Jurnal Teknologi Informasi dan Pendidikan
Volume 16, No.1, March 2023
https://doi.org/10.24036/jtip.v16i1.668

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

# Comparison Of Clustering Levels Of The Learning Burnout Of Students Using The Fuzzy C-Means And K-Means Methods

**Winarno[1*], Arif Satria Mustaqim[1], Denis Eka Cahyani[2]**
[1]Informatics Universitas Sebelas Maret, Indonesia
[2]Mathematics, Universitas Negeri Malang, Indonesia
**Corresponding Author: win@staff.uns.ac.id*

**Article Information**

***Article history:***

*No. 668*
*Rec. December 15, 2022*
*Rev. May 4, 2023*
*Acc. May 14, 2023*
*Pub. May 22, 2023*
*Page. 38 - 53*

***Keywords:***

- Clustering
- Fuzzy C-Means
- KMean
- Burnout learning

**ABSTRACT**

*This study aims to investigate the impact of learning burnout resulting from continuous and monotonous work, leading to physical and emotional fatigue. Learning burnout can have a detrimental impact on the productivity of students and hinder their potential when it is not adequately addressed. To address this issue, this study proposes a clustering method for group-level saturation of the learning of students. Fuzzy C-Means and K-Means clustering algorithms are used to produce the best results. Furthermore, it compares the performance of the Fuzzy C-Means and K-Means methods using a dataset of student boredom, and the testing is performed with 3, 4, and 5 clusters. The results show that Fuzzy C-Means yielded a score of 0.224 for the Davies Bouldin Index, while K-Means obtains a score of 0.384. Additionally, the global silhouette coefficient for Fuzzy C-Means is 0.278, and K-Means produces a score of 0.287. Based on these findings, it can be concluded that Fuzzy C-Means generate more precise clusters than K-Means.*



## 1. INTRODUCTION

The phenomenon of learning burnout poses a significant challenge in academic settings, specifically during the process of acquiring knowledge. The negative impact of boredom and disinterest in studying can have severe psychological consequences, resulting in disruptions to the performance of an individual and ability to achieve academic goals [1]. Moreover, the issue of saturation learning results in unproductive students and impedes their potential growth [2]. One possible approach to determine the level of this learning is to group students based on their saturation levels. The conventional method used by guidance counselors may not always result in accurate and efficient grouping, leading to subjective evaluations.

By utilizing information technology, different studies propose to group students based on their level of learning burnout using clustering methods, which is expected to generate more accurate groupings. Students' clustering offers several advantages, such as improved placement, accompaniment, and overall development [3]. Additionally, arranging students from heterogeneous classes can contribute to creating a better learning atmosphere [4].

This study utilizes two clustering methods, namely K-Means and Fuzzy C-Means. The K-Means method groups data based on the distance between the data and the cluster center, while the Fuzzy C-Means method is based on the degree of membership, which ranges from 0 to 1 [5]. Previous studies showed that both methods generated significant results with distinct patterns in different clusters. K-Means is particularly useful for classifying large amounts of data in mining applications [6].

Numerous analyses have been conducted using both clustering methods. Mingoti and Lima compared Fuzzy C-Means with several nonhierarchical clustering algorithms, such as K-Means and Self-Organization Map (SOM). The study found that Fuzzy C-Means produced better and more significant results compared to the nonhierarchical clustering algorithms. The method was also shown to have better overall performance, stability, and robustness in data clustering. Additionally, it was not affected by overlapping results and outliers, with an average recovery rate of 90% [7]. The study conducted by Velmurungan and Santhanam also compared the two clustering methods and found that K-Means have better performance due to its light computing burden and faster grouping time [8].

Based on the above descriptions, a comparison between the two methods is necessary to determine the most suitable for grouping data. In this study, a comparison between the clustering methods was conducted to assess the effectiveness of classifying saturation data of future students for determining their level of burnout.

## 2. RESEARCH METHOD

This study employs a methodology that is outlined in Figure-1.

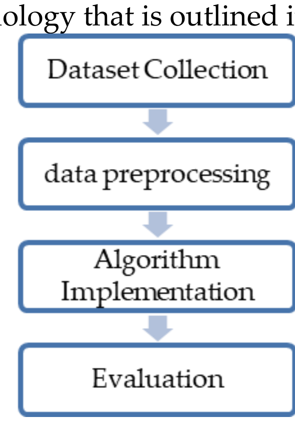


**Figure-1.** Study Flow Diagram

### 2.1 User Requirements Designing

This study utilized the burnout dataset of 498 students from SMA N 1 Boyolali gathered between October 1st and December 1st, 2020. Among these students, 228 were from 10th grade, 179 were from 11th grade, and 91 were from 12th grade. To collect the data, a scale that assessed burnout in learning was employed, which was developed in consultation with guidance teachers, counselors, and educational psychology experts. The scale consisted of four indicators, namely fatigue emotions, fatigue physical, exhausted cognitive, and loss of motivation [9]. Meanwhile, the four-choice answer format, which ranged from "not relevant" to "very relevant" and was graded on a scale of 1 to 4, was used.

### 2.2 System Design

Before preprocessing, the models were created based on the total score of each indicator from the questionnaire responses, which were selected by students and multiplied by every statement. The maximum value for each indicator was 100, and the first stage of data processing was data cleaning, which involved the removal of inconsistent and blank data. In this study, the cleaning process was performed manually using an Excel application. Out of 498 records, 3 were found to be inconsistent and were removed, leaving 495 available for analysis. The subsequent stage was the attribute selection process based on expert consultation. The questionnaire data included eight attributes, but after consulting with experts, only five were selected [10].

### 2.3 Development

*Fuzzy C-Means* algorithm facilitates the creation of clusters based on the degrees of data membership. In the approach, the determination of data membership can be achieved through score degrees with a range value between 0 and 1 [11]. Specifically, a data point has a score membership of 0 when it does not enter the member and 1 when it enters the member from a set fuzzy. To perform clustering with Fuzzy C-Means, the following algorithm can be carried out:

1. Construct a data matrix with dimensions n x m, where n and m represent the number of data points and attributes.
2. Determine the initial values for calculating scores, count of *cluster*, rank, iteration maximum, the smallest expected error( $\xi$), the initial objective function, and initial iteration.
3. Set random $\mu_{ik}$value as element matrix partition initial ( $U$) with a score of $\mu_{ik}$ with each element of U falling in the range between 0 and 1. The number of columns in the partition matrix (U) should be equal to the clusters and can be determined using Equation 1.

$$U = \begin{pmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & .. & \mu_{1k}(x_k) \\ \mu_{21}(x_1) & \mu_{22}(x_2) & .. & \mu_{2k}(x_k) \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{i1}(x_1) & \mu_{i2}(x_2) & .. & \mu_{ik}(x_k) \end{pmatrix} \quad (1)$$

4. Count center *clusters* iteration to k with Equation 2.

$$V_{kj} = \frac{\sum_{i=1}^{n}(((\mu_{ik})^{w} * X_{ij})}{\sum_{i=1}^{n}(\mu_{ik})^{w}} \quad (2)$$

5. Count score function objective on iteration to -t with Equation 3.

$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}(X_{ij} - V_{kj})^2\right](\mu_{ik})^{w}\right) \quad (3)$$

6. Count change matrix partition ( $U$) with Equation 4.

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}(X_{ij}-V_{kj})^2\right]^{\frac{-1}{w-1}}} \quad (4)$$

7. Check | Pt – (Pt-1) | < ξ or (t > iteration maximum) to determine when to stop after the condition is satisfied.
8. Repeat step d when the condition is not satisfied [9].

## 2.4 Implementation

Algorithm *K-Means* use distance to group data into some *clusters* with the same characteristics [1]. The process of algorithm *K-means* is shown below.

1. Determine many *clusters* to be formed
2. Determine center *clusters* beginning randomly. Change center *clusters* using Equation 5.

$$V = \frac{\sum_{i=1}^{n} x_i}{n} \quad (5)$$

3. Calculate the Euclidean distance from each data point to the centroids of each cluster using Equation 6.

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (6)$$

4. Group the data based on their closest distance to each centroid, considering their shared characteristics.
5. Perform an iteration by determining the new positions of the centroids.
6. The grouping is considered complete when the clusters do not move or change. However, the process is repeated when there is displacement [12].

## 2.5 Evaluation

Evaluation is a crucial aspect in ensuring the quality of an algorithm [13]. To evaluate the clustering performance of Fuzzy C-Means and K-Means algorithms, two common methods are the silhouette coefficient and the Davies-Bouldin Index (DBI).


P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

## 2.6 Silhouette Coefficient

*Silhouette coefficient* involves the use of two components, namely ai, and bi. The value of ai is the average distance of a data point to all other points within the same cluster [14], while bi is the minimum average distance to all data points in a different cluster, as shown in Equation 7.

$$a_j^i = \frac{1}{m_j - 1} \sum_{r=1\, r \neq 1}^{mj} d(x_i^j, x_r^j) \quad (7)$$

Value *bi* can be calculated with Equation 8.

$$b_j^i = \left\{ \frac{1}{m_n} \sum_{r=1\, r \neq 1}^{mn} d(x_i^j, x_r^n) \right\} \quad (8)$$

Value SI can be calculated with Equation 9.

$$SI_i^j = \frac{b_i^j - a_i^j}{max\left\{a_i^j, b_i^j\right\}} \quad (9)$$

The range of the silhouette coefficient is between -1 and +1, where a value closer to 1 indicates that the data point is well-matched with its cluster. On the other hand, a negative value shows that the data point is likely to belong to a different cluster.

$$SI_j = \frac{1}{m_j} \sum_{i=1}^{mj} SI_i^j \quad (10)$$

Criteria SI measurement can be seen in Table 1.

**Table 1.** Silhouette Coefficient Criteria [15]

| Score | Interpretation |
|---|---|
| 0.71-1.00 | Structure strong |
| 0.51-0.70 | Structure good |
| 0.26-0.50 | Structure weak |
| <=0.25 | Structure bad |

## 2.7 Davis Bouldin Index

The DBI involves the calculation of two components, namely the Sum of Squares Between Clusters (SSB) and the Sum of Squares Within Clusters (SSW) [16]. SSB is used to measure the separation or distance between each cluster and can be formulated using Equation 12.

$$\mathrm{SSB_{ij}} = \mathrm{d(c_i, c_j)} \quad (12)$$

SSW is used to measure the cohesion distance between a member of the cluster, as shown in Equation 13

$$\mathrm{SSW_i} = \frac{1}{\mathrm{m_i}} \sum_{\mathrm{j=i}}^{\mathrm{m_i}} \mathrm{d(x_j, c_i)} \quad (13)$$

The ratio value is used for comparing each cluster, as shown in Equation 14

$$\mathrm{R_{ij}} = \frac{\mathrm{SSW_i + SSW_j}}{\mathrm{SSB_{i,j}}} \quad (14)$$

The DBI value can be calculated using Equation 15.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j} (R_{i,j}) \tag{15}$$

The quality of clustering is considered better when the DBI value is closer to 0, indicating that the clustering is more valid [15]. The index can also be used to determine the optimal number of clusters for a given dataset [17].

### 2.8 Burnout

Burnout learning refers to a state in which an individual feels emotionally or physically exhausted due to their work responsibilities [2]. Various factors contribute to this problem, including high demands from educational institutions, inadequate opportunities for creativity, insufficient incentives, inadequate interaction between students and teachers, excessive parental expectations, and conflicts between the school environment and student family dynamics. These factors can manifest as indicators of burnout learning such as emotional, physical, and cognitive exhaustion, as well as loss of motivation [11]. Meanwhile, academic burnout may also significantly affect the level of self-efficacy [18].

## 3. RESULTS AND DISCUSSION

### 3.1 Data Collection

The dataset used pertains to the learning burnout experienced by students of SMA N 1 Boyolali. The data were collected from 1st October to 1st December 2020 and comprised a total of 498 students, with 228, 179, and 91 from 10th, 11th, and 12th grades, respectively. To collect the data, a scale was used to measure burnout learning, which was developed in consultation with guidance teachers, counseling experts, and educational psychologists. The scale comprised four indicators, namely emotional fatigue, physical fatigue, cognitive exhaustion, and loss of motivation [9]. In this study, a four-choice answer format was employed, which included options such as "not relevant", "somewhat relevant", "relevant", and "very relevant", with a scale ranging from 1 to 4.

### 3.2 Fuzzy C-Means

The summary of data for this dataset is presented in Table 2.

**Table 2.** Data Matrix

| X1 | X2 | X3 | X4 |
|---|---|---|---|
| **60** | 56 | 64 | 72 |
| 58 | 85 | 57 | 52 |



P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

| | | | |
|---|---|---|---|
| 49 | 37 | 52 | 49 |
| | ... | | |
| **58** | 78 | 78 | 81 |

Description:
$X_1$ = emotional exhaustion
$X_2$ = physical exhaustion
$X_3$ = cognitive exhaustion
$X_4$ = loss of motivation

Determine values beginning calculation as follows:
Amount *clusters* : 3
Rank ( $w$) : 2
Maximum iteration ( *MaxIter* ) : 1000
*error* smallest expected ( $\xi$) _ : $10^{-6}$
Function objective : 0
Iteration beginning : 1

The next step determine random $\mu_{ik}$value in form element matrix partition $U$, as in Table 3.

**Table 3** Matrix U Membership (Random)

| $\mu_{ik1}$ | $\mu_{ik2}$ | $\mu_{ik3}$ |
|---|---|---|
| **0.4** | 0.6 | 0.0 |
| **0.5** | 0.4 | 0.1 |
| **0.3** | 0.5 | 0.2 |
| | ... | |
| **0.0** | 0.9 | 0.1 |

Determine three center clusters $Vkj$ with Equation 2, the results of calculating the first iteration as shown in Tables 4, 5, and 6.

**Table 4** Calculating Cluster Center Results for First Iterations on Cluster 1

| No | $(\mu_{ik1})^2$ | $(\mu_{ik1})^2 .X_{i1}$ | $(\mu_{ik1})^2 .X_{i2}$ | $(\mu_{ik1})^2 .X_{i3}$ | $(\mu_{ik1})^2 .X_{i4}$ |
|---|---|---|---|---|---|
| **1** | 0.16 | 9.6 | 8.96 | 10.24 | 11.52 |
| **2** | 0.25 | 14.5 | 21.25 | 14.25 | 13 |
| **3** | 0.09 | 4.41 | 3.33 | 4.68 | 4.41 |
| | | | ... | | |
| **495** | 0 | 0 | 0 | 0 | 0 |
| **count** | 77.61 | 4623.0 | 4121.9 | 4307.0 | 4204.05 |

**Table 5.** Calculating Cluster Center Results for First Iterations on Cluster 2

| No | $(\mu_{ik1})^2$ | $(\mu_{ik1})^2 .X_{i1}$ | $(\mu_{ik1})^2 .X_{i2}$ | $(\mu_{ik1})^2 .X_{i3}$ | $(\mu_{ik1})^2 .X_{i4}$ |
|---|---|---|---|---|---|
| **1** | 00.36 | 21.06 | 20.16 | 23.04 | 25.92 |
| **2** | 00.16 | 09.28 | 13.06 | 09.12 | 08.32 |
| **3** | 00.25 | 12.25 | 09.25 | 13 | 12.25 |
| | | | ... | | |
| **495** | 0,05625 | 46.98 | 63.18.00 | 63.18.00 | 65.61 |
| **count** | 79.69 | 4801.09.00 | 4374.03.00 | 4480.07.00 | 4249.00.00 |

**Table 6.** Calculating Cluster Center Results for First Iterations on Cluster 3

| No | $(\mu_{ik1})^2$ | $(\mu_{ik1})^2 .X_{i1}$ | $(\mu_{ik1})^2 .X_{i2}$ | $(\mu_{ik1})^2 .X_{i3}$ | $(\mu_{ik1})^2 .X_{i4}$ |
|---|---|---|---|---|---|
| **1** | 0 | 0 | 0 | 0 | 0 |
| **2** | 00.01 | 00.58 | 0.059 | 00.57 | 00.52 |
| **3** | 00.04 | 0.108 | 01.48 | 02.08 | 0.108 |
| | | | ... | | |
| **495** | 0.01 | 0.58 | 0.054 | 0.054 | 0.056 |
| **count** | 90.03 | 5372.44 | 4712.09 | 4997.31 | 4842.0 |

*The Center of the Cluster* on the first iteration determined with Equation 2 is shown below

$$V = \begin{bmatrix} 59.567 & 0.891 & 1.0449 & 0.976 \\ 60.257 & 0.910 & 1.0243 & 0.948 \\ 59.673 & 0.877 & 1.0603 & 0.969 \end{bmatrix}$$

The outcomes of computing the objective function using Equation 3 are illustrated in Table 7.

**Table 7**. Result of Calculation Objective Function at First Iteration

| No | $L_1$ | $L_2$ | $L_3$ | $LT$ |
|---|---|---|---|---|
| **1** | 256.59 | 146.229 | 146.229 | 413.05 |
| **2** | 36.914 | 120.845 | 120.845 | 173.01 |
| **3** | 4.445 | 25.537 | 120.845 | 41.16 |
| | | ... | | |
| **495** | 0 | 1.441.229 | 1.441.229 | 1460 |


P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

| |
|---|
| 183119.16 |

According to Equation 3, matrix U can be determined into a new matrix, as shown in Table 8.

**Table 8**. Calculation Results Degrees Membership New ( Matrix Partition New )

| | | | |
|---|---|---|---|
| **1** | 0.00291 | 0.00589 | 0.00264 |
| **2** | 0.00215 | 0.00111 | 0.00109 |
| **3** | 0.00111 | 0.00159 | 0.00216 |
| | | ... | |
| **495** | 0.00054 | 0.00056 | 0.00052 |

The updated partition matrix for the initial iteration is presented in Table 9.

**Table 9.** New Partition Matrix ($U_1$)

| $\mu_{ik1}$ | $\mu_{ik2}$ | $\mu_{ik3}$ |
|---|---|---|
| **0.00291** | 0.00589 | 0.00264 |
| **0.00215** | 0.00111 | 0.00109 |
| **0.00111** | 0.00159 | 0.00216 |
| | ... | |
| **0.00054** | 0.00056 | 0.00052 |

After the first iteration, the results obtained were P1 = 183119.163 and P0 = 0, since ∣P1 - P0∣ = ∣183119.163 - 0∣ = 183119.163 which is less than the pre-defined tolerance level of ξ (10^(-6)), and the current iteration count is 1, less than the maximum limit of 1000. Therefore, the algorithm will proceed to the next iteration, and the final result yielded the following matrix representing the cluster centers.

$$V = \begin{bmatrix} 60.448 & 52.253 & 56.113 & 54.435 \\ 72.039 & 67.682 & 69.593 & 65.791 \\ 47.693 & 41.309 & 42.399 & 41.673 \end{bmatrix}$$

The calculated value of the objective function $|P_{33} - P_{32}| = |\ 95866.085 - 95872.179\ |$ = 6.94E-07 >ξ$(10^{-6})$. Therefore, it can be concluded that the objective value is smaller than the minimum error, as depicted in Figure 2.

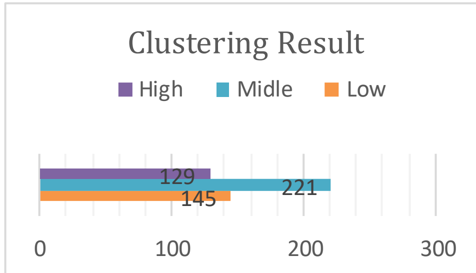


**Figure 2**. Learning Burnout Clustering Result With Fuzzy C-Means

Based on Figure 2, the cluster located in the middle corresponds to the learning burnout experienced by 221 students in SMA N 1 Boyolali.

### 3.3 K-Means

The centroid of the data is selected randomly at the 165th, 330th, and 400th data. The result of the centroid is shown in Table 10.

**Table 10**. Result of Centroid

| Cluster | Data1 | Data2 | Data3 | Data4 |
|---|---|---|---|---|
| $C_1$ | 63 | 54 | 59 | 60 |
| $C_2$ | 71 | 67 | 71 | 62 |
| $C_3$ | 37 | 31 | 49 | 47 |

The subsequent step involved calculating the distances between the data and the centroid, as presented in Table 11.

**Table 11.** Distance of Centroid to the Data

| No | d1 | d2 | d3 |
|---|---|---|---|
| 1 | 32.465 | 28.089 | 58.702 |
| 2 | 25.592 | 43.749 | 13.892 |
| 3 | 4.690 | 17.291 | 42 |
| | | … | |
| 495 | 37.456 | 26.457 | 68.168 |

Each record needs to be assigned to the cluster with the closest centroid. Furthermore, when a data point is closer to a certain centroid, it will be assigned to the cluster, as shown in Table 12.


P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

**Table 12**. Clustering Result First Iteration

| No | Low | Middle | High |
|---|---|---|---|
| 1 | | 1 | |
| 2 | | | 1 |
| 3 | 1 | | |
| | | ... | |
| 495 | | 1 | |

After all the data points have been assigned to their respective clusters, the next step involves computing new centers using Equation 5. The iteration process continues until the data points become stable. In this case, the algorithm stopped at 45 iterations, and the final centroid matrix is presented below.

$$V = \begin{bmatrix} 46.274 & 41.5 & 40.559 & 39.744 \\ 72.635 & 69.627 & 70.481 & 65.788 \\ 60.318 & 50.202 & 56.100 & 54.679 \end{bmatrix}$$

The clustering results are illustrated in Figure 2.

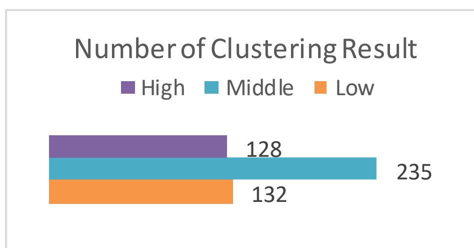


**Figure 2.** Result of Clustering with K Means

Based on Figure 2, there are 235, 132, and 123 students in the middle, low, and high clusters. By comparing Figures 1 and 2, it is evident that there is a discrepancy in the number of students experiencing burnout when using the Fuzzy C-Means and K-Means algorithm.

### 3.4 Evaluation

In this study, the Silhouette coefficient and DBI methods were employed to assess the quality of the clustering results [19]. The result is calculated by measuring the cohesion and separation of the data points in each cluster. On the other hand, the DBI evaluates the distance between each data point and the centroid of its respective cluster. The Silhouette coefficient is typically used to evaluate the homogeneity of the clustering results [20].

### 3.4.1 Evaluation with Silhouette Coefficient at Fuzzy C-Mean Case

Table 13 displays the calculation of the Silhouette coefficient for each cluster obtained from the Fuzzy C-Means algorithm.

**Table 13.** Result from Silhouette Coefficient Calculation with Fuzzy C-Means

| SI 1 | SI 2 | SI 3 |
|---|---|---|
| 0.184 | 0.300 | 0.041 |
| 0.137 | 0.457 | 0.467 |
| 0.472 | 0.094 | 0.342 |
| | … | |
| 0.092 | 0.442 | 0.342 |

After the silhouette coefficient for the clusters has been determined, it is possible to calculate the global value by using Equation 11, as shown in Table 14.

**Table 14.** Silhouette Coefficient Global

| Cluster | Data Count | SI |
|---|---|---|
| 1 | 145 | 0.284 |
| 2 | 221 | 0.299 |
| 3 | 129 | 0.252 |
| SI Global | | 0.278 |

According to the computed global silhouette coefficient value of 0.278, it can be inferred that the structure of the Fuzzy C-Means clustering algorithm is weak.

### 3.4.2 Evaluation with Davies Bouldin Index at Fuzzy C-Mean Case

Table 15 displays the results of the computation of SSW for each cluster, according to Equation 13.

**Tabel 15**. SSW Cluster

| Cluster | SSW |
|---|---|
| C1 | 16.33 |
| C2 | 14.839 |
| C3 | 18.059 |

The subsequent step involves the calculation of SSB value for each cluster, using Equation 12, as shown in Table 16.

**Tabel 16**. SSB Calculation

| | **Centroid** | | |
|---|---|---|---|
| **SSB** | C1 | C2 | C3 |



P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

| | | | |
|---|---|---|---|
| **1** | 0 | 25.160 | 51.164 |
| **2** | 25.160 | 0 | 26.214 |
| **3** | 51.164 | 26.214 | 0 |

Equation 14 can calculate the ratio value $R_{12}, R_{13}$ *and* $R_{23}$ as $1.239, 0.672$ and $1.254$, respectively. DBI value is obtained by finding the greatest ratio value and dividing the result by the count of the cluster, as shown below.

$$DBI = \frac{1}{3}(1.254) = 0.224$$

From the calculation, it can be seen that the DBI for the Fuzzy C-Means cluster results is 0.224, hence, classified in good condition.

### 3.4.3 Evaluation with Silhouette Coefficient at K- Means

Table 17 displays the calculation of the silhouette coefficient for the 495 students cluster.

**Table 17.** Calculation Result using Silhouette Coefficient with K-Means Cluster 1

| SI 1 | SI 2 | SI 3 |
|---|---|---|
| **0.099** | 0.098 | 0.292 |
| **0.077** | 0.474 | 0.460 |
| **0.466** | 0.342 | 0.171 |
| | … | |
| **0.330** | 0.347 | 0.464 |

The subsequent step involves obtaining the Global Silhouette Coefficient value, which is presented in Table 18.

**Table 18.** Calculation Result Global Silhouette Coefficient

| Cluster | Data Count | SI |
|---|---|---|
| 1 | 132 | 0.299 |
| 2 | 128 | 0.264 |
| 3 | 235 | 0.296 |
| SI Global | | 0.287 |

The computation has yielded a Global Silhouette Coefficient value of 0.287, indicating that the structure of the K-Means clustering algorithm is weak. This is consistent with the evaluation of the Fuzzy C-Means algorithm using the silhouette coefficient.

### 3.4.4 Evaluation with Davies Bouldin Index at Fuzzy C-Means Case

Using Equation 13, it is possible to obtain the value of SSW for each cluster, as shown in Table 19.

**Table 19.** SSW Cluster

| Cluster | SSW |
|---|---|
| C1 | 16.304 |
| C2 | 18.247 |
| C3 | 15.00 |

The subsequent step involves obtaining the value of SSB for each cluster, using Equation 12, as shown in Table 20.

**Table 20.** Calculation SBB value

| | Centroid | | |
|---|---|---|---|
| **SSB** | C1 | C2 | C3 |
| **1** | 0 | 55.314 | 27.157 |
| **2** | 55.314 | 0 | 29.314 |
| **3** | 27.157 | 29.314 | 0 |

The subsequent step involves obtaining the ratio value by using Equation 14, where $R_{12} = 0.624$, $R_{13} = 1.152$ and $R_{23} = 1.134$. It can be observed that $R_{13}$ has the largest value among the three ratios. Using Equation 15, the value of the DBI can be obtained and is presented below.

$$DBI = \frac{1}{3}(1.152) = 0.384$$

Based on the calculation, the DBI value for the K-Means clustering results is 0.384, where the clustering results are in good condition. The DBI result shows that the clustering of Fuzzy C-Means is better than K-Means, and the DBI value is also lower. Therefore, the clustering result for Fuzzy C-Means is closer to the ideal condition where the DBI value is equal to zero.

## 4. CONCLUSION

This study has successfully applied the Fuzzy C-Means and K-Means methods to cluster the levels of learning burnout among SMA N 1 Boyolali students. The system test results using the Silhouette coefficient method yielded a value of 0.278 and 0.287 for Fuzzy


P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

C-Means and K-Means, respectively. Based on the evaluation of the Silhouette coefficient, K-Means outperformed Fuzzy C-Means since its value was closer to 1. Furthermore, K-Means and Fuzzy C-Means obtained a score of 0.384 and 0.224, respectively, when evaluated with the DBI. It can be concluded that Fuzzy C-Means was superior to K-means since the DBI results were lower, indicating a good value close to zero. In future studies, it is recommended to include additional factors such as the GPA score of students and their count of credits achieved.

## REFERENCES

[1] P. Stiles, "Joan Turville-Petre: A Bibliographical Appreciation." Appeared in Old English Newlsetter, 2007.
[2] B. Slivar, "The syndrome of burnout, self-image, and anxiety with grammar school students," Psihološka obzorja / Horizons of Psychology, vol. 2, no. 10, pp. 21–32, 2001.
[3] Y. Muharmi, "Pengelompokan Siswa Berdasarkan Faktor-faktor yang Mempengaruhi Keberhasilan Siswa Dalam Belajar Menggunakan Metode Clustering K-Means," Jurnal Teknologi Informasi & Pendidikan, vol. 9, no. 1, pp. 94–101, 2016.
[4] D. Kurniadi, R. Safitri, D. Irfan, and K. Budayawan, "Determining Study Groups Based on Student ProfileCriteria Using K-Means Method," Jurnal Teknologi Informasi dan Pendidikan, vol. 14, no. 1, pp. 271–276.
[5] A. Kapoor and A. Singhal, "A comparative study of K-Means, K-Means++ and Fuzzy C-Means clustering algorithms," in 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, India, Feb. 2017, pp. 1–6. doi: 10.1109/CIACT.2017.7977272.
[6] S. T. Siska, "Analisa Dan Penerapan Data Mining Untuk Menentukan Kubikasi Air Terjual Berdasarkan Pengelompokan Pelanggan Menggunakan Algoritma K-Means Clustering," Jurnal Teknologi Informasi & Pendidikan, vol. 9, no. 1, pp. 86–93, 2016.
[7] S. A. Mingoti and J. O. Lima, "Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms," European Journal of Operational Research, vol. 174, no. 3, pp. 1742–1759, Nov. 2006, doi: 10.1016/j.ejor.2005.03.039.
[8] T. Velmurugan and T. Santhanam, "A Comparative Analysis between K-Medoids and Fuzzy C-Means Clustering Algorithms for Statistically Distributed Data Points," Journal of Theoretical and Applied Information Technology, vol. 27, no. 1, pp. 19–30, 2011.
[9] I. Vitasari, "Kejenuhan (Burnout) Belajar Ditinjau dari Tingkat Kesepian dan Kontrol Diri pada Siswa Kelas XI SMA Negeri 9 Yogyakarta," 2012. doi: 10.1017/CBO9781107415324.004.
[10] P.-N. Tan, M. Steinbach, and V. Kumar, "Basic Concepts, Decision Trees, and Model Evaluation," in Introduction to Data Mining, 2006.
[11] G. Chen and T. T. Pham, Introduction to Fuzzy Sets, Fuzzy Logic, and Fuzzy Control Systems. 2000. doi: 10.1201/9781420039818.
[12] M. Pokharel, J. Bhatta, and N. Paudel, "Comparative Analysis of K-Means and Enhanced K-Means Algorithms for Clustering," NUTA Jnl, vol. 8, no. 1–2, pp. 79–87, Dec. 2021, doi:

10.3126/nutaj.v8i1-2.44044.
[13] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," Electronics, vol. 9, no. 8, p. 1295, Aug. 2020, doi: 10.3390/electronics9081295.
[14] H. Guo, J. Ma, and Z. Li, "Active Semi-supervised K-Means Clustering Based on Silhouette Coefficient," in Advances in Intelligent, Interactive Systems and Applications, vol. 885, F. Xhafa, S. Patnaik, and M. Tavana, Eds. Cham: Springer International Publishing, 2019, pp. 202–209. doi: 10.1007/978-3-030-02804-6_27.
[15] S. Petrovic, "A Comparison Between the Silhouette Index and the Davies-Bouldin Index in Labelling IDS Clusters," 11th Nordic Workshop on Secure IT-systems, 2006.
[16] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, doi: 10.1109/TPAMI.1979.4766909.
[17] A. Badruttamam, Sudarno Sudarno, and Di Asih I Maruddani, "Penerapan Analisis Klaster K-Modes Dengan Validasi Davies Bouldin Index Dalam Menentukan Karakteristik Kanal Youtube Di Indonesia (Studi Kasus: 250 Kanal Youtube Indonesia Teratas Menurut Socialblade)," Jurnal Gaussian, vol. 9, no. 3, pp. 263–272, 2020.
[18] D. Permatasari, L. Latifah, and P. R. Pambudi, "Studi Academic Burnout dan Self-Efficacy Mahasiswa," JPP, vol. 4, no. 2, Dec. 2021, doi: 10.24176/jpp.v4i2.7418.
[19] M. Shutaywi and N. N. Kachouie, "Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering," Entropy, vol. 23, no. 6, p. 759, Jun. 2021, doi: 10.3390/e23060759.
[20] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering," tc, vol. 20, no. 2, pp. 186–197, May 2021, doi: 10.33633/tc.v20i2.4556.


P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id