

# Implementation of Fuzzy C-Means and K-Medoids in Grouping People's Welfare Indicators (Case Study in Riau Province in 2020)

Safira Naila Farafisha<sup>1</sup>, Achmad Fauzan<sup>2</sup>, Afdelia Novianti<sup>3</sup>  
{18611144@students.uii.ac.id<sup>1</sup>, achmadfauzan@uui.ac.id<sup>2</sup>, 18611082@students.uui.ac.id<sup>3</sup>}

Statistics Department, Faculty of Mathematics and Natural Science, Universitas Islam Indonesia,  
Yogyakarta, Indonesia<sup>1</sup>  
Statistics Disaster Management Laboratory, Statistics Department, Faculty of Mathematics and  
Natural Science, Universitas Islam Indonesia, Yogyakarta, Indonesia<sup>2</sup>  
Statistics Department, Faculty of Mathematics and Natural Science, Universitas Islam Indonesia,  
Yogyakarta, Indonesia<sup>3</sup>

**Abstract.** Riau Province is one of the provinces that has social inequality in its society. This is due to the uneven distribution of development carried out in remote areas that are difficult for the government to access. Therefore, the researchers grouped districts in Riau Province with a low welfare index by clustering using Fuzzy C-Means and K-Medoids methods. Clustering is a method of grouping objects or events into groups based on their similarity values. Later, a comparison will be made between the Fuzzy C-Means and K-Medoids methods to get the best result on the Riau Province community welfare data in 2020. From the analysis that has been carried out, there is a grouping of data into 3 clusters for Fuzzy C-Means method and 2 clusters for K-Medoids. Based on the result of the variances comparison of the two methods, namely Fuzzy C-Means and K-Medoids, each of which has a variance of 0.375681 and 3.248903, it can be concluded that the best cluster method for this case is Fuzzy C-Means. Based on the results of clustering on districts/cities by Fuzzy C-Means method can be applied as a reference by related parties to improve the welfare of the people of the Riau province.

**Keywords:** Riau Province, Clustering, Fuzzy C-Means, K-Medoids, Variance.

## 1. Introduction

The government's development has not all been evenly distributed to small districts that are difficult to access in Riau Province, which causes inequality in the welfare of society. People's welfare is the right of citizens to enjoy life in the country they live in. A country certainly has a goal in the development of the country itself, one of which is the welfare of its people. Prosperity here is defined as a state of peace and prosperity which is also interpreted as always sufficient and not lacking in life, not lacking in physical, material, and spiritual.

In measuring people's welfare, it is necessary to calculate people's welfare indicators, namely the level of population welfare as a reflection of the quality level of a country's human resources, which includes indicators of population, environment, consumption, housing, education, nutritional health, expenditure, employment, etc.

The house as a place to live is one of the primary human needs, where a decent home or place to live in the dream of every household in urban and rural areas. A comfortable and peaceful place to live is one level to improve life in the family. A habitable house is a house that meets several criteria, namely building resilience (building materials, roofs, walls, and floors of the house that meet the requirements), minimum adequacy of living space (floor area per capita 7.2 m<sup>2</sup>), has electricity as a source of lighting. Primary care have access to proper drinking water, and have access to adequate sanitation. A household is categorized as occupying a livable house if there are a maximum of two criteria that are not met from the seven forming indicators [1].

This study will compare the Fuzzy C-Means and K-Medoids Clustering methods to discover the best clustering method for grouping districts in Riau based on welfare indicators. Fuzzy C-Means is a clustering technique that uses weighting on each member of the set [2]. K-Medoids is a clustering technique that can handle objects among high values that may deviate from their distribution [3].

## 2. Literature Study

### 2.1. Assumptions in Clustering

There is an assumption that must be met in cluster analysis, and there is no multicollinearity [4]. A multicollinearity test is needed to determine the existence of a linear relationship between 2 or more variables. One way to see the symptoms of multicollinearity can be detected by calculating the tolerance value or Variance Inflation Factor (VIF) [5]. If the VIF value exceeds 10, it indicates that there is multicollinearity between variables. If multicollinearity occurs, it is recommended to eliminate one of the two variables with a fairly significant correlation [6]. VIF value is calculated as Equation 1.

$$VIF = \frac{1}{1 - R_i^2} \quad (1)$$

Where  $R_i^2$  is the coefficient of determination.

### 2.2. Fuzzy C-Means Clustering

Fuzzy C-Means is a clustering approach in which the existence of each data point in a cluster is defined by the degree of membership [7]. The goal of this approach is to obtain the cluster's center, which will then be utilized to determine the data that enter the cluster [8]. The explanation of the Fuzzy C-Means algorithm is as follows [9]:

- 1) Input data to be grouped, namely  $X$  in the order of a sized matrix  $n \times m$  ( $n$  = quantity of data samples and  $m$  = attributes of any data).  $X_{ij}$  =  $i^{\text{th}}$  sample data ( $i = 1, 2, \dots, n$ ),  $j^{\text{th}}$  attribute ( $j = 1, 2, \dots, m$ ).
- 2) The number of clusters ( $c$ ), the partition matrix's power ( $w$ ), the maximum iteration (MaxIter), the least anticipated error ( $\xi$ ), the initial objective function ( $P_0 = 0$ ), and the starting iteration ( $t = 1$ ) are all variables to consider.
- 3) As elements of the initial partition matrix  $U$ , generate random integers  $\mu_{ik}$ , where  $i = 1, 2, \dots, n$  and  $k = 1, 2, \dots, c$ . Then count how many columns of random numbers you've created. Count how many of each column (attribute) by the equation 2.

$$Q_i = \sum_{k=1}^c \mu_{ik} \quad (2)$$

Calculate the first partition matrix  $U$  after adding up each column of the random number determined. Equation 3 may be used to obtain the initial partition matrix  $U$ .

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (3)$$

- 4) Based on Equation 3, calculate the  $k^{\text{th}}$  cluster center for the  $j^{\text{th}}$  attribute  $V_{kj}$ , with  $k = 1, 2, \dots, c$ ; and  $j = 1, 2, \dots, m$ .

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w \cdot X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (4)$$

$\mu_{ik}$  membership degree for the  $i^{\text{th}}$  sample in the  $k$ -cluster and  $X_{ij}$  is the  $i^{\text{th}}$  data with  $j^{\text{th}}$  attribute.

- 5) Calculate the objective function at the iteration  $t$  using the Equation 5:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (5)$$

$P_t$  is the objective function in the  $t^{\text{th}}$  iteration.

- 6) Calculate the change in the partition matrix using the Equation 6:

$$\mu_{ik} = \frac{\left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{1}{w-1}}}{\sum_{k=1}^c \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{1}{w-1}}} \quad (6)$$

- 7) Check the stop condition: if  $t > \text{MaxIter}$ , stop; if not, continue to step 4 with  $t = t + 1$ .

The initial fuzzy clustering approach is to find the cluster's center, which will serve as the average position for every cluster. After a series of upgrades to the cluster center, the cluster center will be moved to its proper position [10]. Each data will be given a value or weight to enter each cluster. Any weighted data that has the closest rarity to a cluster center will have a higher value. In fuzzy, the degree of membership is known, which has a value range of 0 to 1, in contrast to the firm set with a value of 1 or 0 [11]. The Fuzzy C-Means method is one of the simplest and is frequently used in data grouping approaches because it produces accurate estimates with few parameters. The Fuzzy C-Means method may be used to categorize data depending on specific qualities, according to various studies [12].

### 2.3. K-Medoids Clustering

K-Medoids or The Partitioning Around Medoid (PAM) clustering algorithm reduces the distance between the labeled point in the cluster and the point chosen as the cluster center [13]. The K-Medoids technique is being used to address the shortcomings of the k-means algorithm, which is extremely sensitive to outliers. Outliers are pretty far away from most data; thus, they can alter the cluster's average value (mean) [14]. K-Medoids Clustering will arrange a set of  $n$  items into multiple  $k$  clusters through clustering partitions. This technique works by employing medoids, which are items in a collection of objects that constitute a cluster. Then cluster is formed by calculating the closeness between medoids and non-medoids objects [15]. The steps of K-Medoids clustering in grouping the data are as follows [16].

- 1) Create a random number of clusters ( $k$ ).
- 2) Using the euclidean distance measure equation, calculate each object to the nearest cluster:

$$D_{ik} = \sqrt{\sum_k^n (x_{ij} - c_{kj})^2} \quad (7)$$

- $D_{ik}$  : Euclidean Distance,  $x_i$  : Data ( $i$ ),  $x_j$  : Data ( $j$ ),  $x_{ij}$  : Data ( $i$ ) attribute ( $j$ ),  $c_{kj}$  : Data ( $k$ ) attribute ( $j$ ).
- 3) Select a random object from each cluster as a candidate for new medoids.
  - 4) With new medoids, calculate the distance from each object in each cluster.
  - 5) By calculating new total distance - old total distance. Calculate standard deviation (S). If  $S < 0$ , swap objects with data cluster medoids to form a new set of  $k$  objects as medoids.
  - 6) Repeat steps 3 to 5 until there is no medoid change so that clusters and their respective cluster members are obtained.

The Within Sum Square (WSS) method defines the number of clusters formed by internal criteria. Validation with internal criteria focuses on measuring how compact the resulting cluster is either indicated by intra-cluster homogeneity, inter-cluster separation, or a combination of both using only internal data [11]. WSS is a method of evaluating intra-cluster variability. Each observation will be assigned to the closest cluster. The distance between the observations and the cluster will be determined using the Cosine Equation between the observations and the centroid. The next centroid will be changed as the average for observations on every cluster [17].

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{kj})^2 \quad (8)$$

$x_{ij}$  : Data ( $i$ ) cluster ( $j$ ) and  $\bar{x}_{kj}$ : Mean cluster ( $j$ )

#### 2.4. Determining the Best Method

To compare two or more methods in clustering, the average standard deviation within the cluster and the standard deviation between clusters will be use [18]. Amount cluster optimum by minimizing the index value[19]. Standard deviation within-cluster ( $v_w$ ):

$$v_w = \frac{1}{N-k} \sum_{i=1}^k (n_i - 1) \times v_i^2 \quad (9)$$

$v_i^2$  : Variance cluster ( $i$ ),  $n_i$ : Count of data in a cluster ( $i$ ),  $k$  : Count of a cluster,  $N$ : Count of data

Standard deviation between cluster ( $v_b$ ):

$$v_b = \frac{1}{k-1} \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2 \quad (10)$$

$\bar{x}_i$ : mean data ( $i$ ) in cluster and  $\bar{x}$ : mean data ( $i$ ). The variance of a cluster:

$$v = \frac{v_w}{v_b} \quad (11)$$

$v$  : variance of a cluster,  $v_w$ : variance within a cluster,  $v_b$ : variance between clusters.

### 3. Research Methods

#### 3.1. Data and Sources

Riau Province is the focus of this study. The information utilized came from the Riau Province's Central Statistics Agency's website. Secondary data is the type of data used in this study. The variables that were sampled were eight variables consisting of the percentage of households with walls and wood ( $X_1$ ); percentage of households with roofs of concrete, tile, zinc, asbestos, wood, and bamboo ( $X_2$ ); the percentage of families according to their housing

ownership status ( $X_3$ ); percentage of households with clean drinking water sources ( $X_4$ ); percentage of non-ground floor households ( $X_5$ ); the percentage of households with the main source of lighting ( $X_6$ ); percentage of labor force participation rate (TPAK) ( $X_7$ ); percentage of open unemployment rate (TPT) ( $X_8$ ). The data used in this study were 12 locations in Riau Province.

### 3.2. Data Analysis Stage

The first step is to collect secondary data that has been taken from the [riau.bps.go.id](http://riau.bps.go.id) website, then pre-processing the data on people's welfare indicators in Riau Province in 2020. The stages are continued by analyzing the characteristics of the people's welfare indicator data in Riau Province in 2020. Before clustering analysis is carried out, we will check the assumption of the multicollinearity test. After that, a grouping or clustering analysis will be carried out using the Fuzzy C-Means and K-Medoids methods. After comparing the goodness of clusters by comparing the variance of two methods, the best approach will be chosen, and profiling will be made—all analysis using the Rstudio application.

## 4. Results and Discussion

### 4.1. Descriptive Analysis

This descriptive analysis of the Riau Province people's welfare indicator data in 2020 is displayed in the form of a descriptive statistical table containing the value of median, mean, variance, standard deviation, maximum and maximum, and the upper quartile value and that of the data; it looks like the Table 1.

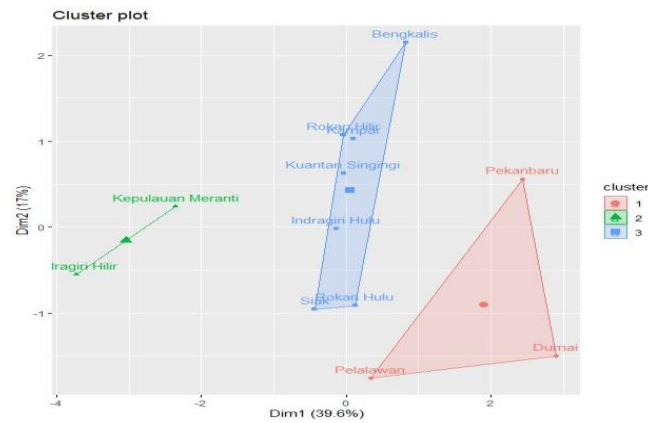
**Table 1.** Descriptive analysis of the welfare of people in Riau in 2020

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$
Mean	97.58	94.42	72.21	68.70	99.16	84.74	65.57	5.80
Median	99.12	96.38	74.61	74.15	99.08	87.59	65.53	4.96
Std. Deviation	4.31	4.83	11.72	24.24	0.51	8.77	2.71	1.56
Variance	18.59	23.30	137.47	587.46	0.265	76.97	7.36	2.45
Minimum	84.51	81.66	51.26	12.06	98.20	68.63	60.90	4.33
Maximum	100	98.12	89.89	94.16	100	95.81	69.28	9.62
1st Qu.	97.94	93.04	65.45	66.70	98.91	80.70	64.37	4.85
3rd Qu.	99.58	97.68	80.61	83.78	99.54	90.59	67.74	6.26

Based on Table 1, the highest average value is in the  $X_5$  variable or the percentage of non-land floor households of 99.16, while the lowest average value is in the  $X_8$  variable or the percentage of the open unemployment rate of 5.80. Variable  $X_8$  has an average of 97.58; the average variable  $X_2$  is 94.42; the average variable  $X_3$  is 72.21; the average variable  $X_4$  is 68.70; the average variable  $X_6$  is 84.74; the average variable  $X_7$  is 65.57, and the average variable  $X_8$  is 5.80. Of these eight variables, the one with the highest variance value is in the  $X_4$  variable or the percentage of households with clean drinking water sources of 587.46, which means that the diversity of data on this variable is significant.

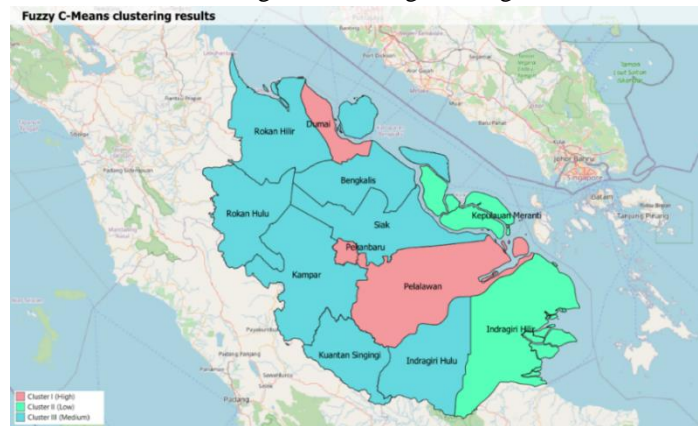
#### 4.2. Fuzzy C-Means Clustering

For the assumption of multicollinearity from the results of the VIF value, no variable has a VIF value > 10; then, the assumption is met. Fuzzy C-Means cluster analysis outcomes are presented in the form of a cluster plot with three clusters, as illustrated in Figure 1.



**Fig. 1.** Plot Fuzzy C-Means Clustering

Based on the cluster plot in Figure 1, the results obtained are districts/cities in the province of Riau based on the welfare of the people using the Fuzzy C-Means method divided into three colors. Cluster 1 is represented by a red area, cluster 2 is represented by a green area, and cluster 3 is represented by a blue area. Based on Fig 1, it is arranged in Figure 2.



**Fig. 2.** Maps Fuzzy C-Means Clustering

As indicated in Table 2, the profiling result is based on the computation of the average value from each cluster's variables.

**Table 2.** Fuzzy C-Means cluster profiling

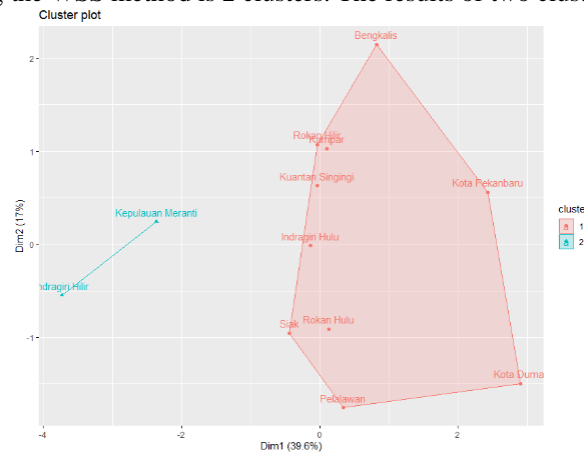
Variable	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>
Cluster 1 (High)	99.7	97.1	55.8	90.2	97.5	87.1	68.2	6.40

Variable	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>
Cluster 2 (Low)	92.3	90.5	86.4	21.2	99.8	71.3	66.1	5.48
Cluster 3 (Medium)	98.2	94.4	75.2	73.0	98.8	87.5	64.3	5.63

From table 2, it can be explained that cluster 1 is a grouping with variables that have the highest average value, almost all of the variables have the highest value among the three clusters except for the variables X<sub>3</sub> and X<sub>5</sub>. Cluster 2 is a grouping with variables with the lowest average value, and almost all the variables have the lowest values among the three clusters except for variables X<sub>3</sub> and X<sub>5</sub>. And cluster 3 is a grouping with variables that have a medium average value.

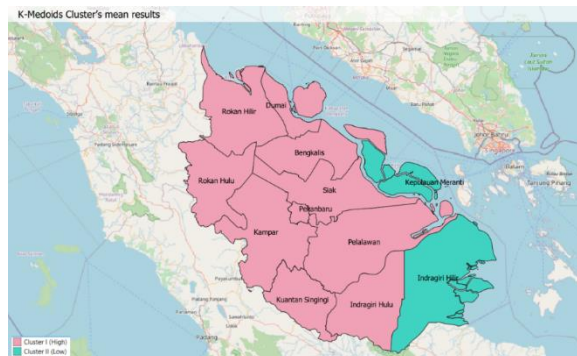
### 4.3. K-Medoids Clustering

Using K-Medoids method analysis, the best cluster that describes people's welfare in Riau Province using the WSS method is 2 clusters. The results of two clusters, as shown in Fig 3.



**Fig. 3.** Plot K-Medoids Clustering

Based on the computation of the similarity between observations of people's wellbeing in Riau Province, the output of K-Medoids reveals that there are two clusters. Based on Figure 3, the results of the clustering groups are in Figure 4.



**Fig 4.** Maps K-Medoids Clustering

According to the K-medoids method, cluster one consists of ten districts, whereas cluster two consists of two districts. The results of the profiling of the average value computation of each cluster group occurrence are presented in Table 3.

**Table 3.** K-Medoids Cluster's mean

Variable	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>
Cluster 1 (High)	98.64	95.201	69.358	78.186	98.424	87.422	65.467	5.865
Cluster 2 (Low)	92.25	90.54	86.445	21.25	99.825	71.345	66.07	5.475

From table 3, the results of profiling using the K-medoids clustering method in cluster 1 consisting of 10 districts have the best average for variables X<sub>1</sub>, X<sub>2</sub>, X<sub>4</sub>, and X<sub>6</sub>. While cluster 2 has two district members with the best average on X<sub>3</sub>, X<sub>5</sub>, X<sub>7</sub>, and X<sub>8</sub>. Based on the high average of each indicator, because cluster 1 and cluster 2 have the same score, next see the average distance. For example, even if cluster two is better than cluster one in X<sub>5</sub>, but the average distance is not far, it is different from X<sub>4</sub>. Cluster one and cluster two has a very long distance. It will be determined that cluster 1 has a high level of people's welfare and cluster 2 is a low.

#### 4.4. Comparison of Variance of Fuzzy C-Means and K-Medoids Method

Based on equation 6-8, it is continued by determining the best method presented in Table 4-6

**Table 4.** Fuzzy C-Means Variance

Fuzzy C-Means	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>
Vw	13.17	15.06	39.90	168.86	10.57	27.46	8.63	0.98
Vb	72.36	50.96	1270.94	6004.01	7.27	423.2	30.77	1.46
Vw/Vb	0.18	0.29	0.031	0.028	1.45	0.06	0.28	0.66
Var	<b>0.375</b>							

**Table 5.** K-Medoids Variance

K-Medoids	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>
Vw	13.63	22.01	102.56	105.92	3.42	41.59	8.04	2.68
Vb	68.14	36.21	486.61	5402.85	3.27	430.78	0.61	0.25
Vw/Vb	0.20	0.61	0.21	0.02	1.04	0.10	13.26	10.55
Var	<b>3.25</b>							

**Table 6.** Comparison of Variance

Method	Variance
Fuzzy C-Means	0.375
K-Medoids	3.25

Based on Table 6, the value of the variance ratio shows that the Fuzzy C-Means method has better performance than the K-Medoids method because the variance of the Fuzzy C-Means method is more petite than the K-Medoids method.



## 5. Conclusion

Based on the explanation of the previous sections, it can be concluded that the Fuzzy C-Means method's results obtained 3 clusters with cluster one is a group with a relatively high state of community welfare, cluster two is a group with low community welfare conditions, and cluster three is a group with moderate community welfare conditions in Riau Province in 2020. At the same time, the K-Medoids method's result obtained 2 clusters with cluster 1 of 10 districts with relatively high community conditions and cluster two with low community conditions. The best method is the method with minor variance. In this study, the Fuzzy C-Means method is the best method with the value of variance is 0.375. From the results of the research that has been done, it can be seen that several areas in Riau Province have a relatively low level of people's welfare. Based on this research, policymakers can pay more attention to the welfare conditions of the community so that all people in the area receive equitable welfare.

## REFERENCE

- [1] I. BPS-Statistics, *Welfare Indicators*. BPS-Statistics Indonesia, 2018.
- [2] J. Tamaela, E. Sedyono, and A. Setiawan, "Cluster analysis menggunakan algoritma Fuzzy C-means dan K-means untuk klasterisasi dan pemetaan lahan pertanian di Minahasa Tenggara," *J. Buana Inform.*, vol. 8, no. 3, pp. 151–160, 2017, doi: 10.24002/jbi.v8i3.1317.
- [3] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartama, "Analisis algoritma K-Medoids clustering dalam pengelompokan penyebaran Covid-19 di Indonesia," *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 166–173, 2020.
- [4] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate data analysis (seventh edition)*. 2009.
- [5] S. Santoso, *Menguasai statistik parametrik konsep dan aplikasi dengan SPSS*. Jakarta: PT Elex Media Komputindo, 2015.
- [6] A. Fauziah, "Hierarchical cluster analysis industri manufaktur besar dan sedang berdasarkan status penanaman modal (studi kasus: industri manufaktur besar dan sedang di Jawa Tengah tahun 2015)," Universitas Islam Indonesia, 2019.
- [7] S. Kusumadewi and P. Hery, *Aplikasi logika fuzzy untuk pendukung keputusan edisi 2*. Yogyakarta: Graha Ilmu, 2010.
- [8] A. S. Rizal, "Metode K-Means cluster dan fuzzy C-Means cluster," in *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS*, 2015.
- [9] A. K. Wijaya, "Implementasi data mining dengan algoritma fuzzy C - Means (studi kasus penjualan di UD Subur Baru)," *Jur. Tek. Inform. FASILKOM UDINUS*, pp. 1–8, 2014.
- [10] D. L. Rahakbauw, V. Y. I. Ilwaru, and M. H. Hahury, "Implementasi Fuzzy C-Means Clustering dalam penentuan beasiswa," vol. 11, pp. 1–12, 2017.
- [11] V. A. P. Sangga, "Perbandingan algoritma K-Means dan algoritma K-Medoids dalam pengelompokan komoditas peternakan di provinsi Jawa Tengah tahun 2015," Universitas Islam Indonesia, 2015.
- [12] E. R. Syahputra, Y. A. Dalimunthe, and Irvan, "Application of fuzzy C-Means algorithm for determining field of interest in information system study STTH Medan," *J. Phys. Conf. Ser.*, vol. 930, no. 1, 2017, doi: 10.1088/1742-6596/930/1/012014.
- [13] N. L. Anggreini, "Teknik clustering dengan algoritma K-Medoids untuk menangani strategi promosi di politeknik TEDC Bandung," *J. Teknol. Inf. dan Pendidik.*, vol. 12, no. 2, pp. 1–7, 2019.
- [14] N. . Kaur, "K-medoids clustering algorithma," *Int. J. Comput. Appl. Technol.*, pp. 42–45.
- [15] A. W. Setiyawati, "Implementasi algoritma Partitioning Around Medoids (PAM) untuk pengelompokan sekolah menengah atas di DIY berdasarkan daya serap ujian nasional," Universitas Sanata Dharma, 2017.

- [16] Y. H. Chrisnanto and G. Abdillah, "Penerapan algoritma partitioning around medoids ( PAM ) clustering untuk melihat gambaran umum kemampuan akademik mahasiswa," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2015, no. Sentika, pp. 444–448, 2015.
- [17] Q. Zhao, M. Xu, and P. Fränti, "Sum-of-squares based cluster validity index and significance analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5495 LNCS, pp. 313–322, 2009, doi: 10.1007/978-3-642-04921-7\_32.
- [18] M. J. Bunkers and J. R. Miller, "Definition of climate regions in the Northern Plains using an objecting cluster modification technique," *J. Clim.*, vol. 9, pp. 130–145, 1996.
- [19] D. R. Ningrat, D. A. I. Maruddani, and T. Wuryandari, "Analisis cluster dengan algoritma K-Means dan Fuzzy C-Means clustering untuk pengelompokan data obligasi korporasi," *None*, vol. 5, no. 4, pp. 641–650, 2016.