



Contents lists available at DJM

DESIMAL: JURNAL MATEMATIKA

p-ISSN: 2613-9073 (print), e-ISSN: 2613-9081 (online), DOI 10.24042/djm
<http://ejournal.radenintan.ac.id/index.php/desimal/index>



K-means and fuzzy c-means algorithm comparison on regency/city grouping in Central Java Province

Ummu Wachidatul Latifah, Sugiyarto*, Suparman

Ahmad Dahlan University Yogyakarta, Indonesia

ARTICLE INFO

Article History

Received : 23-05-2022
Revised : 18-08-2022
Accepted : 21-08-2022
Published : 30-08-2022

Keywords:

K-Means; Fuzzy C-Means;
Clustering.

*Correspondence: E-mail:
Sugiyarto@math.uad.ac.id

Doi:
[10.24042/djm.v5i2.12204](https://doi.org/10.24042/djm.v5i2.12204)

ABSTRACT

The Human Development Index (HDI) is very important in measuring the country's success as an effort to build the quality of life of people in a region, including Indonesia. The government needs to make groupings based on the needs of a city/district. To facilitate data grouping based on the similarity of existing characteristics, it is necessary to have a data grouping method, namely the clustering technique. There are several algorithms that are often used in clustering techniques, namely K-Means and Fuzzy C-Means. Each algorithm has advantages and disadvantages. Therefore, in this research, the comparison of the best clustering algorithms will be discussed. The purpose of this research is to compare the K-Means and Fuzzy C-Means algorithms in the grouping of Regencies/Cities in Central Java Province in 2021 based on the Human Development Index (IPM) indicator. The method used is descriptive qualitative. The data used is obtained from the Central Java Statistics Agency, which is in the form of the Central Java Province HDI indicator in 2021 which consists of 35 Regency/City members and has 4 variables, namely Life Expectancy (AHH) (X1), Expectation of School Years (HLS) (X2), Average Length of School (RLS) (X3) and Expenditure Per Capita (X4). The results showed that after a comparative analysis with the standard deviation ratio method, the FCM clustering method was better than the K-Means method. The value of the FCM standard deviation ratio is 0.460093 and the K-Means standard deviation ratio is 0.473601.

<http://ejournal.radenintan.ac.id/index.php/desimal/index>

INTRODUCTION

The Human Development Index (HDI) is critical in determining a country's progress as an endeavor to improve the quality of life of its population/community in a region/country, including Indonesia (Kirana et al., 2019). The emergence of qualified, competent, and competitive

human resources is the human development index that is the focus and hope of progress in the aspect of education. Each city or district has a unique human development index. As a result, the government must create groupings based on the demands of a city/district. A data grouping method,

specifically the clustering methodology, is required to assist data grouping based on the similarity of existing characteristics.

Clustering is the process of grouping data based on similarities (Ghazal et al., 2021). The clustering algorithm's purpose is to divide data into several groups based on similar characteristics. Thus, the data will be divided into numerous clusters using the appropriate clustering technique. Each data in one cluster has high and low similarity to data in other clusters.

Hierarchical algorithms and partition algorithms are two algorithms that are frequently employed in clustering approaches. The dataset is divided into smaller subsets hierarchically in the hierarchical algorithm, whereas the dataset is partitioned into the desired number of sets in one step in the partitioning algorithm (Windarto et al., 2019). K-Means and Fuzzy C-Means are two popular partition methods.

The K-Means algorithm is a partition-based technique that attempts to divide data into two or more clusters by utilizing the average value as the cluster center (Oktarina et al., 2020). The basic idea behind the K-Means algorithm starts from determining the number of groups (k), then randomly selecting the cluster center (centroid). Second, compute the distance between each data point and the cluster's center. Third, divide the data into clusters based on their proximity. Fourth, calculate the new group center and then continue the procedures until the cluster center does not change or the maximum iteration is reached. It is possible to begin the clustering process by identifying the grouped data using the Euclidean Distance formula (Utomo, 2021). The K-Means algorithm has the advantage of being simple and efficient, but it also has the disadvantage of requiring parameters and being sensitive to foreign data (Kolay et al., 2017).

Fuzzy C-Means (FCM) is a clustering method that is part of the Hard K-Means algorithm. FCM employs a fuzzy grouping model, which makes each data point a member of all classes or clusters produced with varying degrees or levels of membership ranging from 0 to 1. The degree of membership determines the level of data existing in a class or cluster (Rahakbauw et al., 2017). The basic principle behind FCM is to first establish the cluster's center, which will mark the average location for each cluster. The cluster center is still inaccurate in the initial conditions. Each data set has varying degrees of participation in each cluster. It can be shown that by repeatedly fixing the cluster center and the membership value of each data, the cluster center will go to the correct location. This iteration is based on minimizing the objective function that describes the distance from a given data point to the cluster's center, which is weighted by the degree of membership of that data point (Irawan et al., 2021). However, fuzzy C-Means have several weaknesses, including the need for a large number of groups and a predetermined group membership matrix (Haqiqi & Kurniawan, 2015; Le & Altman, 2011). The advantage of Fuzzy C-Means over other algorithms is that cluster placement is more exact (Agustini, 2017).

Several previous research, for example Zhou & Yang (2020) did a comparative analysis between the K-Means clustering method and FCM clustering, with results indicating that the K-Means clustering approach is superior to FCM clustering. Another research, conducted by Ramadhan et al. (2019), focused on categorizing the occurrence of flood disasters in Indonesia. The FCM method is used in the grouping process. Hassan et al. (2020) published a paper entitled "*Evaluate the performance of K-Means and the Fuzzy C-Means algorithms to formation balanced clusters in wireless*",

which showed that Fuzzy C-Means outperformed K-Means.

Based on the information presented, it is clear that each clustering algorithm has distinct advantages and disadvantages. As a result, we will explore the grouping of districts/cities based on the Human Development Index using the K-Means and FCM algorithms in this paper.

METHOD

The quantitative research method was applied in this research. A quantitative method is a research approach in which the data to be evaluated is in the form of numbers (numeric). The Central Java Statistics Agency provided the Human Development Index data. The subjects of this research were residents of the Central Java province. The variables employed are derived from the Human Development Index (BPS Jateng, 2021), namely, Life Expectancy (AHH) (X_1), Expected Years of Schooling (HLS) (X_2), Average Years of Schooling (RLS) (X_3), and Expenditures Per Capita (X_4). The method used in this research was to evaluate the number of clusters formed using the standard deviation ratio method. Data processing was done using the Python programming language.

K-Means Algorithm

The steps in the K-Means clustering algorithm are as follows (Syakur et al., 2018):

1. Determine the number of K clusters and the maximum number of iterations.
2. Determines the value of the centroid (center point). The initial centroid value is chosen randomly, for the next iteration using the following equation.

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j \quad (1)$$

where:

C_i = average of i^{th} cluster

M = the amount of data that is a

member of the i^{th} cluster

i, j = cluster index

x_j = the value of the j^{th} data in the cluster

3. Calculate the distance between the centroid point and the point of each object. To calculate the distance, you can use Euclidean Distance, namely

$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (2)$$

where:

d = distance

(x_i, y_i) = object coordinates

(x_j, y_j) = centroid coordinates

4. Reallocation of data to each group is carried out based on the comparison of the distance between the data and the centroid of each group. Determination of cluster members is done by considering the minimum distance of the object. The value obtained in the data membership at the distance matrix is 0 or 1 where the value 1 is for data allocated to the cluster, while the value 0 is for data allocated to other clusters (Wakhidah, 2010).

$$a_{ij} \begin{cases} 1 & d = \min \{D(x_i, c_i)\} \\ 0 & \text{lainnya} \end{cases} \quad (3)$$

5. Recalculate the position of the centroid. a_{ij} is the membership value of x_i to the center point of $cluster c_1$, d is the shortest distance from point x_i to $cluster K$ after comparison and c_1 is the center of the 1st cluster. The objective function used for the K-Means method is determined based on the distance and value of data membership in the group. The objective function according to J. Mac Queen (1967) can be determined using the equation n is the number of data, k is the number of clusters, a_{i1} is the membership value of the data x_i to $cluster c_1$ followed by a , a has two possibilities, namely 1 or 0. If a is 1 or 0 then the data is a

member of the group and if a is 0 then the data is allocated to another cluster.

$$J = \sum_{i=1}^n \sum_{l=1}^k a_i c D(x_i c_i)^2 \quad (4)$$

6. If there is a change in the centroid value or the number of iterations is less than the maximum number of iterations, then do step 3, otherwise, the iteration

stops, and the results are clustered.

A flowchart is required to make it easier to identify the plot to locate the results of the cluster implementation when determining the cluster based on the available data using the K-Means clustering technique. The K-Means algorithm flowchart is shown in Figure 1 (Purba et al., 2018).

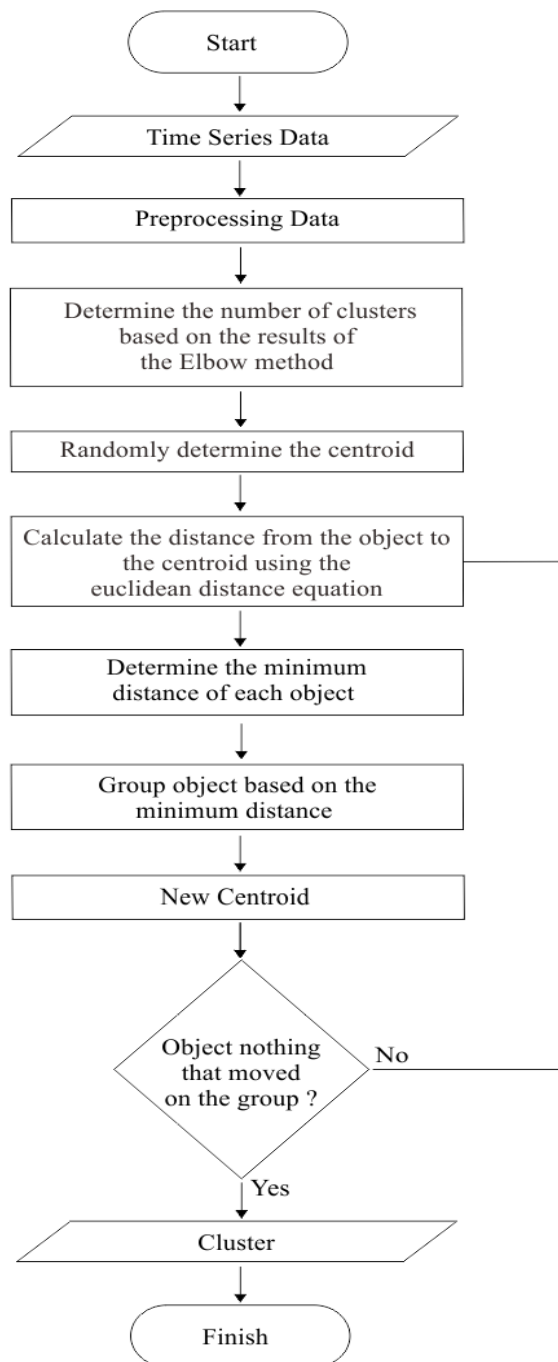


Figure 1. K-Means Flowchart

Fuzzy C-Means Algorithm

The basic concept behind FCM is to first establish the cluster's center, which will mark the average location for each cluster. However, the center of this cluster is still inaccurate in the initial conditions. For each cluster, each data point has a membership degree. The following is the FCM algorithm (Sutoyo & Sumpala, 2016):

1. The input data to be in *clusterX* is a matrix of size $n \times m$ (n is the number of data samples and m attributes for each data. X_{ij} is the i^{th} ($i = 1, 2, 3, \dots, n$) sample data, j^{th} attribute ($j = 1, 2, 3, \dots, m$).
2. Determine the following:
 - a. Number of clusters = c
 - b. Rank = w
 - c. Maximum Iteration = *MaxIter*
 - d. Smallest expected error = ξ
 - e. The objective function, $P_0 = 0$
 - f. Initial iteration, $t = 1$
3. Generate random numbers of $\mu_{ik}, i = 1, 2, 3, \dots, n; k = 1, 2, 3, \dots, c$; seb as elements of the initial partition matrix of U .

$$U_0 = \begin{pmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{c1} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{1i} & \dots & \dots & \mu_{ci} \end{pmatrix} \quad (5)$$

The Fuzzy Clustering partition matrix must meet the following conditions:

$$\mu_{ik} = [0, 1]; 1 \leq i \leq c; 1 \leq k \leq n \quad (6)$$

$$\sum_{i=1}^n \mu_{ik} = 1; 1 \leq i \leq c \quad (7)$$

Count the number of each column:

$$Q_j = \sum_{k=1}^c \mu_{ik} \quad (8)$$

with $j = 1, 2, 3, \dots, m$

Then calculate:

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j} \quad (9)$$

4. Calculate the center of the k^{th} : V_{kj} with $k = 1, 2, 3, \dots, c$ and $j = 1, 2, 3, \dots, m$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n \mu_{ik}} \quad (10)$$

5. Calculate the objective function at t^{th}, P_t :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (11)$$

6. Calculate the partition matrix change:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}} \quad (12)$$

with: $i = 1, 2, 3, \dots, n$; and $k = 1, 2, \dots, c$

7. Check the stop condition if: ($|P_t - P_{t-1}| < \xi$) or ($t > MaxIter$) then stop, if not: $t = t + 1$, repeat step 4.

A flowchart is required to make it easier to identify the plot to locate the results of the cluster implementation when determining the cluster based on the available data using the Fuzzy C-Means clustering technique. The Fuzzy C-Means algorithm flowchart is shown in Figure 2.

Evaluation Stage of Determining the Best Method

Bunkers et al. (1996) stated that the performance of clustering results can be seen from the ratio value of the average standard deviation within the cluster (S_w) and the standard deviation between clusters (S_b). The formula for the standard deviation in the cluster (S_w) is as follows:

$$S_w = \frac{1}{K} \sum_{k=1}^K S_k \quad (13)$$

(Bunkers et al., 1996)

where,

S_k = Standard Deviation of the k^{th} cluster
 K = number of clusters

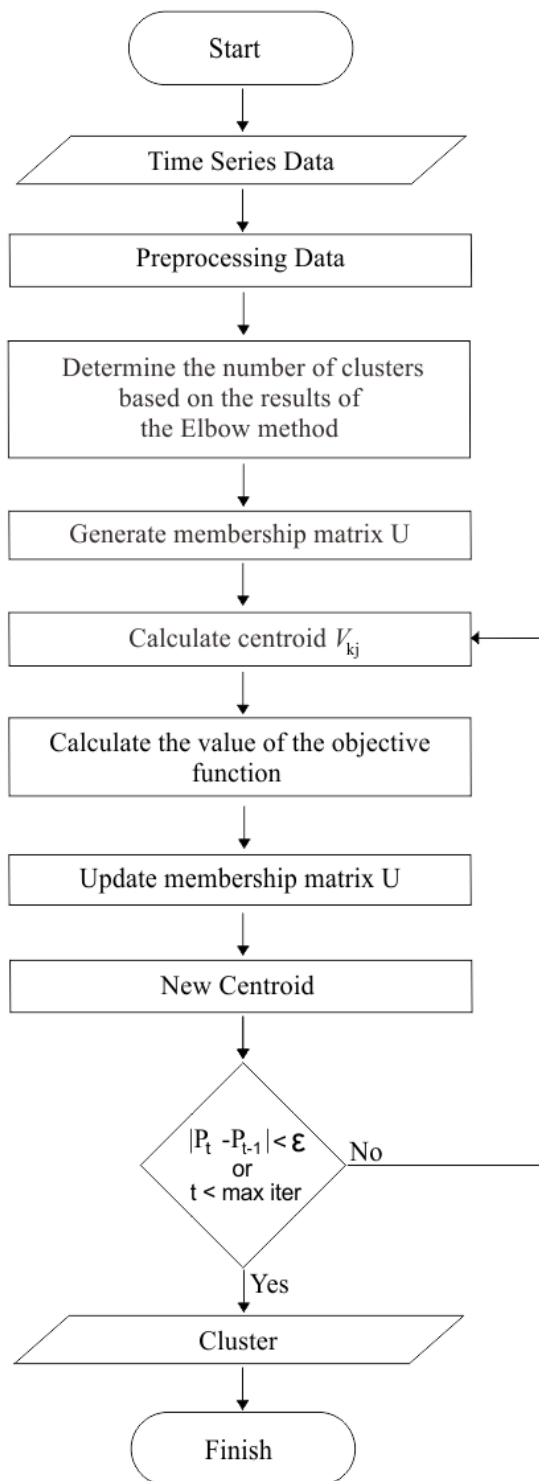


Figure 2. Fuzzy C-Means Flowchart

If given cluster c_k , d where $k = 1, \dots, p$ and each cluster has members of x_i , where $i = 1, \dots, n$ and n is the number of members from each cluster, and \bar{x}_k is the average of cluster k then to find the value of the k^{th}

(S_k) standard deviation, the following formula is used:

$$S_k = \sqrt{\frac{1}{N_k - 1} \sum_{k=1}^{N_k} (x_i - \bar{x}_k)^2} \quad (14)$$

where,

N_k = the number of k^{th} clusters

x_i = average member variable of i^{th}

\bar{x}_k = average cluster variable k^{th}

The standard deviation between clusters (S_b) is stated as follows:

$$S_b = \left[\frac{1}{K-1} \sum_{k=1}^K (\bar{x}_k - \bar{x})^2 \right]^{\frac{1}{2}} \quad (15)$$

where,

\bar{x} = mean of all clusters

The smaller the value of S_w and the greater the value of S_b it means that the method has a good performance or high homogeneity. The method chosen is the one that gives the smallest S_w/S_b (Ningrat et al., 2016). The following is the formula for the standard deviation ratio:

$$S = \frac{S_w}{S_b} \times 100\% \quad (16)$$

RESULTS AND DISCUSSION

The findings of testing the K-Means and Fuzzy C-Means algorithms on the Human Development Index (IPM) data for Central Java Province in 2021 will be described in this subsection. The total data is 36 regencies/cities with 4 indicators consisting of Life Expectancy (AHH) (X_1), Expected Years of School (HLS) (X_2), Average Years of Schooling (RLS) (X_3), and Expenditures Per Capita (PPK) (X_4). Central Java Province HDI indicator data in 2021 can be seen in Table 1.

Table 1. Central Java Province HDI Indicators in 2021

Regency/ City	AHH (X_1)	HLS (X_2)	RLS (X_3)	PPK (X_4)
Central Java Province	74.47	12.77	7.75	11034
Cilacap Regency	73.90	12.63	7.09	10534
Banyumas Regency	73.80	13.03	7.63	11546
Purbalingga Regency	73.21	12.00	7.25	10032
Banjarnegara Regency	74.28	11.63	6.75	9407
...
Tegal City	74.54	13.07	8.73	13143

The initial stage before doing clustering is checking for outliers in the data. In this research, the data were

checked for outliers using the Boxplot method.

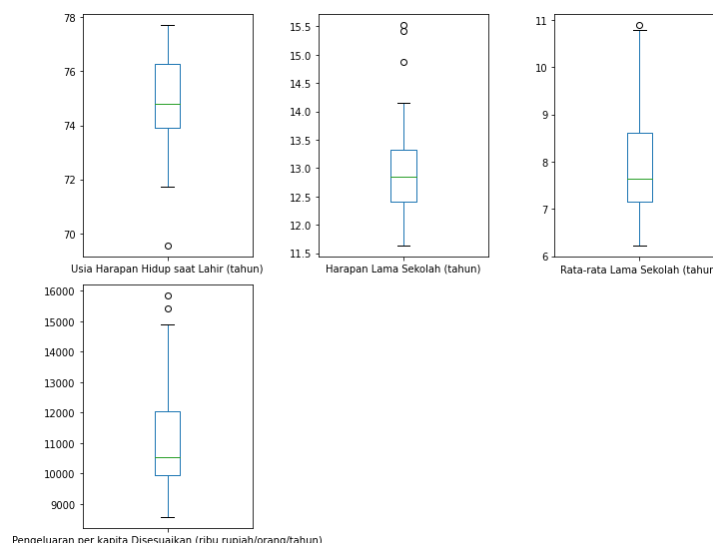


Figure 3. Outliers of Research Data

Figure 3 demonstrates that each variable has an outlier based on the Box Plot results. As a result, the data must be normalized. The Min-Max Scaler approach

was used to normalize the data in this investigation. The following table shows the normalization results:

Table 2. Normalization Result Data

AHH (X_1)	HLS (X_2)	RLS (X_3)	PPK (X_4)
0.6019536	0.2923076	0.3269230	0.33854
0.5323565	0.2564102	0.18589744	0.26973
0.5201465	0.3589743	0.3012820	0.40894
0.4481074	0.0948717	0.2200854	0.20067
0.5787545	0.0	0.1132478	0.11472
...
0.61050061	0.3692307	0.53632479	0.628617

The next step is to determine the best number of clusters using the Elbow method, as presented in Figure 4.

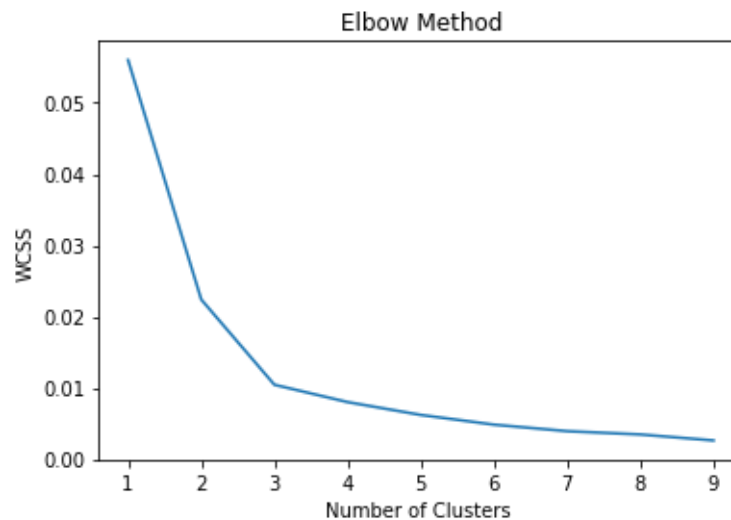


Figure 4. The Best Number of Clusters Based on the Elbow Method

According to the Elbow method, the value of 3 makes a right angle, indicating that the value of 3 is the optimum or best number of clusters (Winarta & Kurniawan, 2021).

The Analysis Results of the K-Means Algorithm

The first step in the K-Means clustering analysis is to determine the best

number of clusters using the Elbow method. Based on the Elbow method, the best number of clusters is 3 clusters. The next step is to perform cluster analysis using the K-Means method with the number of clusters 3 with a maximum iteration of 300 and a tolerance of 0.0001. The results of clustering using K-Means are shown in Figure 5.

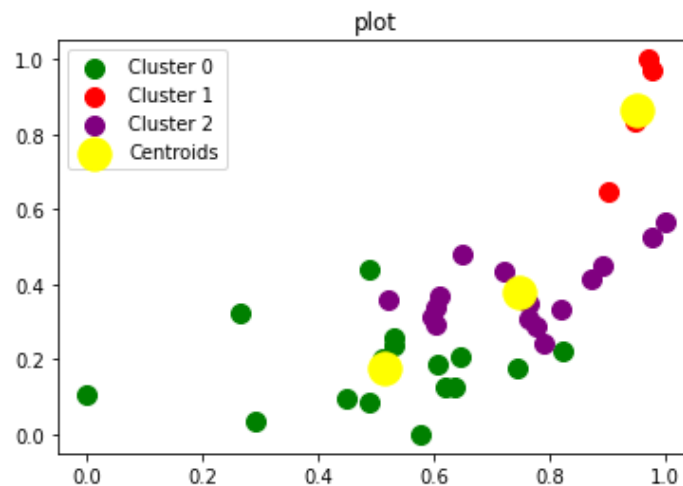


Figure 5. The Plot of K-Means Clustering Results

Figure 5 shows that cluster 0 is the lowest cluster, with 16 regencies/cities, cluster 1 is the highest, with 4 regencies/cities, and cluster 2 is an intermediate cluster, with 16 regencies/cities. Table 3 shows the values of the centroids of each variable as well as the members of each cluster.

Table 3. Centroid Value

Cluster	Centroid Value			
	X_1	X_2	X_3	X_4
0	0.513354	0.176602	0.179353	0.169446
1	0.950854	0.862820	0.965811	0.833425
2	0.748092	0.379487	0.422008	0.415508

Based on Table 3, the characteristics of each cluster can be described as follows:

1. Cluster 1 is a cluster that has the highest centroid value compared to other clusters. It implies that the districts/cities included in cluster 1 have Life Expectancy (AHH) (X_1), Expected Years of Schooling (HLS) (X_2), Average Years of Schooling (RLS) (X_3), and Per Capita Expenditure (PPK) (X_4). In other words, Regencies/Cities belonging to cluster 1 have high population growth.
2. Cluster 2 is a cluster that has the value of the second centroid. Therefore, the districts/cities included in cluster 2 have Life Expectancy (AHH) (X_1), Expected Years of Schooling (HLS) (X_2),

Average Years of Schooling (RLS) (X_3), and Per Capita Expenditure (PPK) (X_4). In other words, regencies/cities that are included in cluster 2 have fairly good population growth.

3. Cluster 0 is a cluster that has the lowest centroid value compared to other clusters. Thus, the districts/cities that are included in cluster 0 have Life Expectancy (AHH) (X_1), Expected Years of Schooling (HLS) (X_2), Average Years of Schooling (RLS) (X_3), and Expenditures Per Capita (PPK) (X_4) are low. In other words, Regencies/Cities belonging to cluster 0 have low population growth.

Members of the Regency/City of each cluster are presented in Table 4.

Results of the Fuzzy C-Means Algorithm Analysis

Clustering analysis with the FCM algorithm begins by determining the $n \times m$, size matrix, namely the size matrix of 35×4 . The number of clusters is determined based on the Elbow method so that the number of k clusters to be used is 3 with a maximum iteration of 300 and a tolerance of 0.0001. The results of clustering using FCM are presented in Figure 6.

Table 4. Results of Regency/City K-Means Clustering

Cluster	Regency/City	Total
0	Cilacap Regency, Purbalingga Regency, Banjarnegara Regency, Kebumen Regency, Wonosobo Regency, Magelang Regency, Wonogiri Regency, Grobogan Regency, Blora Regency, Rembang Regency, Temanggung Regency, Batang Regency, Pekalongan Regency, Pemaslang Regency, Tegal Regency,	15
1	Magelang City, Surakarta City, Salatiga City, Semarang City	4
2	Central Java Province, Banyumas Regency, Purworejo Regency, Boyolali Regency, Klaten Regency, Sukoharjo Regency, Karanganyar Regency, Sragen Regency, Pati Regency, Kudus Regency, Jepara Regency, Demak Regency, Semarang Regency, Kendal Regency, Pekalongan City, Tegal City, Regency Brebes	17

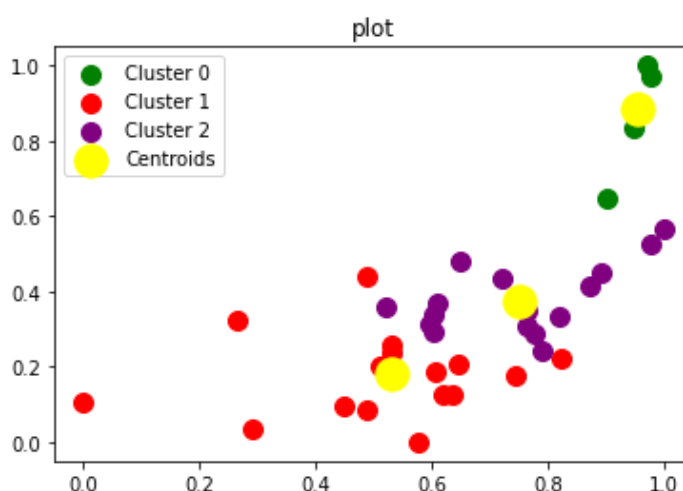


Figure 6. FCM Clustering Results

Figure 6 shows that cluster 0 is the highest cluster with four regencies/cities, cluster 1 is the lowest cluster with sixteen regencies/cities, and cluster 2 is an

intermediate cluster with sixteen regencies/cities. Table 5 shows the values of the centroids for each variable.

Table 5. The Centroid Value of FCM

Cluster	Centroid Value			
	X_1	X_2	X_3	X_4
0	0.9546563	0.8856978	0.9554999	0.8733735
1	0.5321224	0.1837012	0.1935349	0.191009
2	0.7516965	0.3728712	0.4198855	0.410793

According to Table 5, the characteristics of each cluster are as follows:

1. Cluster 0 is a cluster that has the highest centroid value compared to other clusters. Thus, the districts/cities that are included in cluster 0 have Life Expectancy (AHH) (X_1), Expected Years of Schooling (HLS) (X_2), Average Years

of Schooling (RLS) (X_3), and Per Capita Expenditure (PPK) (X_4). In other words, districts/cities belonging to cluster 0 have high population growth.

2. Cluster 2 is a cluster that has the value of the second centroid. Thus, the districts/cities included in cluster 2 have Life Expectancy (AHH) (X_1), Expected Years of Schooling (HLS) (X_2),

Average Years of Schooling (RLS) (X_3), and Per Capita Expenditure (PPK) (X_4). In other words, districts/cities that are included in cluster 2 have fairly good population growth.

3. Cluster 1 is a cluster that has the lowest centroid value compared to other clusters. Therefore, the districts/cities that are included in cluster 1 have Life

Expectancy (AHH) (X_1), Expected Years of Schooling (HLS) (X_2), Average Years of Schooling (RLS) (X_3), and Expenditures Per Capita (PPK) (X_4) which are low. In other words, Regencies/Cities belonging to cluster 1 have low population growth.

Members of the Regency/City of each cluster are shown in Table 6.

Table 6. Results of Regency/City Fuzzy C-Means Clustering

Cluster	Regency/City	Total
0	Magelang City, Surakarta City, Salatiga City, Semarang City	4
1	Demak Regency, Purbalingga Regency, Banjarnegara Regency, Kebumen Regency, Wonosobo Regency, Magelang Regency, Wonogiri Regency, Grobogan Regency, Blora Regency, Purworejo Regency, Temanggung Regency, Batang Regency, Pekalongan Regency, Pemalang Regency, Tegal Regency, Brebes Regency	16
2	Central Java Province, Banyumas Regency, Rembang Regency, Boyolali Regency, Klaten Regency, Sukoharjo Regency, Karanganyar Regency, Sragen Regency, Pati Regency, Kudus Regency, Jepara Regency, Cilacap Regency, Semarang Regency, Kendal Regency, Pekalongan City, Tegal City	16

Evaluation of Determining the Best Clustering Method

The results of the evaluation using the S_w and S_b methods from the K-Means and Fuzzy C-Means methods are presented in Table 7.

Table. 7 Evaluation Results of the Standard Deviation Ratio Method

Number of Clusters	Clustering Method	Ratio Value S_w/S_b
3	K-Means	0.478369
3	Fuzzy C-Means	0.460093

Table 7 shows that the value of the S_w/S_b ratio of the FCM method is smaller than that of the K-Means method, with values of 0.460093 and 0.473601, respectively. This indicates that the best clustering method that can be used in grouping districts/cities based on the HDI indicator is the FCM method. According to research, the Fuzzy C-means algorithm outperforms the K-means algorithm for segmenting blood veins in the retina (Wiharto & Suryani, 2020). According to research, the FCM algorithm is superior to the K-means method for creating clusters

on wireless sensor networks (Hassan et al., 2020).

CONCLUSIONS AND SUGGESTIONS

Based on the results of the research, the best number of clusters from the Elbow method was 3. In the K-Means clustering analysis, it was found that cluster 1 is the best cluster with 4 regencies/cities, cluster 2 is the second cluster with 16 regencies/cities, and cluster 0 is the lowest cluster with 16 regencies/cities as members. Cluster 0 is the best cluster with 4 regencies/cities, cluster 2 is the second-best cluster with 16 regencies/cities, and cluster 1 is the lowest cluster with 16 regencies/cities. After analyzing the standard deviation ratio method, it was found that the FCM clustering method was better than the K-Means clustering method. FCM standard deviation ratio is 0.460093, whereas K-Means standard deviation ratio is 0.478369.

To continue evaluating the performance of the *clustering* model, here are two suggestions for future research as

the proposed model some more effective optimization techniques to deal with more complicated and multifactor problems for decision making such as Fuzzy K-Medoids or Fuzzy Subtractive Clustering.

REFERENCES

- Agustini, F. (2017). Implementasi algoritma fuzzy c-means studi kasus. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, 3(1), 127–132.
- BPS Jateng. (2021). *Indeks pembangunan manusia (metode baru) 2019-2021*. <https://jateng.bps.go.id/indicator/26/83/1/indeks-pembangunan-manusia-metode-baru-.html>
- Bunkers, M. J., Miller, J. R., & DeGaetano, A. T. (1996). Definition of climate regions in the northern plains using an objective cluster modification technique. *Journal of Climate*, 9(1). [https://doi.org/10.1175/1520-0442\(1996\)009<0130:DOCRIT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<0130:DOCRIT>2.0.CO;2)
- Ghazal, T. M., Hussain, M. Z., Said, R. A., Nadeem, A., Hasan, M. K., Ahmad, M., Khan, M. A., & Naseem, M. T. (2021). Performances of k-means clustering algorithm with different distance metrics. *Intelligent Automation and Soft Computing*, 30(2), 735–742. <https://doi.org/10.32604/iasc.2021.019067>
- Haqiqi, B. N., & Kurniawan, R. (2015). Analisis perbandingan metode fuzzy c-means dan subtractive fuzzy c-means. *Media Statistika*, 8(2), 59–67. <https://doi.org/10.14710/medstat.8.2.59-67>
- Hassan, A. A. hussian, Shah, W. M., Othman, M. F. I., & Hassan, H. A. H. (2020). Evaluate the performance of k-means and the fuzzy c-means algorithms to formation balanced clusters in wireless sensor networks. *International Journal of Electrical and Computer Engineering*, 10(2), 1515–1523. <https://doi.org/10.11591/ijece.v10i2.pp1515-1523>
- Irawan, J., Handayani, A. A. A. T., & Zohri, L. H. N. (2021). Operasionalisasi ibm spss 21 untuk meningkatkan kemampuan dan keterampilan olah data penelitian mahasiswa. *Jurnal Pengabdian Magister Pendidikan IPA*, 4(2). <https://doi.org/10.29303/jpmpi.v4i2.660>
- J. Mac Queen. (1967). *Some methods for classification and analysis of multivariate observations*. <https://doi.org/10.1007/s11665-016-2173-6>
- Kirana, I. O., Nasution, Z. M., & Wanto, A. (2019). Proyeksi indeks pembangunan manusia di indonesia menggunakan metode statistical parabolic dalam menyongsong revolusi industri 4.0. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 16(2), 202. <https://doi.org/10.23887/jptk-undiksha.v16i2.18178>
- Kolay, S., Ray, K. S., & Mondal, A. C. (2017). K+ means: An enhancement over k-means clustering algorithm. In *arXiv*.
- Le, T., & Altman, T. (2011). A new initialization method for the fuzzy c-means algorithm using fuzzy subtractive clustering. *Proc. Intl'Conf. on Information and Knowledge Engineering, Las Vegas USA*, 144–150.
- Ningrat, D. R., Maruddani, D. A. I., & Wuryandari, T. (2016). Analisis cluster dengan algoritma k-means dan fuzzy c-means clustering untuk pengelompokan data obligasi korporasi. *None*, 5(4), 641–650.
- Oktarina, C., Notodiputro, K. A., & Indahwati, I. (2020). Comparison of k-means clustering method and k-medoids on twitter data. *Indonesian Journal of Statistics and Its Applications*, 4(1), 189–202. <https://doi.org/10.29244/ijisa.v4i1.599>

- Purba, W., Tamba, S., & Saragih, J. (2018). The effect of mining data k-means clustering toward students profile model drop out potential. *Journal of Physics: Conference Series*, 1007(1). <https://doi.org/10.1088/1742-6596/1007/1/012049>
- Rahakbauw, D. L., Ilwaru, V. Y. I., & Hahury, M. H. (2017). Implementasi fuzzy c-means clustering dalam implementation of fuzzy c-means clustering in. *Jurnal Ilmu Matematika Dan Terapan*, 11, 1-12.
- Ramadhan, A., Mustakim, & Handinata, R. (2019). *Implementasi algoritma fuzzy c means dan moora untuk pengelompokan dan penentuan wilayah penanggulangan bencana banjir*.
- Sutoyo, M. N., & Sumpala, A. T. (2016). Penerapan fuzzy c-means untuk deteksi dini kemampuan penalaran matematis. *Scientific Journal of Informatics*, 2(2), 129. <https://doi.org/10.15294/sji.v2i2.5080>
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conference Series: Materials Science and Engineering*, 336(1), 0-6. <https://doi.org/10.1088/1757-899X/336/1/012017>
- Utomo, W. (2021). The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in indonesia. *ILKOM Jurnal Ilmiah*, 13(1), 31-35. <https://doi.org/10.33096/ilkom.v13i1.763.31-35>
- Wakhidah, N. (2010). Clustering menggunakan k-means algorithm. *Jurnal Transformatika*, 8(1), 33. <https://doi.org/10.26623/transformatika.v8i1.45>
- Wiharto, W., & Suryani, E. (2020). The comparison of clustering algorithms k-means and fuzzy c-means for segmentation retinal blood vessels. *Acta Informatica Medica*, 28(1), 42-47. <https://doi.org/10.5455/AIM.2020.28.42-47>
- Winarta, A., & Kurniawan, W. J. (2021). Optimasi cluster k-means menggunakan metode elbow pada data pengguna narkoba dengan pemrograman python. *Jurnal Teknik Informatika Kaputama (JTik)*, 5(1), 113-119.
- Windarto, A. P., Hasan Siregar, M. N., Suharso, W., Fachri, B., Supriyatna, A., Carolina, I., Efendi, Y., & Toresa, D. (2019). Analysis of the k-means algorithm on clean water customers based on the province. *Journal of Physics: Conference Series*, 1255(1). <https://doi.org/10.1088/1742-6596/1255/1/012001>
- Zhou, K., & Yang, S. (2020). Effect of cluster size distribution on clustering: A comparative study of k-means and fuzzy c-means clustering. *Pattern Analysis and Applications*, 23(1), 455-466. <https://doi.org/10.1007/s10044-019-00783-6>

