Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

Identifying Academic Excellence: Fuzzy Subtractive Clustering of Student Learning Outcomes

Muhammad Bagas Satrio Wibowo*[1], Kartika Maulida Hindrayani[2], Trimono[3]

[1,2,3]Data Science, University of Pembangunan Nasional "Veteran" East Java, Indonesia

Email: [1]21083010071@student.upnjatim.ac.id, [2]kartika.maulida.ds@upnjatim.ac.id, [3]trimono.stat@upnjatim.ac.id

Received : Apr 15, 2025; Revised : Jun 25, 2025; Accepted : Jun 26, 2025; Published : Dec 11, 2025

Abstract

Education forms a vital foundation for a nation's future. In this digital era, while the use of Information and Communication Technology (ICT) in education is increasing, it brings increasingly complex challenges in education data management and analysis. The growing number of students each year results in a large volume of data, which would be difficult to manage if still relying on manual methods. Manual approaches are inefficient, time-consuming, prone to inconsistencies and human error, especially when identifying outstanding students in large and complex data. This research aims to implement a clustering system to group outstanding students at XYZ elementary school using the Fuzzy Subtractive Clustering (FSC) method. FSC was chosen for its ability to identify data groups based on the density of data points. FSC involves several important parameters, including radius, squash factor, acceptance ratio, and rejection ratio. Added variabel of social and spiritual values aims to enhance grouping quality by offering a broader perspective on students' character, attitudes, and social interactions. Parameter exploration shows an increase in the silhouette score from 0.20–0.45 to 0.45-0.57 and variable addition spiritual and social values, which indicates clearer cluster separation and provides better insights. The best parameters results were achieved with radius 0.3, accept ratio 0.5, reject ratio 0.04, and squash factor 1.25, resulting in a Silhouette Score of 0.57 and forming 5 student groups. Cluster results can guide special mentoring for students with low academic, spiritual, and social values, and support personalized learning programs based on each cluster's characteristics.

Keywords : *Cluster, Fuzzy Subtractive Clustering, Students.*

1. INTRODUCTION

Education is an important and fundamental component of human life [1]. Besides being the foundation for the development of every individual, education also has a crucial role in shaping the nation's next generation of quality. This makes education a long-term investment for the nation's progress [2]. In Indonesia, the evaluation of student learning achievements through academic grades is a significant indicator in assessing the effectiveness of the education system [3]. However, the management of large and complex student grade data is often underutilized. This underutilization leads to difficulties in the early identification of potential students, targeted planning of intervention programs, and the overall evaluation of educational program effectiveness [4].

The rapid progress of Information and Communication Technology (ICT) in the contemporary era requires its strategic integration into the transformation of the education sector [5]. ICT presents a spectrum of opportunities, potentials, and challenges for educational innovation, facilitating a paradigm shift towards a progressive learning paradigm [6].

According to Dapodik records, the total of students in the 2019/2020 was around 44.69 million, while in the 2024/2025 it reached around 52.14 million. This data indicates a notable increase in student participation within the Indonesian education system [7]. The growing amount of data makes manual

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

data management increasingly inefficient and prone to errors. Consequently, an automated approach is needed to overcome the complexity of data management[8]. The amount of information available has the potential to generate valuable and relevant data insights for strategic decision making [9].

One of them is at XYZ Elementary School, an educational institution in Sidoarjo with 262 students and 15 teachers. The process of searching for outstanding students is still done manually, which tends to be less effective, takes a long time, and risks causing bias or human error [10]. In addition, the variation in grades between students makes the process of grouping based on achievement difficult without the support of a systematic data analysis method. Each teacher also records student grades manually in the report card and evaluates learning outcomes individually in their respective classes with an understanding of student achievement limited to the classroom teacher's perspective, without a comprehensive evaluation involving all teachers and the principal [11]. This learning outcome is very important as a basis for school planning and sustainable development of education quality. Therefore, student learning outcomes must be conducted fairly, objectively and openly so that all interested parties have a clear understanding of student achievement [12]. The lack of available data analysis systems results in the unavailability of comprehensive data for objective assessment and comparison between students. This hinders the school's efforts to identify overall student achievement patterns and plan effective education quality improvement programs [13]. Identifying outstanding students is crucial not only for recognizing academic excellence but also for guiding targeted educational support and maximizing student potential. By accurately identifying outstanding students, educators can provide enrichment programs, leadership opportunities, and personalized mentoring that foster continued growth [14]. Early identification helps schools develop role models who can inspire their peers and contribute to a positive learning environment. In the broader context, supporting outstanding students ensures that their talents are nurtured effectively[15].

This condition is that XYZ elementary school needs to implement a system for analyzing student learning outcomes data, because evaluating student learning outcomes is the key to improving the quality of education [10]. In this case, Fuzzy Subtractive Clustering (FSC) is applied, which plays a role in processing data more effectively, so that it can be a tool for educators in utilizing existing information to the maximum[16] .

The research conducted by [17] aims to group students based on report card grades using the K-Means Clustering method, as well as determining top students (top rank) among high-achieving groups with the help of the Simple Additive Weighting (SAW) method. The limitations in this research are the very limited amount of data, which is only 25 records, and only using the grade variable as the basis for grouping, so it does not reflect other aspects that also affect student achievement such as attitude or social factors.

Similar research conducted also by [18] aims to find out and form student data clusters using the k-means method based on academic grades, attitude scores, and discipline scores so that they become a cluster so that the results of student clusters can be a reference in improving student grades in the next learning process. The best result with the application of the K-Means Clustering algorithm produces 3 clusters with a silhouette score of 0.489. Limitations of this research include the use of only one final grade variable without a detailed explanation of the type or source of the grade, making interpretation of the results difficult. The imbalance in the number of members between clusters, especially from 155 data in only cluster 2 which contains only one student, also shows potential inaccuracies in data separation.

Research conducted by [19] find out which algorithm is more effective between K-Means and DBScan in grouping junior high school student data based on their academic achievement. The results showed that the DBSCAN algorithm with epsilon 11 and 24 data samples produced 2 clusters with a Silhouette Score of 0.258030877243884. While K-Means algorithm with Elbow method produces 4

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

clusters with Silhouette Score of 0.5697019340266847. From the comparison, K-Means algorithm proved to be more efficient. K-Means also shows clusters that are cleaner, structured, and have better similarity between data. The limitation of this research is that there has been no further exploration of validation techniques or performance comparisons with other fuzzy clustering algorithms that are potentially more suitable for the characteristics of the data used.

Research conducted by [20] applying Fuzzy Subtractive Clustering (FSC) for imputation of medical data. The data used is hypertension data. The variables used are age, gender, systolic pressure, diastolic pressure, and body weight. This study produced a Partition Coefficient (PC) value of 0.5369 for 2 clusters, 0.4801 for 3 clusters, and 0.5473 for 4 clusters. The limitation of this research is the simulation process that only focuses on the radius parameter (r) in the formation of the number of clusters, so it has not explored the influence of other important parameters such as squash factor (q), accept ratio, or reject ratio.

A similar study [21] applied Fuzzy Subtractive Clustering (FSC) to impute medical data of heart failure patients. The results showed that the most optimal number of clusters was 3, which was selected based on the most significant Partition Coefficient (PC) value, namely PC = 0.7393. The limitations of this research still focus on the use of one type of dataset that has been widely explored before, so the generalization of the results is still limited. It is necessary to apply this Fuzzy Subtractive Clustering method to datasets from different fields or domains to test its consistency and reliability more broadly.

As well as research conducted by [22] This research aims to prove the performance of Fuzzy Subtractive Clustering (FSC) and Fuzzy C-Means (FCM) methods for solving imputation problems. This research is limited by the need to justify some clustering parameters, including acceptance/rejection ratios, as with other soft computing techniques. Therefore, to determine the ideal parameters, testing should be conducted.

Previous research generally used the K-Means method as the most common method of data grouping because of its simplicity. However, the K-Means grouping method has significant drawbacks, especially its sensitivity to the initial selection of the cluster centers [23]. To overcome this limitation, the FSC method was chosen because in identifying the cluster center, FSC is more adaptive based on data density, and emphasizes the importance of optimizing parameters such as radius and acceptance ratio. However, some of these studies acknowledge limitations in the exploration of several other important parameters, including radius, squash factor, acceptance ratio, and rejection ratio. Therefore, the research explores the determination of ideal parameters, which is done by utilizing soft computing techniques [24].

The application of Fuzzy Subtractive Clustering (FSC) is expected to group data more effectively, so that it can be a tool for educators to make maximum use of existing information. This method's ability to create clusters based on the maximum permitted distance between cluster members and the cluster center is one of its benefits. The cluster center is a component of the clustered data in this method [25].

Based on the background that has been explained, this study aims to group outstanding students based on report card grades by applying the Fuzzy Subtractive Clustering (FSC) method. As an innovative approach, this research application of Fuzzy Subtractive Clustering (FSC) integrated with parameter exploration carried out specifically tailored to the dataset, specifically to identify outstanding students based on report card grades.

2. METHOD

This research utilizes primary data sourced from SDI XYZ. Python and JupyterLab were used as the main software and programming language for the analysis. The research flow, starting from data collection to model evaluation is illustrated in detail in Figure 1.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

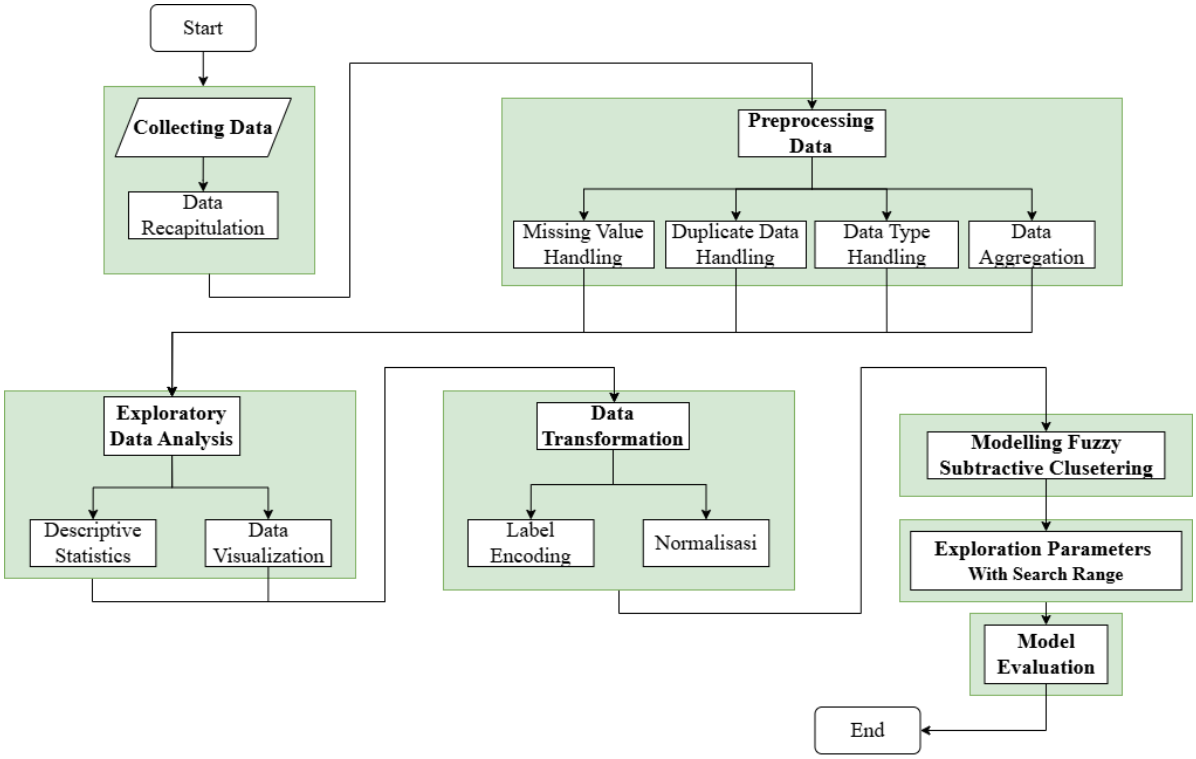

Figure 1. Research Flow

2.1. Collecting Data

The research data was obtained directly from SDI XYZ through data recapitulation by manually recording each student's report card into excel. The dataset used was primary data consisting of report cards of students in grades 1 to 6, with a total of 260 students and 16 research variables in the 2021/2022 school year. The details of the features used are presented in Table 1.

Table 1. Details of the features

Feature	Description
Name	Student's full name
Class	Class
P_Agama	Religious Education (Knowledge score)
Ket_Agama	Religious Education (Skill score)
P_PPKN	Civic Education (Knowledge score)
'Ket_PPKN	Civic Education (Skill score)
'P_B.Indonesia'	Indonesian Language (Knowledge score)
'Ket_B.Indonesia'	Indonesian Language (Skill score)
'P_Matematika'	Mathematics (Knowledge score)
'Ket_Matematika'	Mathematics (Skill score)
'P_IPA'	Natural Science (Knowledge score)
'Ket_IPA'	Natural Science (Skill score)
'P_IPS'	Social Science (Knowledge score)
'Ket_IPS'	Social Science (Skill score)
'P_B.Inggris'	English Language (Knowledge score)
'Ket_B.Inggris'	English Language (Skill score)
'N_ Spiritual'	Spiritual Value (Competence in spiritual attitude including acceptance, application, and appreciation of religious teachings)
'N_Sosial'	Social Score (Competency in social such as honesty, discipline, responsibility, politeness, care, and self-confidence in interactions)

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

2.2. Preprocessing Data

Data preprocessing is the process of transforming, cleaning, and integrating raw data into a format suitable for analysis [26]. Data preprocessing is also useful for ensuring consistency in the dataset, so that modeling tasks can be performed and further analysis can be done effectively [27]. This stage includes several stages which will be carried out as follows:

a. Handling missing values
b. Handling duplication data
c. Aggregation of skill value data (40%) and knowledge (60%)

2.3. Exploratory Data Analysis

Exploratory Data Analysis (EDA) in this research is conducted through descriptive statistics and data visualization, the goal is to uncover hidden data structures, discover new insights, identify anomalies or outliers, to build efficient models [28]. The variables that EDA is performed in are average final subject grade, grade and number of students.

2.4. Data Transformation

Data transformation converts data into a format that is suitable and needed for analysis and includes the label encoding [29]. The encoding label to be applied is that categorical data must be converted into numerical data so that it can be processed by the model, because the model cannot process categorical data [30]. Label encoding will be applied to the social and spiritual value variables as a data transformation step to facilitate further processing.

2.5. Modeling Fuzzy Subtractive Clustering

The method in this study focuses on the application of a clustering model using the Fuzzy Subtractive Clustering (FSC) method. This method starts with normalization to calculate the data density around each point in the dataset using parameters such as radius, squash factor, rejection ratio, and acceptance ratio to adjust the sensitivity of the model to the data [31]. The details of the FSC algorithm [32] used are presented in Table 2.

Table 2. FSC Algorithm

FSC Algorithm
Input: Dataset (x), Radius (r), Squash factor (q), Accept ratio, Reject ratio, Maximum Value, Minimum Value Output: Cluster Center (C), Cluster data
a. Normalize (x) data point is between (0 and 1) based on equation (1). b. Calculate the initial potential data of the cluster center based on equation (2) or (3). Select one data point with the highest potential to become the first cluster center. c. Reduce the potential of each potential data point and determine the next cluster center with equations (5) - (8). d. Check the conditions of each potential cluster center result. If condition 1 (ratio > accept ratio) continues to be iterated, start at step c. If condition 2 (ratio value < accept ratio and ratio value > reject ratio) , calculate Mds (minimum distance to the segment) with equation (6), (8)-(9). If condition 3 (ratio value < accept ratio and ratio value < reject ratio) iteration will be stopped because no more cluster center candidates have been found. e. Calculate the sigma of each cluster center using equations (10) and (11). f. Calculate the degree of membership of each data with equation (12) g. Determine the location of the cluster with the degree of membership obtained

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

a. Data normalization aims to homogenize data of the same scale in the range of 0 to 1 [33], so that no variable dominates the clustering results. This process is very important because FSC uses the density measure of data points. Normalization calculation uses min-max normalization in equation (1).

$$x_{ijnorm} = \frac{x_{ij} - x_{jmin}}{x_{jmax} - x_{jmin}} \quad (1)$$

In equation (1) x_{ij} is the original value of the data in the-*i* row and the-*j* variable, with x_{jmin} being the minimum value of the-*j* variable and the x_{jmax} is maximum value of the-*j* variable.

b. Calculate the initial potential data of the cluster center based on equation (2). Select one data point with the highest potential to become the first cluster center.

$$D_k = \sum_{k=1}^{n} e^{-4\left((Dist_{kj})^2\right)} \quad (2)$$

T_j is the value of the reference data point (center or target point) in the-*j* dimension, X_{kj} is the value of the-*k* data point in the-*j* dimension. Where r is the radius of the value previously set. From the results of the D_k calculation, the highest potential was chosen to be the first cluster center and was used as the initial Z value.

c. Reduce the data potential of each point based on equations (3)-(7).

$$R = \frac{Z}{M}, \ Only\ iteration\ 1\ is\ Z = M \quad (3)$$

$$Z = Max[D_i^t |\ i = 1,2,3 \ldots., n] \quad (4)$$

Initialize Z or M based on the previous highest potential value. In equation (4) and (5), the ratio value obtained is denoted by R, while the highest potential point value for the first iteration is M. For the second iteration and so on, the highest potential point value is represented by Z.

$$D_{c_{l_i}} = \ \mathrm{M} \times e^{-4(\sum_{j=1}^{m} Dist_{ij}(S_{ij})^2)} \quad (5)$$

Continue by calculating the value of $D_{c_{l_i}}$ as the potential reduction value for each previous data point so that the calculation results will be used to calculate the new potential. The description in the equation (5) explains the formula for calculating the potential of data in the-*l* cluster for the-*i* sample, denoted as $D_{c_{l_i}}$. This equation uses the variable M which represents the highest potential data from the first iteration. The symbol s_{ij} which indicates the reduction in potential data of the-*i* sample at the-*j* attribute.

$$D_i^t = D_i^{t-1} - D_{c_{l_i}} \quad (6)$$

Then the new potential is calculated. The new potential of the data D_i^t is the old potential D_i^{t-1} minus the reduction in the potential of the data $D_{c_{l_i}}$. The description in the equation (6) explains the formula for calculating the new potential data for the-*i* sample at iteration *t*, denoted as D_i^t. this is calculated by subtracting the potential data of the-*l* cluster from the sample at iteration *t*, represented as $D_{c_{l_i}}$. From the potential data of the-*i* sample in the previous iteration, denoted as D_i^{t-1}. The potential for finding potential cluster centers is calculated using equation (5) and then the results are continued with equation (6) so that new potential is obtained. Prospective cluster

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

centers are selected based on the highest value D_i^t, while values D_i^t less than 0 the potential data will be set to 0 and will no longer be considered as cluster centers.

d. The conditions of 1 candidate for a cluster center are checked if the ratio value is > accept ratio, the candidate point for the cluster center is accepted as a new cluster center and is labeled as C_i. The steps will be repeated as in the previous procedure is to look for a candidate cluster center in equations (3) to (6).

Condition 2 If the ratio value is < accept ratio and the ratio value is > reject ratio, new candidates will be accepted as cluster centers if they are located far enough from existing cluster centers. Procedure in this condition Md = -1, continue with Equations (7)-(8).

$$Sd_i = \sum_{j=1}^{m} \left(\frac{V_j - C_{lj}}{r}\right)^2 \quad (7)$$

$$Mds = \sqrt{Md} \quad (8)$$

The description in the equation (8) explains the formula for calculating Sd_i, which is the sum of squared distances between the candidate cluster center V_j and the-l cluster center C_{lj}, which is divided by the r (radius) of each data attribute. If $Md < 0$ or $Sd < Md$, then $Md = Sd_i$. Continued with if ≥ 1 $(Ratio + Mds) \geq 1$ then the data is accepted as the center of the cluster, while if $(Ratio + Mds) < 1$ then the data will not be reconsidered as the center of the cluster and the value set to 0. Then it will look for other potential highest points, continue with iterations in equations (5), (7) - (8).

Condition 3 If the ratio value is < accept ratio and the ratio value is < reject ratio, the iteration will be stopped because no more cluster center candidates have been found. So the search for the cluster center will stop if it reaches condition 3.

e. Calculate the sigma of each cluster center using equations (9) and (10). After no cluster centers are found, the sigma calculation is continued to find the degree of membership to be calculated.

$$C_{lj_{denorm}} = C_{lj} \times \left(x_{\max j} - x_{minj}\right) + x_{minj} \quad (9)$$

To obtain this value, the normalized cluster center C_{lj} is multiplied by the difference between the maximum value $x_{\max j}$ and the minimum value x_{minj} of the-j attribute. Then, the minimum value of the data for that attribute x_{minj} added to restore the cluster center to its original data scale. The description in the equation (10) explains the formula for calculating $C_{ljdenorm}$ represents the cluster center for the-k cluster at the-j attribute after denormalization.

$$\sigma_j = r_j \times \frac{x_{jmax} - X_{jmin}}{\sqrt{8}} \quad (10)$$

Specifically, σ_j represents the standard deviation of the cluster, which is determined by multiplying the radius of each data attribute r_j, with the difference between the maximum value x_{jmax} and the minimum value X_{jmin} of the-j attribute. The description in the equation (10) explains the formula for calculating σ_j.

f. Degree of membership is calculated using the Gauss function at equation (11), this degree of membership will be used to determine whether the data is included in the cluster based on the largest one.

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

$$\mu_{kj} = e^{-\sum_{j=1}^{m} \frac{(X_{ij} - C_{lj})^2}{2\sigma_j^2}} \quad (11)$$

The equation (12) calculates the degree of membership μ_{kj} of the-k cluster for the-i sample. It is determined by using the exponential function, where the squared difference between the sample data X_{ij} and the cluster center C_{lj} is divided by twice the squared standard deviation σ_j^2.

g. Determine the location of the cluster using the membership degree value obtained previously. The location of the data cluster is where the cluster with the highest membership degree value in the data.

2.6. Parameter Exploration

To find the best parameter configuration, the modeling process is conducted by exploring the initialization of key parameters, including the radius, squash factor, acceptance ratio, and rejection ratio, in order to obtain optimal evaluation results. The parameter ranges in the optimization process of the Fuzzy Subtractive Clustering (FSC) method were selected based on standard values commonly used in previous studies, such as squash factor of 1.25, accept ratio of 0.5, and reject ratio of 0.15. Furthermore, an empirical exploration was conducted to assess whether the performance of the model can be improved by using parameter values that are smaller or larger than these standards. The aim is to evaluate whether the variation in values produces better clustering results compared to the use of standard values [34].

In Fuzzy Subtractive Clustering (FSC), two comparative thresholds are used: the accept ratio and the reject ratio, both of which are fractional values ranging from 0 to 1. The accept ratio serves as a lower threshold, meaning a data point can be accepted as a potential cluster center if its potential exceeds this value. Conversely, the reject ratio serves as an upper threshold, indicating that a data point cannot be considered a cluster center if its potential falls below this value. The squash factor is a constant used to determine the extent of the potential suppression around an accepted cluster center, thereby influencing the selection of subsequent centers. Meanwhile, the radius is a vector that defines the neighborhood influence of a cluster center on surrounding data points [20].

2.7. Model Evaluation

The Silhouette method basically evaluates how well an object is integrated into the cluster in which it belongs. It states that for every object in a cluster[35]. Silhouette Score is a clustering model evaluation technique that has a value range between -1 to 1 [36]. The weakness of using the silhouette coefficient is that to obtain its optimal value it is more suitable for methods such as those that must be run repeatedly for various values of k (number of clusters). The higher silhouette score,is the better the cluster. The details of the silhouette score used are presented in Table 3.

Table 3. Silhouette score details

Intepretation	Silhouette score
Strong cluster structure	0,71 – 1
Good cluster structure	0,51 – 0.70
Weak cluster structure	0,26 – 50
Bad cluster structure	0 – 0,25

Silhouette Score is a metric to assess how well objects in one cluster are separated from objects in other clusters [37].

$$S_i \frac{b_i - a_i}{\max(b_i, a_i)} \quad (12)$$

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

From equation (12) is the equation for calculating the silhouette score value, with a_i reference to the average distance between one point and all data in the same cluster, while b_i referring to the smallest average distance between one point and points in different clusters.

3. RESULT

3.1. Collecting Data

Table 4 shows the results of the dataset, where each data has the same format. The total number of data is 260 rows with 18 columns. The details of the features used are presented in Table 4.

Table 4. Datasets Details

Nama	Kelas	P_Agama	K_Agama	P_PPKN	K_PPKN	P_B.Indo	K_B.Indo	P_Math	K_Math	...	Spiritual	Sosial
Siswa 1	1B	90	88	91	92	95	82	93	90	...	B	B
Siswa 2	1B	85	84	86	83	90	90	80	85	...	B	B
Siswa 3	1B	72	75	73	76	72	75	72	74	...	B	B
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Siswa 260	2A	95	94	95	95	95	95	90	92		B	A

3.2. Preprocessing

Data preprocessing refers to the process of converting raw data into a format suitable for analysis, the results of which are shown in Table 5. The following are the results of the steps involved in preprocessing, which are outlined for a more comprehensive explanation:

a. Handling missing value, the number of missing values found is = 0 which means there are no missing values in the dataframe. All columns have complete data.
b. Handling duplicate data, there are no duplicate rows in the dataset. Each row is unique, which is a positive sign for data quality, as this means there are no repetitive entries that can break analysis or modeling.
c. The results of data aggregation are obtained from the value of skills (40%) and knowledge (60%) in each variable value of each subject. After aggregating the dataset, the results were taken into 11 features including ‘Nama', 'Kelas', 'NA_Agama', 'NA_PPKN', 'NA_B.Indonesia', 'NA_Matematika', 'NA_IPA', 'NA_IPS', 'NA_Bhs.Inggris', 'Nilai_Spiritual', dan 'Nilai_Sosial'.

Table 5. Preprocessing results

Nama	Kelas	NA_Agama	NA_PPKN	NA_B.Indo	NA_Math	NA_IPA	NA_IPS	NA_B.Inggris	Spiritual	Sosial
Siswa 1	1B	89.2	91.4	89.8	91.8	92	93	90	B	B
Siswa 2	1B	84.6	84.8	90	82	88	90	79.2	B	B
Siswa 3	1B	73.2	74.2	73.2	72.8	77.2	77.8	70.6	B	B
Siswa 4	1B	89.2	87.4	90.8	90.8	86.8	85	84.4	B	B
Siswa 5	1B	90.8	89.8	90.8	90	90.8	91.2	80	B	B
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

Siswa 260	2A	94.6	95	95	90.8	91.8	89.4	94.6	B	A

3.3. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an essential step in understanding the underlying patterns, structures, and relationships within a dataset before conducting further analysis. The following are the EDA steps that have been carried out, with a detailed explanation of the procedures performed:

a. The descriptive statistics

The descriptive statistics presented provide an overview of the distribution of final grades (NA) These data show that the majority of students have good grades, with averages above 80 for all subjects. The results of the descriptive statistics are shown in Table 6.

Table 6. Preprocessing results

	NA_Agama	NA_PPKN	NA_B.Indo	NA_Math	NA_IPA	NA_IPS	NA_B.Inggris
Count	260	260	260	260	260	260	260
Mean	86.2	86.8	87.3	85.4	86.3	86.4	83.9
Min	70	69.4	70	67	70	70	66.2
Max	100	96.2	99.2	98	97.2	96	96

b. Data visualization

As shown in Figure 2, The boxplot visualization shows that students generally scored in the high range (85-90) in all subjects, with Indonesia language having the highest mean and median, indicating that students performed best in that subject. English language having the lowest mean and median, indicating that students are having difficulty. The imbalance between the mean and median in mathematics and english language also indicates a skewed distribution of scores, this can be a reference for educators to give more attention through guidance or remedial programs in subjects with low score distribution.

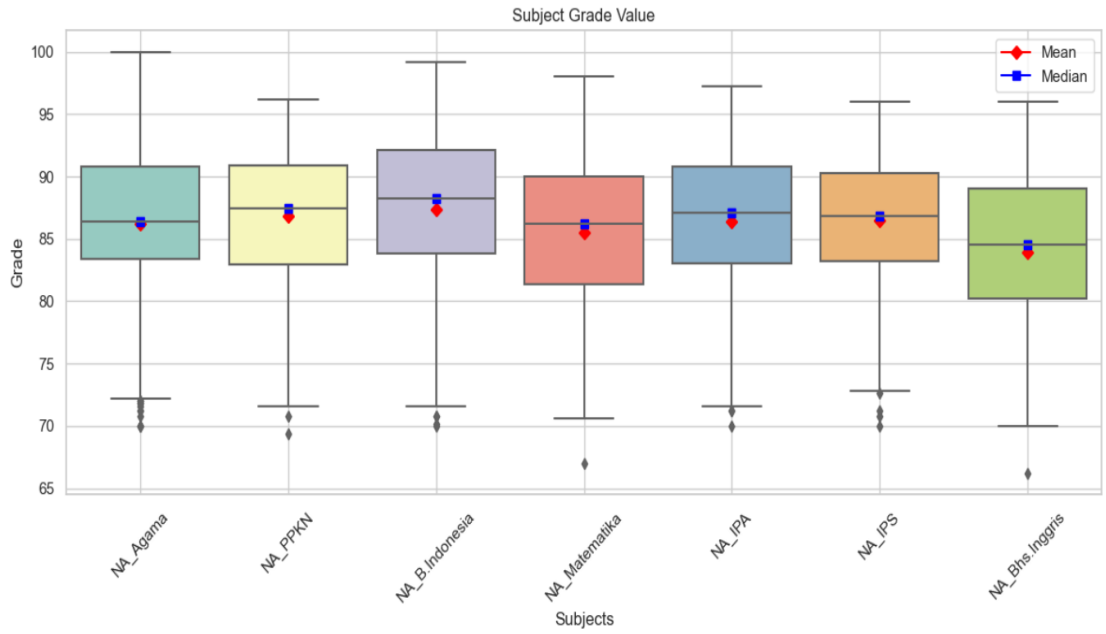

Figure 2. Visualization of Final Subject Grades

From the heatmap in Figure 3 based on the correlation heatmap, there are several ways to interpret Spearman correlation values. A correlation value below 0.4 is considered to indicate a weak relationship.

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

A value between 0.4 and 0.7 represents a moderate correlation [38]. The grades of each subject have a moderate correlation with students' spiritual and social grades. Overall, this shows that although there is a correlation between academic achievement and student social or spiritual grades, the correlation is not very strong, so spiritual and social grades do not fully reflect student academic achievements in each subject.

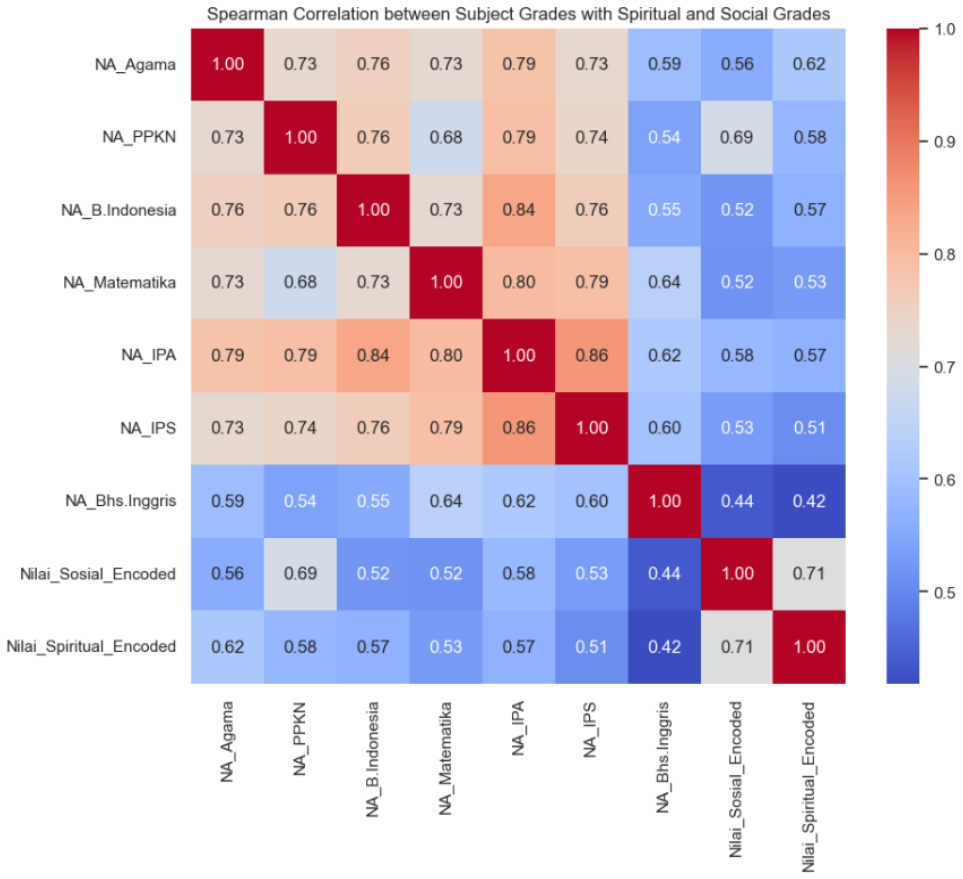

Figure 3. Visualization of Correlation

3.4. Data Transformation

This encoding method was applied to transform the categorical data into numerical form, facilitating its use in further statistical analysis and modeling. Table 7 illustrates the label encoded values for the Social and Spiritual Values, where the categorical values have been assigned as follows: ‘A’ = 2, ‘B’ = 1, and ‘C’ = 0.

Table 7. Encoding result

Grade	Social Grade Encoded	Count Social Grade	Spiritual Grade Encoded	Count Spiritual Grade
A	2	54	2	48
B	1	201	1	208
C	0	5	0	4

3.5. Modeling Fuzzy Subtractive Clustering

Modeling experiments were conducted using the parameters accept ratio of 0.5, reject ratio of 0.04, radius of 0.3, and squash factor of 1.25. First, data normalization is performed, with the results shown in Table 8.

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

Table 8. Normalization Result

Nama	Agama	PPKN	B.Indonesia	...	Spiritual	Social
Siswa 1	0.892	0.95	0.90	...	0.5	0.5
Siswa 2	0.846	0.88	0.90	...	0.5	0.5
Siswa 3	0.732	0.77	0.73	...	0.5	0.5
Siswa 4	0.892	0.90	0.91	...	0.5	0.5
Siswa 5	0.908	0.93	0.91	⋮	0.5	0.5

Calculate the potential of the initial data center and select one data point with the highest potential to be the center of the first cluster obtained in the 160th data with a value = 99.21, with the results shown in Table 9.

Table 9. Initial potential data result

Data	Initial potential
1	57.23
2	78.28
3	19.06
4	83.55
5	65.74
⋮	⋮
160	99.21
⋮	⋮
260	6.48

Subtract the potential of each potential data point and determine the next cluster center, leaving no check on the condition of each potential cluster center result. The results shown in Table 10.

Table 10. Each potential iteration

Data	Initial potential	New Potential 1	...	New Potential 4	New Potential 5
1	57.35	13.90	...	0	0
2	78.34	0	...	0	0
3	19.07	11.84	...	0	0
4	83.28	5.34	...	0	0
5	65.84	2.86	...	2.29	2.6
⋮	⋮	⋮	⋮	⋮	⋮
260	6.48	5.90		0	0.03

The cluster center is found after calculating the potential of each point. The results of the calculation found 5 cluster centers, details of the cluster center are shown in Equation 13.

$$C_{lj_{denorm}} = \begin{bmatrix} 85.8 & 86.8 & 87 & 84 & 85.2 & 85.4 & 81.8 & 1 & 1 \\ 92 & 93.6 & 93.2 & 93 & 92.8 & 92.8 & 89.2 & 2 & 2 \\ 73.2 & 77.2 & 75.4 & 75 & 74.6 & 74.8 & 74 & 1 & 1 \\ 90.6 & 92.6 & 88.2 & 88.8 & 90.8 & 90 & 82 & 2 & 1 \\ 92.6 & 90 & 94.4 & 92 & 92.6 & 90 & 88.6 & 1 & 2 \end{bmatrix} \quad (13)$$

Calculate the sigma of each cluster center, details of the calculate the sigma are shown in Equation 14.

$$\sigma = [10.60 \quad 10.20 \quad 10.52 \quad 10.39 \quad 10.30 \quad 10.18 \quad 10.18 \quad 0.21 \quad 0.21] \quad (14)$$

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

Calculate the degree of membership of each data with equation, details of the calculate the sigma are shown in Table 11.

Table 11. Degree of membership result

Data	Degree to cluster 0	Degree to cluster 1	Degree to cluster 2	Degree to cluster 3	Degree to cluster 4
1	0.27	0.000	0.000	0.000	0.000
2	0.77	0.000	0.016	0.000	0.000
3	0.016	0.000	0.80	0.000	0
4	0.68	0.000	0.003	0.000	0.000
5	0.49	0.000	0.000	0.000	0.000
⋮	⋮	⋮	⋮	⋮	⋮
260	0.000	0.005	0.000	0.33	0.000

Determine the location of the cluster using the membership degree value obtained previously. The location of the data cluster is where the cluster with the highest membership degree value in the data, with the results shown in Table 12.

Table 12. Degree of membership result

Data	Cluster
1	1
2	1
3	3
4	1
5	1
⋮	⋮
260	4

Followed by the calculation of the silhouette score obtained is 0.57 with the number of clusters as many as 5. These results indicate that the results of clustering good quality data.

3.6. Exploration Parameters

The test results are followed by various parameter configurations in Fuzzy Subtractive Clustering (FSC) to determine the best combination that produces the optimal number of clusters. From the test parameters, the best combination that produces the most optimal number of clusters with good quality silhouette score. For the FSC algorithm, an optimal parameter range search was performed for four parameters that affect the performance model. The ranges of values used in this parameter search are shown in more detail in Table 13 below.

Table 13. Search Parameter

Parameter	Range Search Parameter	Parameter Details
Radius	0.3 – 0.18	[0.3, 0.27, 0.25, 0.2, 0.18]
Squash factor	1.25 – 0.55	[1.35, 1.25,1.15, 1.05, 0.53]
Accept ratio	0.6 – 0.3	[0.6, 0.5, 0.4, 0.3]
Reject ratio	0.2 – 0.03	[0.2, 0.1, 0.07, 0.04, 0.03]

Search for radius configuration with accept ratio = 0.5 and reject ratio = 0.15. Details of the radius exploration used are presented in Table 14.

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

Table 14. Radius exploration FSC Standart parameter

Radius	Squash Factor	Jumlah Cluster	Silhouette score
0.3	1.25	2	0.62
0.27	1.25	3	0.45
0.25	1.25	3	0.43
0.2	1.25	4	0.23
0.18	1.25	5	0.20

Search for squash factor configuration with accept ratio = 0.5 and reject Ratio = 0.15. Details of the squash factor exploration used are presented in Table 15.

Table 15. Squash factor exploration

Radius	Squash Factor	Jumlah Cluster	Silhouette score
0.3	1.35	2	0.62
0.3	1.25	2	0.62
0.3	1.15	3	0.45
0.3	1.05	4	0.31
0.3	0.53	5	0.17

Search for accept ratio configuration with radius = 0.3 and squash factor = 1.25. Details of the accept ration exploration used are presented in Table 16.

Table 16. Accept ratio exploration

Accept Ratio	Reject Ratio	Jumlah Cluster	Silhouette score
0.6	0.15	2	0.62
0.5	0.15	2	0.62
0.4	0.15	2	0.62
0.3	0.15	2	0.62

Search for reject ratio configuration with radius = 0.3 and squash factor = 1.25. Details of the accept ration exploration used are presented in Table 17.

Table 17. Reject ratio exploration

Accept Ratio	Reject Ratio	Jumlah Cluster	Silhouette score
0.5	0.2	2	0.62
0.5	0.1	3	0.45
0.5	0.07	4	0.51
0.5	0.04	5	0.57
0.5	0.03	6	0.53

3.7. Model Evaluation

The test results are shown in Table 18 with the best silhouette score of each number of clusters, the best silhouette score value obtained is 0.62 but the number of clusters formed is limited to 2 clusters. However, the best parameters chosen in Fuzzy Subtractive Clustering are accept ratio 0.5, reject ratio 0.04, radius 0.3, and squash factor 1.25. These parameters were chosen because they produce a silhouette score of 0.52 which is quite good with a more diverse number of clusters, namely 5 clusters.

Table 18. Best parameters each number of clusters exploration

Radius	Squash Factor	Accept Ratio	Reject Ratio	Jumlah Cluster	Silhouette score

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

0.3	1.25	0.5	0.2	2	0.62
0.3	1.25	0.5	0.1	3	0.45
0.3	1.25	0.5	0.07	4	0.51
0.3	1.25	0.5	0.04	5	0.57
0.3	1.25	0.5	0.03	6	0.53

The distribution of grade point averages is based on clustering results consisting of five clusters. Cluster 2 has the highest average score, indicating the group with the best academic performance. Cluster 1, 4 and Cluster 5 are at the middle level, while Cluster 3 has the lowest average score, indicating relatively lower performance compared to the other clusters. Details of cluster distribution of grade point averages presented in Figure 4.

The difference in spiritual values between clusters, where Clusters 2 and 5 have the highest spiritual values, followed by Cluster 1 and Cluster 4 with lower values. Cluster 3 showed the lowest average spiritual value. Details of cluster distribution of spiritual values presented in Figure 4.

Details of cluster distribution of spiritual values presented in Figure 4. The difference in social values between clusters, where Clusters 2 and 4 have the highest social values, followed by Cluster 1 and Cluster 5 with lower values. Cluster e showed the lowest average spiritual value. Details of cluster distribution of social values presented in Figure 4.

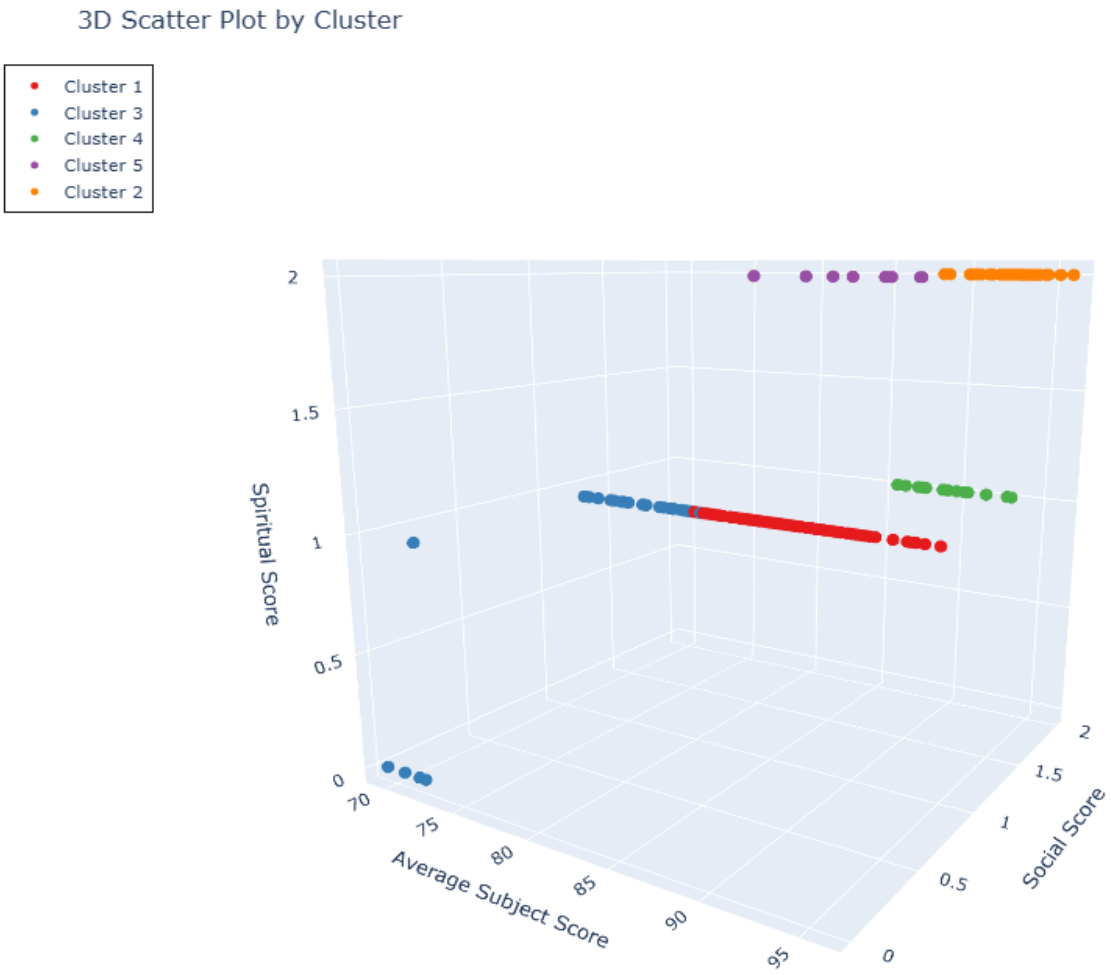

Figure 4. Cluster visualization

The clustering results in Table 19 successfully identified groups of students with different characteristics in terms of academic achievement, spiritual values, and social values. Cluster 2 with 38

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

students stood out as having the highest academic scores and high spiritual and social scores. In contrast, Cluster 3 with 33 students showed the lowest performance in all three aspects (academic, spiritual and social). Cluster 4 with 16 students had average academic performance, high social scores, but average spiritual scores. Cluster 5 with 10 students was also at an intermediate academic level, high spiritual scores, but average social scores. Finally, Cluster 1 with 163 students showed medium academic performance with spiritual and social scores that were among the other groups with medium scores.

Table 19. Cluster Profile

	Count	Mean NA_Agama	Mean NA_PPKN	Mean NA_B. Indo	Mean NA_Math	Mean NA_IPA	Mean NA_IPS	Mean NA_B. Inggris	Mean Spiritual	Mean Social
Cluster 1	163	86.24	86.41	87.29	85.56	86.41	86.69	84.15	1	1
Cluster 2	38	93.31	94.03	94.03	92.15	92.63	91.44	89.93	2	2
Cluster 3	33	75.01	77.37	76.78	74.44	75.87	77.12	74.46	0.87	0.84
Cluster 4	16	88.86	92.28	90.05	88.53	90.13	89.91	85.88	1	2
Cluster 5	10	92.46	89.09	92.98	89.76	90.34	89.64	85.5	2	1

4. DISCUSSIONS

In previous research that have been carried out previously, Table 20 shows a comparison of model evaluation values between the proposed model and previous studies using Silhouette Score was carried out. In the research of Hasibuan et al. [20] K-Means produces the highest Silhouette Score of 0.56, the FSC approach with parameter exploration or tuning is able to produce a Silhouette value that is close to the performance of K-Means, although not exceeding it. DBSCAN shows the lowest performance with a Silhouette Score value of 0.25 which indicates that this method is less suitable for the data structure used. Interestingly, from previous research Yudhistira et al.[19] FSC managed to surpass the performance of K-Means. This shows that the different data used indicates that the clustering results are also different based on the characteristics of the data, so it is important to adjust the method and maximize the data preparation process properly. FSC with parameter optimization can be considered as a good approach, better than DBSCAN and quite competitive with K-Means in clustering students. This research shows the FSC method with parameter exploration can provide better results with a silhouette score of 0.52 for 5 clusters. This highlights the importance of parameter optimization in the application of the FSC method to obtain more better and adaptive clustering outcomes.

Table 20. Comparison Of Model

Research	K	K-Means Silhouette Score	FSC + Explor Parameter Silhouette Score	FSC Silhouette Score	DBScan Silhouette Score
Proposed	2	-	0.62	0.62	-
	3	-	0.45	0.45	-
	4	-	0.51	0.43	-
	5	-	0.57	0.23	-
Yudhistira et al.[18]]	3	0.48	-	-	-

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

Hasibuan et al.[19]][	4	0.56	-	-	-
	2	-	-	-	0.25

5. CONCLUSION

Based on the results, it can be concluded that parameter optimization in the Fuzzy Subtractive Clustering (FSC) method produces the best configuration with acceptance ratio 0.5, reject ratio 0.04, radius 0.3, and squash factor 1.25. This parameter configuration proved effective in producing 5 different clusters with a silhouette score of 0.57 indicating good clustering quality and the reject ratio parameter proved to have a considerable impact on the model evaluation results. Result of clustering successfully identified student groups with diverse characteristics in academic achievement, Cluster 2 with 38 students stood out as having the highest academic scores and high spiritual and social scores. In contrast, Cluster 3 with 33 students showed the lowest performance in all three aspects (academic, spiritual and social). Cluster 4 with 16 students had average academic performance, high social scores, but average spiritual scores. Cluster 5 with 10 students was also at an intermediate academic level, high spiritual scores, but average social scores. Finally, Cluster 1 with 163 students showed medium academic performance with spiritual and social scores that were among the other groups with medium scores. The cluster results can be used as a guide for the school in providing special assistance to students with low academic, spiritual and social scores, as well as supporting the implementation of future learning programs. The limitation of this research lies in the lack of automatic parameter exploration in the Fuzzy Subtractive Clustering (FSC) method to obtain optimal parameters directly. In addition, there has been no comparison of FSC performance with other methods such as K-Means to evaluate the effectiveness of each method in the clustering process on the same data. Therefore, this can be a suggestion for future research so that parameter exploration is carried out automatically and comparisons are made with other clustering methods to obtain more comprehensive results.

REFERENCES

[1] Asmana, Y. A. Wijaya, and Martanti, "CLUSTERING DATA CALON SISWA BARU MENGGUNAKAN METODE K-MEANS DI SEKOLAH MENENGAH KEJURUAN WAHIDIN KOTA CIREBON," *JATI(Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 552–559, Sep. 2022, doi: https://doi.org/10.36040/jati.v6i2.5236.

[2] H. Rahmawati, P. Pujiastuti, and A. P. Cahyaningtyas, "Kategorisasi Kemampuan Berpikir Kritis Siswa Kelas Empat Sekolah Dasar di SD se-Gugus II Kapanewon Playen, Gunung Kidul," *Jurnal Pendidikan dan Kebudayaan*, vol. 8, no. 1, pp. 88–104, Jun. 2023, doi: https://doi.org/10.24832/jpnk.v8i1.3338.

[3] S. Athoillah, M. Y. Abu Bakar, and N. Kholis, "Inovasi Penilaian Hasil Belajar Model POT di Era Merdeka Belajar," *TA'DIBUNA: Jurnal Pendidikan Agama Islam*, vol. 7, no. 1, pp. 39–51, Jul. 2024, doi: http://dx.doi.org/10.30659/jpai.7.1.39-51.

[4] Siti Asiyah and Novebri, "Manajemen Peserta Didik dalam Meningkatkan Prestasi Akademik dan Non Akademik Siswa SMPN 1 Lembah Sorik Marapi," *Hikmah : Jurnal Studi Pendidikan Agama Islam*, vol. 1, no. 4, pp. 213–224, Dec. 2024, doi: https://doi.org/10.61132/hikmah.v1i4.329.

[5] M. Alenezi, S. Wardat, and M. Akour, "The Need of Integrating Digital Education in Higher Education: Challenges and Opportunities," *Sustainability (Switzerland)*, vol. 15, no. 6, p. 4782, Mar. 2023, doi: https://doi.org/10.3390/su15064782.

[6] A. Haleem, M. Javaid, M. A. Qadri, and R. Suman, "Understanding the role of digital technologies in education: A review," *Sustainable Operations and Computers*, vol. 3, pp. 275–285, Jan. 2022, doi: https://doi.org/10.1016/j.susoc.2022.05.004.

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

[7] Kementerian Pendidikan Dasar dan Menengah, "Data Pokok Peserta Didik," Data Pokok Pendidikan (DAPODIK). Accessed: Apr. 08, 2025. [Online]. Available: https://dapo.dikdasmen.go.id/sp

[8] R. A. Ariyaluran Habeeb *et al.*, "Clustering-based real-time anomaly detection—A breakthrough in big data technologies," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 8, Aug. 2022, doi: https://doi.org/10.1002/ett.3647.

[9] A. L. Maukar, F. Marisa, A. A. Widodo, N. Kamilaningtyas, D. Novian, and D. Nugraha, "ANALISIS DATA PENERIMAAN MAHASISWA BARU BERBASIS K-MEANS," *JIKO (Jurnal Informatika dan Komputer)*, vol. 6, no. 2, pp. 142–147, Sep. 2022, doi: http://dx.doi.org/10.26798/jiko.v6i2.558.

[10] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, Dec. 2022, doi: https://doi.org/10.1186/s40561-022-00192-z.

[11] A. Pasaribu, D. Prasetya Kristiadi, and C. Lea Taryono, "PENGEMBANGAN SISTEM PENILAIAN SISWA DENGAN MODEL RAPID APPLICATION DEVELOPMENT PADA SMP WAHANA HARAPAN," *Jurnal Sistem Informasi dan Teknologi (SINTEK)*, vol. 3, no. 1, pp. 9–13, 2023, doi: https://doi.org/10.56995/sintek.v3i1.50.

[12] Y. M. Gultom, F. Syahputra, and S. Syahrial, "Pengaruh Evaluasi Pembelajaran terhadap Kualitas Pembelajaran Guru di Sekolah Dasar," *Jurnal Pendidikan Guru Sekolah Dasar*, vol. 1, no. 3, pp. 1–8, May 2024, doi: https://doi.org/10.47134/pgsd.v1i3.543.

[13] M. Hilman, A. Rinaldi Dikananda, and A. Rifai, "K-Means Algorithm for Clustering High-Achieving Student at Madrasah Tsanawiyah Yami Waled," *Journal of Artificial Intelligence and Engineering Applications*, vol. 4, no. 3, pp. 1538–1548, Jun. 2025, doi: https://doi.org/10.59934/jaiea.v4i3.771.

[14] T. Andayani and F. Madani, "Peran Penilaian Pembelajaran Dalam Meningkatkan Prestasi Siswa di Pendidikan Dasar," *Jurnal Educatio FKIP UNMA*, vol. 9, no. 2, pp. 924–930, Jun. 2023, doi: https://doi.org/10.31949/educatio.v9i2.4402.

[15] V. Valentine *et al.*, "Penerapan Kurva Normal dalam Analisis Nilai Ujian Akhir Siswa Propinsi Kalimantan Tengah," *Informatech: Jurnal Ilmiah Informatika dan Komputer*, vol. 1, no. 2, pp. 157–162, 2024, doi: https://doi.org/10.69533/y4f5rd94.

[16] M. Qusyairi, Zul Hidayatullah, and Arnila Sandi, "Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow," *Infotek: Jurnal Informatika dan Teknologi*, vol. 7, no. 2, pp. 500–510, Jul. 2024, doi: https://doi.org/10.29408/jit.v7i2.26375.

[17] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *Journal of Information Systems and Informatics*, vol. 3, no. 3, Sep. 2021, doi: https://doi.org/10.51519/journalisi.v3i3.164.

[18] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 1, no. 1, pp. 20–28, Mar. 2023, doi: https://doi.org/10.58602/jaiti.v1i1.22.

[19] M. S. Hasibuan, A. H. Lubis, and M. N. Sari, "Perbandingan algoritma clustering dbscan dan k-means dalam pengelompokan siswa terbaik," *INFOTECH : Jurnal Informatika & Teknologi*, vol. 5, no. 2, pp. 301–309, Dec. 2024, doi: https://doi.org/10.37373/infotech.v5i2.1457.

[20] A. E. Haryati and S. Surono, "COMPARATIVE STUDY OF DISTANCE MEASURES ON FUZZY SUBTRACTIVE CLUSTERING," *MEDIA STATISTIKA*, vol. 14, no. 2, pp. 137–145, Jan. 2022, doi: https://doi.org/10.14710/medstat.14.2.137-145.

[21] A. Eka Haryati, S. Surono, T. Tanu Wijaya, G. Khang Wen, and A. Thobirin, "Fuzzy subtractive clustering (FSC) with exponential membership function for heart failure disease clustering," *International Journal Of Artificial Intelegence Research*, vol. 6, no. 1, Jun. 2022, Accessed: Jun. 24, 2025. [Online]. Available: https://ijair.id/index.php/ijair/article/view/306

[22] S. Kusumadewi, L. Rosita, and E. G. Wahyuni, "Performance of Fuzzy C-Means (FCM) and Fuzzy Subtractive Clustering (FSC) on Medical Data Imputation," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 15, no. 1, pp. 29–40, May 2024, doi: https://doi.org/10.21512/comtech.v15i1.11002.

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

[23] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: https://doi.org/10.1109/ACCESS.2020.2988796.

[24] A. E. Haryati and Sugiyarto, "Clustering with Principal Component Analysis and Fuzzy Subtractive Clustering Using Membership Function Exponential and Hamming Distance," *IOP Conf Ser Mater Sci Eng*, vol. 1077, no. 1, p. 012019, Feb. 2021, doi: 10.1088/1757-899x/1077/1/012019.

[25] L. Rosita, S. Kusumadewi, T. Ratnaningsih, N. Kertia, B. D. Purwanto, and E. G. Wahyuni, "Ferritin Level Prediction in Patients with Chronic Kidney Disease using Cluster Centers on Fuzzy Subtractive Clustering," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 403–418, Jul. 2024, doi: http://dx.doi.org/10.12785/ijcds/160132.

[26] M. Idhom, D. A. Prasetya, P. A. Riyantoko, T. M. Fahrudin, and A. P. Sari, "Pneumonia Classification Utilizing VGG-16 Architecture and Convolutional Neural Network Algorithm for Imbalanced Datasets," *TIERS Information Technology Journal*, vol. 4, no. 1, pp. 73–82, Jun. 2023, doi: https://doi.org/10.38043/tiers.v4i1.4380.

[27] B. Hakim, "Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning," *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 2, Aug. 2021, doi: http://dx.doi.org/10.30813/jbase.v4i2.3000.

[28] P. A. Riyantoko, K. M. Hindrayani, T. M. Fahrudin, and M. Idhom, "Exploratory Data Analysis and Machine Learning Algorithms to Classifying Stroke Disease," *Network Security and Information System (IJCONSIST)*, vol. 2, no. 2, pp. 77–82, 2021, doi: https://doi.org/10.33005/ijconsist.v2i02.49.

[29] I. N. Simbolon and P. D. Friskila, "ANALISIS DAN EVALUASI ALGORITMA DBSCAN (DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE) PADA TUBERKULOSIS," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: https://doi.org/10.23960/jitet.v12i3S1.5206.

[30] M. K. Dahouda and I. Joe, "A Deep-Learned Embedding Technique for Categorical Features Encoding," *IEEE Access*, vol. 9, pp. 114381–114391, Aug. 2021, doi: https://doi.org/10.1109/ACCESS.2021.3104357.

[31] R. D. Christyanti, D. Sulaiman, A. P. Utomo, and M. Ayyub, "Clustering Wilayah Kerawanan Stunting Menggunakan Metode Fuzzy Subtractive Clustering," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 17, no. 1, pp. 1–8, Oct. 2023, doi: https://doi.org/10.32815/jitika.v17i1.877.

[32] S. Kusumadewi and H. Purnomo, *APLIKASI LOGIKA FUZZY Untuk Pendukung Keputusan*, 2nd ed. Yogyakarta: Graha Ilmu, 2013.

[33] D. Arman Prasetya, A. P. Sari, M. Idhom, and A. Lisanthoni, "Optimizing Clustering Analysis to Identify High-Potential Markets for Indonesian Tuber Exports," *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 1, pp. 113–122, Feb. 2025, doi: https://doi.org/10.35882/skzqbd57.

[34] N. Azizah, D. Yuniarti, and R. Goejantoro, "Penerapan Metode Fuzzy Subtractive Clustering (Studi Kasus: PengelompokkanKecamatan di Provinsi Kalimantan Timur Berdasarkan Luas Daerah dan Jumlah Penduduk Tahun 2015)," *Jurnal EKSPONENSIAL*, vol. 9, no. 2, Nov. 2018, Accessed: Jun. 23, 2025. [Online]. Available: https://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/316

[35] S. Renaldi. S, D. A. Prasetya, and A. Muhaimin, "Analisis Klaster Partitioning Around Medoids dengan Gower Distance untuk Rekomendasi Indekos (Studi Kasus: Indekos di Sekitar Kampus UPNVJT)," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 3, pp. 2060–2069, Jul. 2024, doi: https://doi.org/10.33379/gtech.v8i3.4898.

[36] M. Yohansa, K. A. Notodiputro, and E. Erfiani, "Dynamic Time Warping Techniques for Time Series Clustering of Covid-19 Cases in DKI Jakarta," *ComTech: Computer, Mathematics and Engineering Applications*, vol. 13, no. 2, pp. 63–73, Nov. 2022, doi: https://doi.org/10.21512/comtech.v13i2.7413.

[37] M. F. Fahrudin, P. A. Riyantoko, K. M. Hindrayani, and H. P. Swari, "Cluster Analysis of Hospital Inpatient Service Efficiency Based on BOR, BTO, TOI, AvLOS Indicators using Agglomerative Hierarchical Clustering," *Jurnal Informatika dan Teknologi Informasi*, vol. 18, no. 2, pp. 194–210, 2021, doi: https://doi.org/10.31315/telematika.v18i2.4786.

Vol. 6, No. 3, Juni 2025, Page. 1569-1588
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4614

[38] I. P. Ginandjar, P. N. S, and R. Ilyas, "PENGARUH SELEKSI FITUR PADA TINGKAT AKURASI METODE RANDOM FOREST UNTUK IDENTIFIKASI AKUN BUZZER TWEET TOKOH POLITIK INDONESIA," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, pp. 3427–3432, Oct. 2023, doi: https://doi.org/10.36040/jati.v7i5.7477.

[39] S. Delimawati, H. Yozza, and Maiyastri, "PENGELOMPOKAN KABUPATEN/KOTA DISUMATERA BARAT BERDASARKAN FAKTORTERKAIT KEJADIAN DEMAM BERDARAH DENGAN METODE FUZZY SUBTRACTIVE CLUSTERING," *Jurnal Matematika UNAND*, vol. 10, no. 1, pp. 150–158, 2021, doi: https://doi.org/10.25077/jmu.10.1.150-158.2021.