

## COMPARISON OF FUZZY C-MEANS AND FUZZY GUSTAFSON-KESSEL CLUSTERING METHODS IN PROVINCIAL GROUPING IN INDONESIA BASED ON CRIMINALITY-RELATED FACTORS

**Bella Destia<sup>1</sup>, Mujiati Dwi Kartikasari<sup>2\*</sup>**

<sup>1,2</sup>Department of Statistics, Universitas Islam Indonesia  
Kaliurang Street, Km 14.5 Sleman, Yogyakarta, 55584, Indonesia

Corresponding author's e-mail: \*[mujiatikartikasari@uii.ac.id](mailto:mujiatikartikasari@uii.ac.id)

### ABSTRACT

#### Article History:

Received: 23th February 2023

Revised: 8<sup>th</sup> May 2023

Accepted: 10<sup>th</sup> May 2023

#### Keywords:

Clustering;

Crime;

Fuzzy;

Fuzzy C-Means;

Fuzzy Gustafson-Kessel.

Indonesia is a country that has a population density that is increasing every year. With the increase in population density, the crime rate in Indonesia is increasing. Criminal acts arise because they are supported by factors that cause crime. To improve the security and welfare of the Indonesian people, the authors grouped each province in Indonesia based on the factors that influence crime. This study uses a comparison of the Fuzzy C-Means Clustering (FCM) and Fuzzy Gustafson-Kessel Clustering (FGK) methods by using the validation index for determining the optimal cluster, namely the Davies Bouldin Index. The data used is secondary data in the form of variables forming factors that affect the crime rate in Indonesia, where the data obtained comes from the website of the Central Statistics Agency (BPS). The results obtained in this study for the FGK method are better than the FCM method because they have a smaller standard deviation ratio. The results of grouping using the best method, namely FGK, it was found that the optimal number of clusters formed was 5 clusters with the results of grouping cluster 1 consisting of 6 provinces, cluster 2 consisting of 4 provinces, cluster 3 consisting of 11 provinces, cluster 4 consisting of 5 provinces, and cluster 5 consisting of 8 provinces.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

#### How to cite this article:

B. Destia and M. Dwi Kartikasari., "COMPARISON OF FUZZY C-MEANS AND FUZZY GUSTAFSON-KESSEL CLUSTERING METHODS IN PROVINCIAL GROUPING IN INDONESIA BASED ON CRIMINALITY-RELATED FACTORS," *BAREKENG: J. Math. & App.*, vol. 17, iss. 2, pp. 1093-1102, June, 2023.

Copyright © 2023 Author(s)

Journal homepage: <https://ojs3.unpatti.ac.id/index.php/barekeng/>

Journal e-mail: [barekeng.math@yahoo.com](mailto:barekeng.math@yahoo.com); [barekeng\\_journal@mail.unpatti.ac.id](mailto:barekeng_journal@mail.unpatti.ac.id)

**Research Article** • **Open Access**

## 1. INTRODUCTION

Indonesia is a country that has a population density that is increasing every year. With increasing population density, the crime rate in Indonesia is expanding. As a country with various ethnicities and different characters, Indonesia certainly makes criminal acts a significant concern.

Criminal acts have become a very familiar thing. Because this action is permanently attached to the community's life in which various individuals and groups mingle. Each individual has a unique and different personality, which can trigger social conflict. This conflict can trigger the emergence of various criminal acts that occur in the community. An example of a conflict that has happened as an act of crime in Indonesia is the Sampit conflict, namely the conflict that occurred between the Dayak and Madurese tribes. The existence of the Dayak and Madura tribes on the island of Kalimantan, especially Central Kalimantan, creates competition between these tribes. The Dayak tribe, a native tribe from Kalimantan, felt rivaled and dissatisfied with the increasing number of Madurese, who, on average, had very competitive personalities, so inter-tribal conflicts arose and led to various criminal acts at that time.

Everyone's attitude and lifestyle usually depend on their views and opinions. One of the traits humans have is that they never feel enough or satisfied with what they have and always feel lacking because this trait makes a person towards an attitude and lifestyle called hedonism [1]. Hedonism can trigger criminal acts because people with a hedonistic view will do everything. They can get what they want regardless of whether the method is right or wrong, even though they don't need it [2]. Such people usually always want to be recognized with a luxurious lifestyle, so they are seen as slang by their social circle.

Criminal acts arise because they are supported by factors that cause crime. To improve the security and welfare of the Indonesian people, the researchers grouped each province in Indonesia based on the factors that influence the crime rate. This is expected to help the government in Indonesia to improve the performance of the region in tackling criminality in Indonesia. In this study, the researcher did not choose a specific type of crime because it was constrained by the data to be used, such as the researcher wanted to research related to crime with the category of crimes against life. In crimes such as murder, where the factors supporting murder generally occur because of a grudge from within a person, revenge is a factor that is difficult to measure, which will be processed as data.

Grouping, commonly known as clustering, is dividing data into several clusters or groups so that one cluster has the highest level and data between clusters has the lowest level [3]. In this study, researchers will classify several factors that are considered to give rise to crime, including the level of education, the level of the poor, the Gross Regional Domestic Product (PDRB) per capita, and the number of crime incidents in each province in Indonesia in 2020, refers to research conducted by Khairani and Ariesa [4].

After mentioning the factors used in this study, the researcher will group them using a flexible grouping method following the form of the data, namely Fuzzy Gustafson-Kessel Clustering. Fuzzy Gustafson-Kessel Clustering (FGK) is an example of clustering, which is the development of Fuzzy C-Means Clustering (FCM) [5]. In fact, there have been studies on crime that were conducted before, such as research conducted by Hapsari and Widodo in 2017 [6] and research conducted by Suriani in 2020 [7], both of which used the K-Means clustering method. However, in this study, FGK is more suitable because the method is more flexible in following data where Indonesia is a country with different ethnic and racial diversity in each province. Fuzzy Gustafson-Kessel Clustering can group data where the existence of data in a group is determined by its membership by converting the distance calculation into an adaptive distance norm function which is always in each iteration using a fuzzy covariance matrix, unlike FCM, which assumes the geometric shape of the cluster is a perfect sphere. The FGK algorithm uses the Mahalanobis distance function better to fit the geometric shapes in the data set.

So, from the description above, we want to know the grouping of the factors that affect the crime rate in Indonesia using a comparison of two methods, namely the Fuzzy C-Means Clustering (FCM) method and the Fuzzy Gustafson Kessel Clustering (FGK) method. So, this study can consider policymakers in making decisions in tackling Indonesia's crime rate.

## 2. RESEARCH METHODS

### 2.1. Data

The data used in this study is secondary data in the form of variables forming factors that affect the crime rate in Indonesia, where the data obtained comes from the website of the Central Statistics Agency (BPS), namely <https://www.bps.go.id>.

This study uses descriptive analysis and grouping using the Fuzzy C-Means Clustering (FCM) and Fuzzy Gustafson Kessel Clustering (FGK) methods using the optimum cluster determinant validity index, namely the Davies Bouldin Index.

### 2.2. Cluster Validation with Davies Bouldin Index

The Davies Bouldin Index is used to find the results of the clustering algorithm which was first introduced by David L. Davies and Donald W. in 1979 [8]. The Davies Bouldin Index is one of the methods used to measure the validity or the most optimal number of clusters in a grouping method where cohesion is defined as the sum of the proximity of the data to the cluster center point of the cluster being followed [9]. The Davies Bouldin Index is the ratio of the sum of the intra-cluster and inter-cluster distances [10]. A good cluster is one that has a minimum Davies Boudin Index value. To calculate the Davies Bouldin Index, the following equation is used

$$DBI = \frac{1}{k} \sum_{i=1}^k R_i \quad (1)$$

with

$$R_i = \max R_{ij}; i = 1, \dots, k; i \neq j \quad (2)$$

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (3)$$

where  $SSW_i$  is sum square within in cluster  $i$ ,  $SSW_j$  is sum square within in cluster  $j$ , and  $SSB_{ij}$  is sum square between inter cluster.

### 2.3. Cluster Analysis

Cluster analysis is the process of grouping a set of data objects into several clusters so that objects in a cluster have high similarity but are very different from objects in other clusters [11]. The characteristics of a good cluster are as follows: 1) homogeneity (within a cluster) is a high degree of similarity between members in the same cluster; and 2) heterogeneity (between clusters) is the difference in height between one cluster and another [12]. The actual quality of the cluster results depends on the method used. The clustering method must also be able to measure its own ability to find hidden patterns in the data being studied. There are many methods for measuring the similarity value between objects being compared [13]. There are many types of cluster analysis, including hierarchical clustering, neural network-based clustering, kernel-based clustering, and sequential data clustering [14].

Before conducting a cluster analysis, the following prerequisite assumptions must be met [15]:

1. The samples taken are representative of the existing population. To determine if the observation data is suitable for analysis, the Kaiser Meyer Olkin (KMO) test can be applied. If the obtained value is 0.5, then the data adequately represents the existing population.
2. Multicollinearity. Multicollinearity refers to the correlation between independent variables. It is preferable if there is none, but if there is, the data must first undergo Principal Component Analysis (PCA) to eliminate multicollinearity.

### 2.4. Principal Component Analysis

Principal Component Analysis (PCA) method or principal component analysis is a statistical technique that transforms the majority of variables that were previously correlated into a new set of smaller and independent variables. Thus, PCA is advantageous for reducing data to facilitate interpretation. Consider variables that consist of objects. The variables are then transformed into principal components (with  $k < p$ ),

a linear combination of these variables. The  $k$  main components can replace the variables that make up the variable without losing a substantial amount of information. Principal component analysis is typically an intermediate analysis, which means that the top component results can be used for additional analysis [16].

Based on the eigenvalue, the number of component factors can be determined, and the eigenvalue indicates the element's contribution to the variance of all initial variables. According to [17], the larger eigenvalue of 1 can be used to determine how many factors will be formed. Moreover, values less than one will be excluded from the analysis. In addition, the cumulative percentage of variance can be used to determine the number of factors formed; factor extraction is terminated when the cumulative percentage of variance reaches at least 60% or 75% of all variants of the original variable.

## 2.5. Fuzzy C-Means Clustering

Jim Bezdek introduced Fuzzy C-Means Clustering (FCM) for the first time in 1981. FCM is a method of grouping in which the presence of each data in each group is determined by a different value or degree of membership [18]. In a fuzzy concept, the membership of an object or data is not specified with a value of 1 indicating membership in a cluster or a value of 0 indicating non-membership, but rather with a degree of membership between 0 and 1. The FCM algorithm is depicted below.

$$P_t(X; U, V) = \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^W (d_{ik})^2 \quad (4)$$

with

$$d_{ik} = \sum_{j=1}^m (X_{ij} - V_{kj})^2 \quad (5)$$

where  $X$  is the clustering data,  $U$  is the initial partition matrix generated by random numbers, and  $V$  is the cluster center matrix.

The following are the steps for Fuzzy C-Means Algorithm Clustering.

Step 1. Input clustering data.

Step 2. Determine the number of clusters to be formed ( $c \geq 2$ ), the weighting ( $W > 1$ ), the maximum iteration (MaxIter), the expected smallest error ( $\varepsilon > 0$ ), and the initial objective function equal to 0 ( $P_0 = 0$ ), as well as the initial iteration ( $t = 1$ ).

Step 3. Generating random numbers to form the initial partition matrix  $U$  with the following equation

$$U = \begin{bmatrix} \mu_{11}(x_1) & \dots & \mu_{1n}(x_n) \\ \dots & \dots & \dots \\ \mu_{c1}(x_1) & \dots & \mu_{cn}(x_n) \end{bmatrix} \quad (6)$$

Step 4. Calculate the cluster center ( $V_{kj}$ ), with  $k = 1, 2, \dots, c$  dan  $j = 1, 2, \dots, m$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^W X_{ij})}{\sum_{i=1}^n (\mu_{ik})^W} \quad (7)$$

Step 5. Calculate Objective function

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^W \right) \quad (8)$$

Step 6. Improve each degree of membership of each data in each cluster or improve the partition matrix with the following equation

$$\mu_{ik} = \left[ \frac{\sum_{j=1}^m (X_{ij} - V_{kj})^2}{\sum_{k=1}^c \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]} \right]^{-\frac{1}{(W-1)}}. \quad (9)$$

Step 7. Specifies the termination criteria, if  $(|P_t - P_{t-1}| \leq \varepsilon$  or  $t > \text{MaxIter}$ ) then it is dismissed. However, if not then the iteration is increased  $t = t + 1$  and repeats to step 4.

## 2.6. Fuzzy Gustafson Kessel Clustering

Fuzzy Gustafson-Kessel (FGK) Clustering is a method for grouping data in which the existence of data within a cluster is determined by its membership value. This is accomplished by modifying the distance calculation function into an adaptive distance function or adaptive distance norm, which is continuously updated using a fuzzy covariance matrix [19]. FGK uses the Mahalanobis distance function to adjust the geometric shape of a data set more precisely than FCM [5], which assumes that the geometric shape of a cluster is perfectly round. The Gustafson-Kessel Clustering Process consists of the following steps:

$$J(X; U, V, A) = \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^W (d_{ikA_i})^2. \quad (10)$$

The distance of mahalanobis can be calculated by the following equation

$$d_{ikA_i} = (X_{ij} - V_{kj})^T A_i (X_{ij} - V_{kj}). \quad (11)$$

where  $X$  is data to be clustered,  $U$  is initial partition matrix by generating random numbers, and  $V$  is cluster center matrix.

Following are the steps for using the Fuzzy Gustafson-Kessel Algorithm to cluster data:

Step 1. Data inputs for clustering.

Step 2. Determine the number of clusters to be formed ( $c \geq 2$ ), weighting ( $W > 1$ ), maximum iteration (MaxIter), (MaxIter), slightest error expected ( $\varepsilon > 0$ ), and initial objective function equal to 0 ( $P_0 = 0$ ), and initial iteration ( $t = 1$ ).

Step 3. Generating random numbers to form the initial partition matrix  $U$  with the following equation

$$U = \begin{bmatrix} \mu_{11}(x_1) & \dots & \mu_{1n}(x_n) \\ \dots & \dots & \dots \\ \mu_{c1}(x_1) & \dots & \mu_{cn}(x_n) \end{bmatrix}. \quad (12)$$

Step 4. Calculate the cluster center ( $V_{kj}$ ), with  $k = 1, 2, \dots, c$  dan  $j = 1, 2, \dots, m$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^W X_{ij})}{\sum_{i=1}^n (\mu_{ik})^W}. \quad (13)$$

Step 5. Calculate the covariance of the grouping matrix ( $F_i$ ) using the following formula

$$F_i = \frac{\sum_{i=1}^n (\mu_{ik})^W (X_{ij} - V_{kj})(X_{ij} - V_{kj})^T}{\sum_{i=1}^n (\mu_{ik})^W}. \quad (14)$$

Step 6. Calculate distance using Mahalanobis distance

$$d_{ikA_i} = (X_{ij} - V_{kj})^T \left[ \rho_i \det(F_i)^{\frac{1}{n}} F_i^{-1} \right] (X_{ij} - V_{kj}). \quad (15)$$

Step 7. Calculate objective function using **Equation (10)**.

Step 8. Improve each degree of membership of each data in each cluster or improve the partition matrix with the following equation

$$\mu_{ik} = \left[ \sum_{j=1}^k \left( \frac{d_{ikA_i}}{d_{jkA_i}} \right)^{\frac{2}{(W-1)}} \right]^{-1} \quad (16)$$

Step 9. Specifies the termination criteria, if  $(|P_t - P_{t-1}| \leq \varepsilon$  or  $t > \text{MaxIter}$ ) then it is dismissed. However, if not then the iteration is increased  $t = t + 1$  and repeats to step 4.

## 2.7. Cluster Performance Evaluation

Measuring the performance of cluster results is helpful for knowing the goodness of the clusters that have been obtained. Comparison of two or more clusters to determine the best cluster can be seen through the value of the standard deviation ratio within the cluster ( $S_w$ ) and the standard deviation between clusters ( $S_B$ ) values can be used to compare two or more clusters and identify the best cluster [20]. The formula for the standard deviation in the cluster ( $S_w$ ) is as follows:

$$S_w = K^{-1} \sum_{k=1}^K S_k \quad (17)$$

where  $S_w$  is average standard deviation in the cluster,  $S_k$  is standard deviation of the  $k$ -cluster, and  $K$  is number of groups formed.

The value ( $S_B$ ) is searched using the following equation:

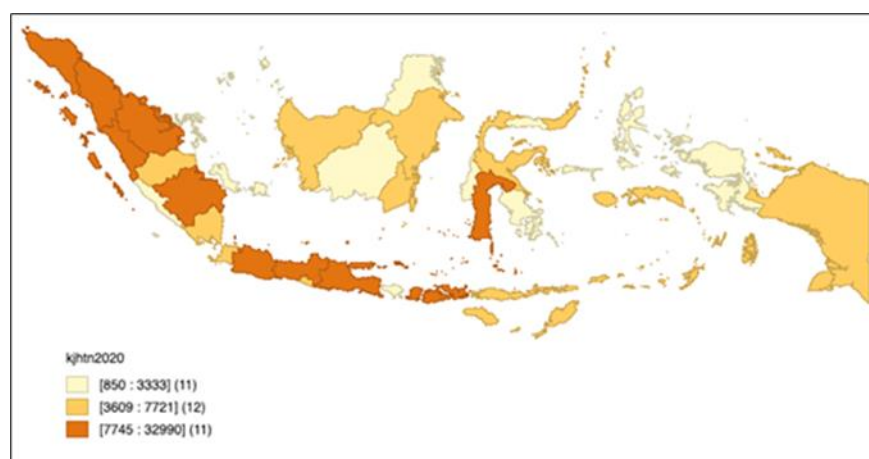
$$S_B = \left[ (K - 1) \sum_{k=1}^K (\bar{x}_k - \bar{x})^2 \right]^{\frac{1}{2}} \quad (18)$$

where  $S_B$  is standard deviation between clusters,  $\bar{x}_k$  is mean of the  $k$ -cluster, and  $\bar{x}$  is mean of the whole clusters.

## 3. RESULTS AND DISCUSSION

### 3.1. Descriptive Analysis

Before further analysis, a description of the data will be carried out to find an overview of the factors that influence the crime rate in Indonesia in 2020.



**Figure 1.** Distribution of crime incidents by province



**Figure 1** demonstrates that the islands of Sumatra and Java have the highest concentration of crime events. Java and Sumatra have Indonesia's greatest populations, demonstrating that the more the population, the higher the region's crime rate. According to the aforementioned findings, 32990 occurrences of crime were reported in North Sumatra Province in 2020, which accounted for the majority of the crime incidents in Indonesia [21]. The poverty factor and a high level of social inequality are two of the key causes of the high crime rate in North Sumatra Province. The next province with the highest number of crimes was Jakarta Province, with 26585 cases, followed by East Java Province with 17642 cases, and so on until North Maluku Province, with 850 cases, had the lowest number of crimes. According to the findings of the 2021 criminal statistics release book report, Indonesia's rate of crime tends to decline, going from 269324 cases in 2019 to 247218 cases in 2020.

### 3.2. Cluster Assumption

This study utilizes a sample of characteristics that influence crime rates in each region of Indonesia. Consequently, the premise of a representative sample is met.

Before undertaking cluster analysis, it is required to conduct multicollinearity tests to determine whether or not there is a significant link between one variable and another. Using Bartlett's test to assess multicollinearity based on the following hypothesis.

$H_0$  : There is no multicollinearity in the data

$H_1$  : There is multicollinearity in the data

The statistical results of the Bartlett's test are presented in **Table 1**.

**Table 1. Statistical results of Barlett's test**

Bartlett's Test	$\chi^2$ Test	P-Values
	51.69	1.302e-07

Existing data reject  $H_0$  with a 95% confidence interval, hence it can be stated that there is multicollinearity in the data. In this instance, multicollinearity can be avoided by reducing the factor on the analyzed variables using Principal Component Analysis (PCA).

Next, use a PCA analysis to eliminate the data's multicollinearity symptoms. According to [17], the eigenvalue larger than 1 indicates the number of factors that will be produced. While values less than one will be omitted from the analysis. In addition, while evaluating the number of components created, it is possible to determine from the background where 60% or 75% of all variations originate. **Table 2** is the outcome of the obtained eigenvalues and background differences.

**Table 2. Eigenvalues**

Factor	Eigenvalues		
	Total	% of Variance	Cumulative (%)
1	2.51	50.2	50.2
2	1.22	24.3	74.5
3	0.55	10.9	85.4
4	0.51	10.3	95.7
5	0.21	4.3	100

**Table 2** demonstrates that both factor 1 and factor 2 have several eigenvalues. Thus, two factors are generated via the main component analysis. Factor 1 accounts for 50.2% of the variation and factor 2 contributes 24.3% of the variance; therefore, the total variance that the two factors can explain is 74.5%, which is greater than 60%, indicating that the two factors were able to describe the data.

Then, new data consisting of two principal components are obtained and examined for multicollinearity. According to the results derived from the newly received data, there is no multicollinearity.

### 3.3. Validation of the Number of Clusters

To determine the number of clusters to be employed, the following cluster validation test is conducted which is represented in **Table 3**.

**Table 3. Validation test**

Number of Cluster	Davies Bouldin	
	FCM	FGK
2	0.823	1.085
3	0.795	1.099
4	0.904	1.312
5	0.715	0.981

Based on **Table 3**, the Davies Bouldin validation test is conducted using the ideal number of clusters, five. This can be demonstrated by comparing the smaller value in other clusters to the Davies Bouldin value.

### 3.4. Clusters Results

**Table 4** displays the results of five clusters obtained using the FCM approach.

**Table 4. Cluster result of FCM**

Cluster	Member	Number of Members
1	Aceh, Kalimantan Tengah, Papua, Sulawesi Barat, dan Sumatera Selatan.	5
2	Jawa Barat, Kalimantan Selatan, Kalimantan Utara, Kep. Bangka Belitung, Kep. Riau, Lampung, Maluku Utara, Riau, dan Sulawesi Selatan.	9
3	Banten, DI Yogyakarta, DKI Jakarta, Jawa Tengah, Kalimantan Barat, Nusa Tenggara Barat, Sulawesi Tenggara, dan Sumatera Barat.	8
4	Bali, Bengkulu, Gorontalo, Jambi, Kalimantan Timur, Maluku, Sulawesi Tengah, dan Sulawesi Utara.	8
5	Jawa Timur, Nusa Tenggara Timur, Papua Barat, dan Sumatera Utara.	4

**Table 4** represents the categorization of Indonesian provinces using FCM with 5 clusters. Cluster 1 comprises five provinces, Cluster 2 comprises nine provinces, Cluster 3 comprises eight provinces, Cluster 4 comprises eight provinces, and Cluster 5 comprises four provinces.

**Table 5** displays the cluster results produced using five clusters and the FGK approach.

**Table 5. Cluster result of FGK**

Cluster	Member	Number of Members
1	Aceh, Kalimantan Tengah, Maluku, Papua, Sulawesi Utara, dan Sumatera Selatan.	6
2	Jawa Timur, Nusa Tenggara Timur, Papua Barat, dan Sumatera Utara.	4
3	Bali, Banten, Bengkulu, Gorontalo, Jambi, Jawa Barat, Kalimantan Timur, Kep. Riau, Riau, Sulawesi Selatan, dan Sulawesi Tengah.	11
4	Kalimantan Selatan, Kalimantan Utara, Kep. Bangka Belitung, Lampung, dan Maluku Utara.	5
5	DI Yogyakarta, DKI Jakarta, Jawa Tengah, Kalimantan Barat, Nusa Tenggara Barat, Sulawesi Barat, Sulawesi Tenggara, dan Sumatera Barat.	8

**Table 5** is the outcome of categorizing Indonesian provinces using FGK and five clusters. Cluster 1 comprises six provinces, cluster 2 comprises four provinces, cluster 3 comprises eleven provinces, cluster 4 comprises five provinces, and cluster 5 comprises eight provinces.



### 3.5. Cluster Performance Evaluation

The following phase consists of identifying the optimal cluster results based on the standard deviation values within and between clusters.

**Table 6. Cluster performance evaluation**

Method	Number of Cluster	$S_W$	$S_B$	Ratio
FCM	5	0.87	0.14	6.21
FGK	5	0.83	0.15	5.53

The smallest  $S_W$  value, the greatest  $S_B$  value, and the smallest ratio value produce the best cluster outcomes. The grouping of provinces in Indonesia based on factors that influence crime rates uses the FGK method, which divides the data into 5 clusters with the smallest  $S_W$  value of 0.83, the largest  $S_B$  value of 0.15, and the smallest ratio value of 5.53; thus, the FGK method is superior to the FCM method for clustering.

### 3.6. Cluster Profiling

The findings of cluster profiling for five clusters using FGK are presented in **Table 7**.

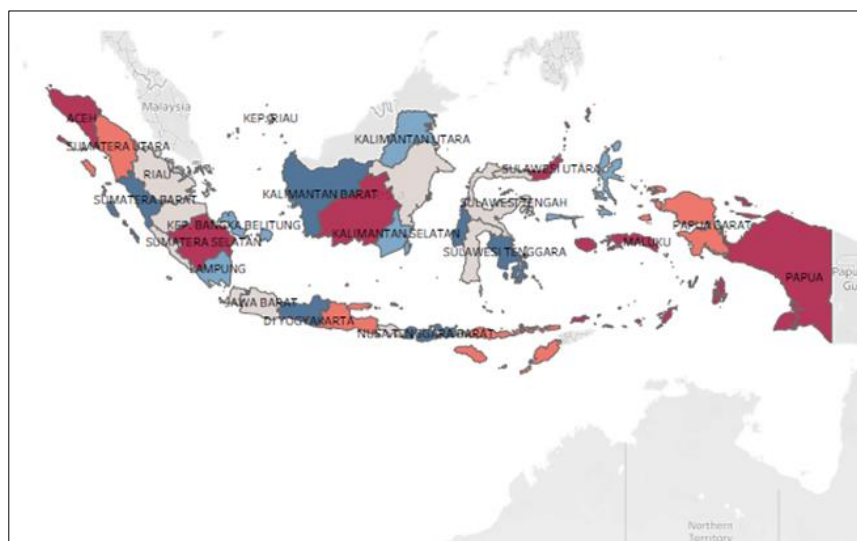
**Table 7. Cluster profiling**

Cluster	Education	Unemployment	Poor Population	PDRB	Crime
1	61.92	5.98	14.37	1.22	6962.5
2	61.52	5.96	15.88	5.23	14646
3	65.93	6.71	9.11	3.19	5598
4	62.33	4.96	7.37	0.95	3319.2
5	67.67	5.85	10.07	3.98	8663.88

According to **Table 7**, the features that distinguish each cluster among those that have been generated are as follows:

- As a result of Cluster 1, the open unemployment rate and the poor are high, and there is a moderately high crime rate, a low level of educational attainment, and GRDP.
- Cluster 2 has a very high incidence of crime, GRDP, and poverty, a very low level of educational attainment, but a comparatively low prevalence of open unemployment.
- Cluster 3 features a very high unemployment rate, a high level of educational attainment, a somewhat high GRDP, a poor population, and a low crime rate.
- Cluster 4 is a group with a very high degree of educational attainment and very low unemployment, poverty, GRDP, and crime rates.
- Cluster 5 has a very high level of educational attainment, a high GRDP and crime rate, a very large proportion of the poor, but a low unemployment rate.

Following is a map depicting the outcomes of cluster analysis using the optimal approach, FGK.



**Figure 2. Results of grouping using FGK**

**Figure 2** is a grouping of the FGK method's results in the form of a map with five colors. Members of cluster 1 are colored red, members of cluster 2 are colored orange, members of cluster 3 are colored gray, members of cluster 4 are young blue, and members of cluster 5 are colored dark blue.

#### 4. CONCLUSIONS

By comparing the value of the standard deviation within and across clusters, the grouping of provinces in Indonesia based on characteristics that affect the crime rate using the FGK technique with the FCM yielded the conclusion that the FGK method is the most effective cluster approach. Using the FGK approach, cluster 1 comprises six provinces, cluster 2 comprises four provinces, cluster 3 comprises eleven provinces, cluster 4 comprises five provinces and Cluster 5 comprises eight provinces.

#### REFERENCES

- [1] A. Kurniawan, "Pengertian Hedonisme beserta Penyebab dan Dampaknya," 2020. <https://www.merdeka.com/jabar/pengertian-hedonisme-beserta-penyebab-dan-dampaknya-klm.html>.
- [2] I. E. Alamsyah, "Jerat Hedonisme di Kalangan Masyarakat," 2019. <https://republika.co.id/berita/q0abgr349/jerat-hedonisme-di-kalangan-masyarakat>.
- [3] F. A. Putri and N. Rochmawati, "Penerapan Algoritma Gustafson-Kessel untuk Clustering Tweets Mention Akun Go-jek dan Grab Indonesia," *J. Informatics Comput. Sci.*, vol. 2, no. 3, pp. 204–212, 2021.
- [4] R. Khairani and Y. Ariesa, "Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Kriminalitas Sumatera Utara (Pendekatan Ekonomi)," *Kaji. Ekon. dan Kebijak. Publik*, vol. 4, no. 2, pp. 99–110, 2019.
- [5] R. K. Putri, B. Warsito, and M. Mustafid, "Implementasi Algoritma Modified Gustafson-Kessel untuk Clustering Tweets pada Akun Twitter Lazada Indonesia," *J. Gaussian*, vol. 8, no. 3, pp. 285–295, 2019.
- [6] D. P. Hapsari and E. Widodo, "Pengelompokan Daerah Rawan Kriminalitas di Indonesia Menggunakan Analisis K-Means Clustering," in *Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika dan Nilai Islami)*, 2017, pp. 147–153.
- [7] L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," *J. Sist. Komput. dan Inform.*, vol. 1, no. 2, pp. 151–157, 2020.
- [8] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 2A, 1979, doi: <https://doi.org/10.1109/TPAMI.1979.4766909>.
- [9] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *Evolusi J. Sains dan Manaj.*, vol. 9, no. 1, pp. 95–100, 2021.
- [10] S. M. Guthikonda, *Kohonen Self-Organizing Maps*. Wittenberg University, 2005.
- [11] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques Third Edition*. United States of America: ELSEVIER, 2012.
- [12] A. Mahmudan, "Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering," *J. Mat. Stat. dan Komputasi*, vol. 17, no. 1, pp. 1–13, 2020.
- [13] B. Wira, A. E. Budianto, and A. S. Wiguna, "Implementasi Metode K-Medoids Clustering untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 di Universitas Kanjuruhan Malang," *RAINSTEK J. Terap. Sains Teknol.*, vol. 1, no. 3, pp. 53–68, 2019.
- [14] R. Xu and D. C. Wunsch, *Clustering*. New Jersey: John Wiley & Sons, Inc., 2009.
- [15] U. Putriana, Y. Setyawan, and Noeryanti, "Metode Cluster Analysis untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Variabel yang Mempengaruhi Kemiskinan pada Tahun 2013," *J. Stat. Ind. dan Komputasi*, vol. 1, no. 1, pp. 38–52, 2016.
- [16] M. S. Delsen, A. Z. Wattimena, and S. Saputri, "Penggunaan Metode Analisis Komponen Utama untuk Mereduksi Faktor-Faktor Inflasi di Kota Ambon," *Barekeng J. Ilmu Mat. dan Terap.*, vol. 11, no. 2, pp. 109–118, 2017.
- [17] Hardisman, *Analisis Faktor, Analisis Jalur, dan Structural Equation Modelling dengan JASP*. Yogyakarta: Bintang Pustaka Mardani, 2021.
- [18] E. T. Luthfi, "Fuzzy C-Means untuk Clustering Data (Studi Kasus: Data Performance Mengajar Dosen)," 2007.
- [19] W. C. Donald and E. Gustafson, "Fuzzy Clustering with a Fuzzy Covariance Matrix," 1978.
- [20] S. B. Purnamasari, H. Yasin, and T. Wuryandari, "Pemilihan Cluster Optimum pada Fuzzy C-Means (Studi Kasus: Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia)," *J. Gaussian*, vol. 3, no. 3, pp. 491–498, 2014.
- [21] S. Purnomo, "Tingkat Kriminalitas Sumut (Masih) Tinggi," 2019. <https://analisadaily.com/berita/arsip/2019/2/19/695959/tingkat-kriminalitas-sumut-masih-tinggi/>.