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 9, no. 3, 2025, doi: 10.29207/resti.v9i3.6515.
[11] A. T. Rumiati, M. Rif'an, N. A. S. Harwanti, Chusna, and H. Annuril, "Clustering of Primary and Secondary School in Indonesia Using The Fuzzy C-Means Method Based on School Self-Evaluation With Imputation Data," *J. Educ.*, vol. 4, no. 8, pp. 20–31, 2021, doi: 10.53819/81018102t2024.
[12] S. Supangat, M. Z. Bin Saringat, G. Kusnanto, and A. Andrianto, "Churn prediction on higher education data with fuzzy logic algorithm," *SISFORMA J. Inf. Syst.*, vol. 8, no. 1, pp. 22–29, 2021, doi: 10.24167/sisforma.v8i1.3025.
[13] Y. Zhang and J. Han, "Differential privacy fuzzy C-means clustering algorithm based on gaussian kernel function," *PLoS One*, vol. 16, no. 3, p. e0248737, 2021, doi: 10.1371/journal.pone.0248737.
[14] M. M. Saeed, "Forecasting the Academic Performance by Leveraging Educational Data Mining," *Intell. Autom. Soft Comput.*, vol. 39, no. 2, pp. 213–231, 2024, doi: 10.32604/iasc.2024.043020.
[15] J. S. Valderama, "Profile Variables Of High And Low Performing Schools: Discriminating Factors Of Mathematics Performance," *J. Comput. Innov. Anal.*, vol. 1, no. 2, pp. 91–110, 2022, doi: 10.32890/jcia2022.1.2.5.
[16] C. Azzahra and Sriani, "Clustering of High School Students Academic Scores Using the K-Means Algorithm," *J. Inf. Syst. Informatics*, vol. 7, no. 1, pp. 572–586, 2025, doi: 10.51519/journalisi.v7i1.1029.
[17] M. Mustakim, D. N. Aini, A. U. Batubara, M. Erkamim, and L. Legito, "Fuzzy Clustering-Based Grouping for Mapping the Distribution of Student Success Data," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 366–372, 2023, doi: 10.57152/malcom.v3i2.1227.
[18] A. Asroni, D. Kurniasari, and S. A. Kurniasari, "The Implementation of Clustering Method With K-Means Algorithm In Grouping Data of Students' Course Scores at Universitas Muhammadiyah Yogyakarta," *Emerg. Inf. Sci. Technol.*, vol. 1, no. 3, pp. 75–83, 2020, doi: 10.18196/eist.v1i3.13172.
[19] M. A. Al Fauzie, Yuliadi, and J. A. Putra, "Clustering Data Menggunakan Metode K-Means untuk Rekomendasikan Pembelajaran Akademik bagi Siswa Aktif dalam Ekstrakurikuler," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 642–648, 2023, doi: 10.30865/klik.v4i1.1116.
[20] S. Balovsyak and H. Kravchenko, "Clustering Students According to their Academic Achievement Using Fuzzy Logic," *Int. J. Mod. Educ. Comput. Sci.*, vol. 15, no. 6, pp. 31–43, 2023, doi: 10.5815/ijmecs.2023.06.03.
[21] R. Jayasree and N A Sheela Selvakumari, "Analyzing Student Performance using Fuzzy Possibilistic C-Means Clustering," *Indian J. Sci. Technol.*, vol. 16, no. 38, pp. 3230–3235, 2023, doi: 10.17485/IJST/v16i38.226.
[22] S. A. B. Telaumbanua, F. Setiadi, and S. Nurjanah, "Analisis Clustering Menggunakan Metode Enhanced Fuzzy C-Means Clustering Dengan Algoritma Rock Pada Student Performance Dataset," *bit-Tech*, vol. 7, no. 3, pp. 984–994, 2025, doi: 10.32877/bt.v7i3.2287.
[23] H. Han, "Fuzzy clustering algorithm for university students' psychological fitness and performance detection," *Heliyon*, vol. 9, no. 8, pp. 1–14, 2023, doi: 10.1016/j.heliyon.2023.e18550.