**Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)**
Vol. 7, No. 2, September 2024, pp. 467 – 475
p-ISSN: 2614-3372 | e-ISSN: 2614-6150 

# Analysis of Social Vulnerability in Java Island using K-Medoids Algorithm with Variation of Distance Measurements (Euclidean, Manhattan, Minkowski)

**[1,2]*Indah Manfaati Nur, [1]Abdurakhman**
[1]Department of Mathematics, Universitas Gadjah Mada, Indonesia
[2]Department of Statistics, Universitas Muhammadiyah Semarang, Indonesia
Email:[1] indahmanfaatinur@mail.ugm.ac.id

**Article Info**
***Article history:***
Received May 26th, 2024
Revised Jun 29th, 2024
Accepted Jul 19th, 2024

***Keywords:***
DBI
Distance
Medoids
Silhouette
Social Vulnerability

**ABSTRACT**
The Social Vulnerability Index (SoVI) measurement to assess social vulnerability is only able to describe conditions in general, without being able to show which factors dominate the score. Therefore, the aim of this research is to fill this gap by applying a correlational approach with a *clustering* method to characterize the dominant factors of social vulnerability at the district level in Java and surrounding areas. The *clustering* method used in this study is the K-Medoids algorithm. This method is more powerful when there are *outliers* in the dataset used. In this study, we considered the use of 3 different distance methods, namely Euclidean distance, Manhattan distance, and Minkowski distance. As a result, the K-Medoids algorithm using Manhattan distance provides the best value based on the Davies Bouldin Index. This research found that social vulnerability exists in every region of Java Island and its surroundings.


***Corresponding Author:***
Indah Manfaati Nur,
Department of Mathematics,
Gadjah Mada University, Yogyakarta, Indonesia
Email: indahmanfaatinur@mail.ugm.ac.id

**DOI: http://dx.doi.org/10.24014/ijaidm.v7i2.31111**

## 1. INTRODUCTION

Recently, a series of natural disasters have hit Indonesia. The National Disaster Management Agency (BNPB) noted that from the beginning of 2022 to November 1, 2022, there have been 3,045 incidents and the majority of these incidents occurred on the island of Java. In addition to these events, not long ago, at the end of November 2022, there was a moderate magnitude 5.6 earthquake in Cianjur Regency which caused enormous damage and losses, both casualties and material losses. Then in early December, another moderate earthquake of magnitude 5.8 and 6.2 occurred in Sukabumi Regency. In addition to earthquakes, in early December 2022 there was also an eruption of Mount Semeru, in Lumajang Regency, East Java [1].

One of these disasters is caused by the geographical location of Java Island, which is located in the Ring of Fire. The term refers to the designation of an area that has a geographical location on the path of volcanoes, so it has a high tendency to experience eruptions and earthquakes due to volcanic activity. In addition, Java Island also has a subduction zone between the earth's tectonic plates, precisely in the southern part of Java Island to the east and west. The consequences that can be caused include earthquakes, tsunamis, volcanic eruptions, and very strong landslides. Based on the results of research conducted by several Indonesian geologists [2], the collision between these plates threatens to trigger a megathrust earthquake with a magnitude of 8.8 in the south of Java Island and has the potential to cause a tsunami as high as 20 meters. Another threat also comes from extreme weather. Indonesia's position in the tropical climate zone means that extreme weather is often followed by very high rainfall, causing flash floods and landslides.

With all the potential threats of natural disasters, it is fitting that the central, provincial, and district / city governments throughout Java Island have good preparedness. Grouping districts/cities based on social vulnerability indicators is needed as an effort to prepare appropriate mitigation to minimize the negative impact of a disaster [3]. Vulnerability is the potential loss incurred in the event of a natural disaster [4]. Then, defines vulnerability as the nature and characteristics of a group of individuals to anticipate the consequences caused by natural disasters [5]. Meanwhile, vulnerability is viewed from how much negative impact a disaster has on people and objects [6]. Thus, if summarized from several experts that have been mentioned, social vulnerability can be interpreted as a set of social capital owned by an area to be able to deal with adverse possibilities due to natural disasters, by minimizing the negative impacts that can be caused.

Measuring the level of social vulnerability has previously been carried out by several experts, including by measuring the level of environmental vulnerability to disasters [4], then [7] by formulating the Social Vulnerability Index (SoVI) based on socio-economic and demographic variables using the principal component analysis (PCA) technique. In Indonesia, SoVI measurements were conducted [8] to determine the level of local assessment of tsunami preparedness in Padang. Later, SoVI across Indonesia was also measured based on the drivers and implications of the policies implemented [9].

However, measurement using this index has several weaknesses, including not explaining the geographical aspects of disaster events [10]. In addition, the index only stops at assessing the vulnerability of an area without providing an in-depth explanation of the impact of natural disasters. Measuring the impact of vulnerability should be able to reveal the reasons why an area may be affected by a particular disaster [11]. The vulnerability level uses the PCA technique and then groups the areas based on the score using the Hierarchical *Clustering* technique [11].

The initial exploration of social vulnerability in Java Island showed the presence of *outliers in* some variables. Therefore, we used cluster analysis with the K-Medoids algorithm which is more robust to *outlier* data [12]. K-Medoids is a partition *clustering* method that combines n objects into k clusters or in another sense K-Medoids is a partition *clustering* method that aims to find a set of k-clusters among the data that best characterizes the objects in the data set. The medoid of each cluster becomes the object of the K-Medoids algorithm. Any objects adjacent to the cluster center will be merged to build a new cluster [13]. Research conducted by Marlina et al, [14] explains that K-Medoids is more suitable for grouping data than K-Means, it can be seen from the results of his research, namely getting a validity of 0.5009 while the validity of K-Means is 0.1143, this shows that grouping data on the distribution of disabled children is better using the K-Medoids method compared to K-Means. The K-Medoids algorithm has the advantage of overcoming the weaknesses in the K-Means algorithm which is sensitive to noise and outliers [15]. The K-Medoids algorithm is better than K-Means because K-Medoids is one of the efficient clustering methods for handling small datasets [16].

In addition, there are several studies that state that K-Medoids can also be applied to various research objects. Research conducted by Sangga (2018) [17], applied K-Medoids for clustering livestock commodities in Central Java province. Then, Sinatrya et al. (2018) [18] conducted research that focused on clustering large amounts of student data by applying the K-Medoids algorithm. Furthermore, research conducted by Damanik et. al. (2019) [19] suggested that the K-Medoids algorithm can be used to cluster villages that have school facilities with good results. Furthermore, Kurmiati et al [20], used the K-Medoids method to analyse the clustering of earthquake-prone areas in Indonesia. The results of this study are earthquake-prone areas based on the time of occurrence and predetermined reference points, namely in the Banda Sea, Southem Molluca Sea, Taulud Island, Halmahera, Minahassa Peninsula, Irian Jaya, Philiphine, Savu Sea, Tanimbar Islands, Sumba, Java, South of Java, Flores with strengths ranging from 3 SR to 5.7 SR. Based on this research, it can be concluded that the K-Medoids algorithm is a fairly flexible algorithm to be applied in various purposes.

Therefore, in this study we want to apply the K-medoids method to social vulnerability data on the island of Java. In this study, we also propose to implement distance measures using Euclidean, Manhattan and Minkowski methods in the K-Medoids algorithm. This is done so that we can get the best clustering results on social vulnerability data in Java. We also consider using the Davies Bouldin Index evaluation to evaluate the clustering results of the three distance measures.

This research is organized as follows. The following section summarizes the motivation for this research, and the second section describes the research method used in this research which includes an explanation of the scope of the research and the steps in conducting the research. The objectives of this research are given in the third section, which consists of the process of finding the best *clustering* method, including the evaluation of each cluster. In addition, the third section also provides many explanations regarding each of the clusters formed and their policy implications. The last section will summarize all the findings of this research.

## 2. RESEARCH METHODOLOGY

In this section we divide it into two parts, first we explain the scope of the research and second the research procedure.

### 2.1. Scope of Research

The scope of this research is to apply the K-Medoids algorithm to cluster districts/cities in Java based on social vulnerability indicators by applying 3 distance measurements, namely Euclidean, Manhattan, and

Minkowski. The variables used in this study were adapted from the *Data in Brief* article entitled "Revisiting Social Vulnerability Analysis in Indonesia Data" [21], with some additions to the updates and data availability. In addition to socioeconomic and demographic indicators, the authors also added the unemployment rate, as a macroeconomic indicator to categorize municipalities in Java and its surroundings. The source of the updated data is the BPS website, www.bps.go.id. The data is generated by Indonesia's routine surveys, the March 2021 National Socio-Economic Survey (SUSENAS) and the August 2021 National Labor Force Survey (SAKERNAS). For reasons of updated data availability, this study only uses 10 variables. The study area covers all districts in Java and surrounding areas based on 2021 data. The variables used in this study are described in Table 1.

**Table 1.** Social vulnerability variables

| Variables | Name | Description | Source |
|---|---|---|---|
| Children | $X_1$ | Percentage of population under 15 years old | SUSENAS March 2021, BPS |
| Parents | $X_2$ | Percentage of population up to 64 years | SUSENAS March 2021, BPS |
| Low education | $X_3$ | Percentage of population aged 15 years and above with low education | SUSENAS March 2021, BPS |
| Illiteracy | $X_4$ | Percentage of the population that cannot read and write | SUSENAS March 2021, BPS |
| Rent | $X_5$ | Percentage of households renting a house | SUSENAS March 2021, BPS |
| Unemployment | $X_6$ | Percentage of the total labor force that is unemployed but actively seeking work and willing to work | SAKERNAS August 2021, BPS |
| Water access | $X_7$ | Percentage of households that do not have access to safe drinking water | SUSENAS March 2021, BPS |
| Poverty | $X_8$ | Percentage of poor population | SUSENAS March 2021, BPS |
| Population growth | $X_9$ | Percentage change in population | BPS |

## 2.2. Research Procedure

K-Medoids algorithm is a variation of the K-Means method. The difference between this method compared to K-Means is that it uses medoids rather than based on the average of each cluster object [22]. K-Medoids is used to reduce the sensitivity of partitions to *outlier* values in the data set [23]. The purpose of K-Medoids *Clustering* is to overcome the shortcomings of K-Means *Clustering in* terms of its sensitivity to *outlier* data that can affect the data distribution [24]. K-Medoids *Clustering* is a partition *clustering* method that combines n objects into k-clusters. The objects in a group of objects representing a cluster are called medoids [25]. The clustering procedure using the K-Medoids algorithm is [26]:

1. Determine the number of clusters ask with theoretical and conceptual considerations.
2. Determine the distance of each object to the nearest cluster. We consider comparing 3 distance measures, including Euclidean distance, Manhattan distance, and Minkowski distance.
   a. Euclidean distance
   Euclidean distance is one of the distance calculation methods used to measure the distance of 2 points in Euclidean space (covering two-dimensional, three-dimensional, or even more Euclidean planes). To measure the level of data similarity with the Euclidean distance formula using the following formula [27]:

$$d_{(x,y)} = |x - y| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

where:
$d$ : distance between x and y data
$x$ : cluster center data
$y$ : data on the attribute
$i$ : each data
$n$ : amount of data
$x_i$ : data at the i-th cluster center
$y_i$ : i-th data

b. Manhattan distance
Manhattan distance is used to calculate the absolute difference between the coordinates of a pair of objects. The formula used is as follows:

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

where:
$d$ : distance between x and y data
$x$ : cluster center data
$y$ : data on the attribute
$i$ : each data
$n$ : amount of data
$x_i$ : data at the i-th cluster center
$y_i$ : i-th data

c. Minkowski distance
Minkowski distance is a metric in vector space where the norm is defined (normed vector space) also considered as a generalization of Euclidean distance and Manhattan distance. In measuring the distance of objects using Minkowski distance, the value of *p is* usually 1 or 2. The following formula is used to calculate the distance in this method.

$$d_{(x,y)} = \left( \sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (3)$$

where:
$d$ : distance between x and y data
$x$ : cluster center data
$y$ : data on the attribute
$i$ : each data
$n$ : amount of data
$x_i$ : data at the i-th cluster center
$y_i$ : i-th data
$p$ : power

3. It then randomly determines a new cluster center to serve as a non-medoid candidate for each object.
4. Calculate the distance between each object in each cluster to the non-medoid candidate.
5. Calculate the deviation amount (S) between the new distance amount - the old distance amount. If S < 0, then replace the objects with non-medoids cluster data to form a new set of k objects called medoids.
6. Next, repeat steps 3 - 5 until the medoid does not change. Then the cluster and each of its cluster members are obtained.

To determine the optimal number of clusters, we consider using the silhouette and elbow methods. The silhouette coefficient value can be obtained through calculating the average distance of an object. While the elbow method is obtained based on the calculation in the sum of squares. In addition, we also consider using the *Davies Bouldin Index* to evaluate the clustering results.

## 3. RESULTS AND DISCUSSION

In this section, we will divide the discussion into several sub-sections. In sub-section 1 we will discuss about *outlier* checking on the data we use. Next, in sub-section 2 we will check the assumptions on *clustering* (i.e. KMO test and multicollinearity with VIF). The determination of the number of *clusters* will be explained in sub-section 3, while the results of our proposed *clustering* comparison will be explained in sub-section 4. Finally, in sub-section 5, we will interpret the results of the best *clustering*.

### 3.1. Outliers Detection

In this section, we will check for *outliers* in the data used, where the K-Medoids method is a method that is more immune if there are *outlier* data. We use the boxplot to detect *outliers* shown in Figure 1. When there are points or data that come out of the boxplot, then the data is *outlier* data. It can be seen in Figure 1 that there are 5 variables that have *outlier* data, namely $X$ , $X$ , $X_{1257}$ , $X$, and $X_8$ . This is because in these variables there is power that comes out of the boxplot. Therefore, the K-Medoids method can be used on these data.

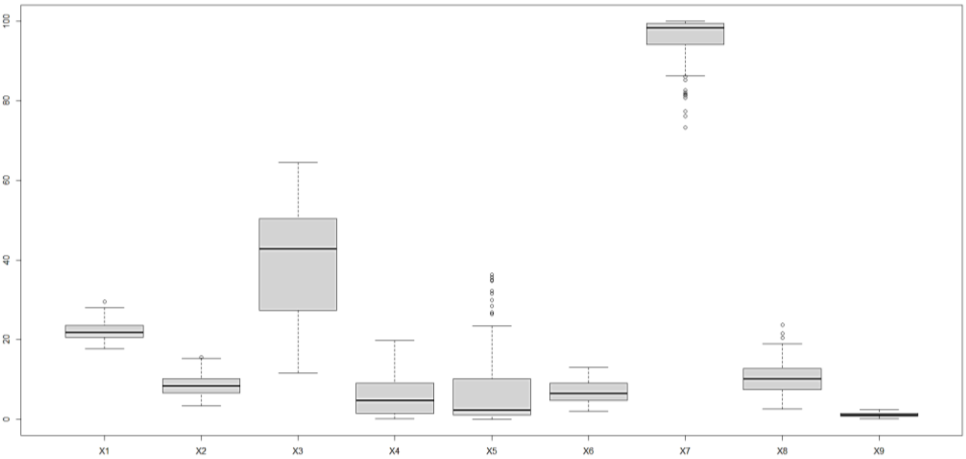



**Figure 1.** *Outlier* detection using boxplot

### 3.2. Cluster Analysis Assumptions

This sub-section will discuss the assumptions of cluster analysis. The assumptions of cluster analysis that must be met are *representative* sample (KMO test) and multicollinearity test (VIF test). A sample is said to be a *representative* sample if the KMO value is between 0.5 - 1 [28]. Meanwhile, if the VIF value is <10, then there is no multicollinearity. The KMO and VIF test results can be seen in Table 2.

**Table 2.** VIF and KMO Test

| Variables | VIF Test | KMO Test |
|---|---|---|
| $X_1$ | 6,29 | |
| $X_2$ | 5,14 | |
| $X_3$ | 5,86 | |
| $X_4$ | 2,93 | |
| $X_5$ | 3,77 | 0,73 |
| $X_6$ | 2,64 | |
| $X_7$ | 1,91 | |
| $X_8$ | 2,53 | |
| $X_9$ | 1,99 | |

Based on Table 2, it is obtained that all variables used do not have multicollinearity because the VIF value is <10. This means that the variables used do not have a high correlation between the variables. Meanwhile, the KMO value obtained is 0.73, meaning that the sample or data used is representative. Therefore, it can be concluded that the data on social vulnerability can be used for *clustering* analysis.

### 3.3. Determination of Number of Clusters

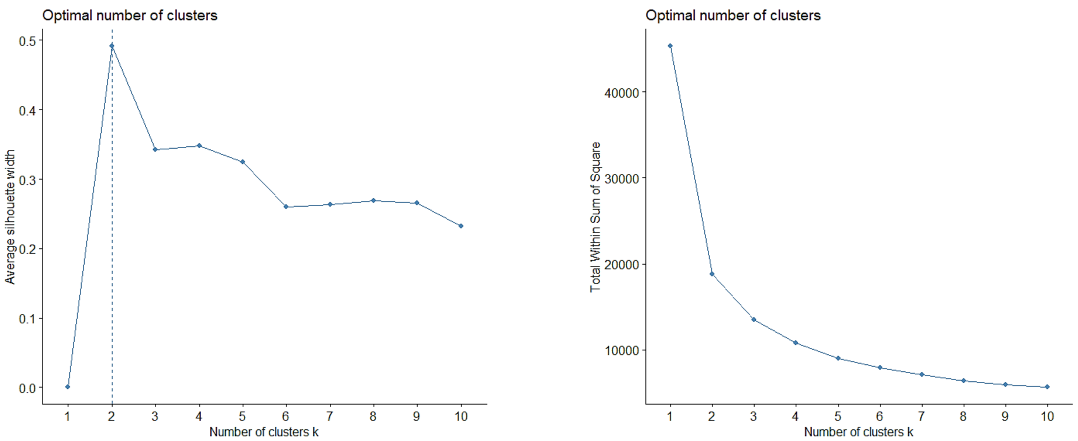
We use two methods to obtain the optimum number of clusters, namely the silhouette and elbow methods. The silhouette method determines the number of clusters by using an average value approach to predict the quality of the clusters formed [29]. When the average value is higher, the better the cluster formed. Meanwhile, the elbow method uses the *within sum square* value approach. The larger the cluster value, the smaller the within sum *square* value. The optimum number of clusters is obtained when the within sum square value decreases the most. The calculation results of the silhouette and elbow methods can be seen in Table 3, while the visualization of the 2 methods can be seen in Figure 2.

**Table 3.** Determination of the number of clusters using the silhouette and elbow methods

| Cluster | Silhouette | Elbow |
|---|---|---|
| 1 | 0,00 | 45263,92 |
| **2** | **0,49** | **18780,92** |
| 3 | 0,34 | 13462,66 |
| 4 | 0,34 | 10769,42 |
| 5 | 0,32 | 8997,80 |
| 6 | 0,25 | 7956,98 |
| 7 | 0,26 | 7110,30 |
| 8 | 0,27 | 6380,48 |

| Cluster | Silhouette | Elbow |
|---|---|---|
| 9 | 0,26 | 5935,35 |
| 10 | 0,23 | 5643,69 |

Based on Table 3, it is found that using the silhouette method, the highest average is when the number of clusters (k) = 2, which is 0.49. Meanwhile, the elbow method also produces the same conclusion, namely the optimum number of clusters obtained is when k = 2. This is because the *within sum square* value that has decreased the most is from k = 1 to k = 2. It can be concluded that in social vulnerability data, the number of clusters used in the K-Modoids algorithm is 2 clusters.




(a) (b)

**Figure 2.** The graph of determining the number of clusters, where (a) silhouette method and (b) elbow method.

### 3.4. Comparison of Cluster Results

We perform cluster analysis using the K-Medoids algorithm using 5 different distance calculations, namely Euclidean distance, Manhattan, Minkowski with p = 1, Minkowski with p = 2, and also Minkowski with p = 3. We use the *Davies Bouldin Index* (DBI) to evaluate the *clustering* results of the K-Medoids algorithm with these 5 distance calculations. Cluster results are said to be good if they have a minimum DBI value. The DBI value of each cluster result is shown in Table 4.

**Table 4.** Evaluation of cluster results using *Davies Bouldin Index*

| Distance | DBI |
|---|---|
| Euclidean | 0,81 |
| Manhattan | 0,79 |
| Minkowski (p=1) | 0,79 |
| Minkowski (p=2) | 0,81 |
| Minkowski (p=3) | 0,82 |

Based on Table 4, the *Davies Bouldin Index* evaluation results for *clustering* using the K-Medoids algorithm using 5 different distance calculation methods are obtained. It can be seen that the Euclidean distance with Minkowski when p = 2 has the same DBI value. This is because the Minkowski distance calculation formula with p = 2 (see equation 3) will produce the same formula as the Euclidean distance. Likewise, when the Minkowski distance with p = 1 will produce the same formula as the Manhattan distance, so these two distances produce the same DBI value. Therefore, based on the DBI value, the best *clustering* result for social vulnerability data in Java with the K-Medoids algorithm is when using the Manhattan or Minkowski distance with p=1 because it produces the smallest DBI value.

### 3.5. Interpretation of Best Cluster Results

The best cluster results from social vulnerability data on the island of Java using the K-Modoids algorithm is to use the Manhattan distance with the number of clusters generated is 2. The results of clustering districts / cities on the island of Java can be seen in Figure 3. While the characteristics of each cluster are presented in Table 5.

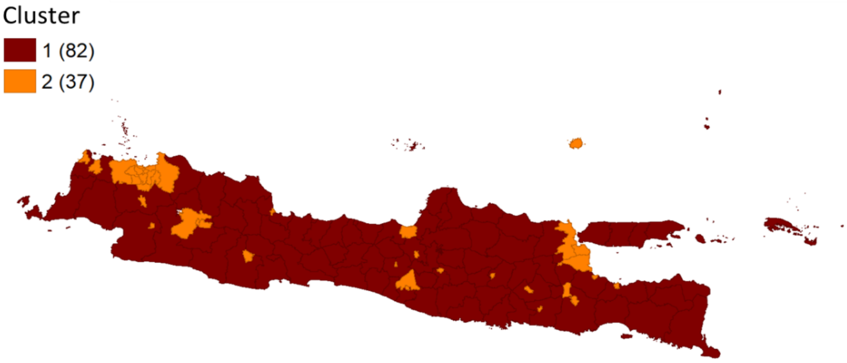



**Figure 3.** Visualization of cluster results using K-Medoids algorithm with Manhattan distance.

**Table 5.** Average characteristics of each cluster

| Variables | Cluster 1 | Cluster 2 |
|---|---|---|
| $X_1$ | 22,2 | 22,2 |
| $X_2$ | 9,57 | 6,40 |
| $X_3$ | 47,0 | 21,2 |
| $X_4$ | 7,57 | 1,58 |
| $X_5$ | 2,09 | 18,4 |
| $X_6$ | 5,98 | 8,81 |
| $X_7$ | 94,5 | 98,3 |
| $X_8$ | 12,1 | 6,38 |
| $X_9$ | 1,14 | 1,04 |

Cluster 1 consists of 82 districts/municipalities in Java. This cluster comprises almost 69% of the districts/municipalities in Java. This cluster is the most vulnerable to the consequences of natural disasters, in terms of education, demographics, economy and health. On the education side, this cluster has an average percentage of population with low education and a high illiteracy rate. Then, on the demographic side, cluster 1 shows a high dependency ratio, which is reflected in the high percentage of elderly people (65 years and above) and people under 15 years old. In addition, this region also has another characteristic in the form of relatively high average population growth among the other clusters. High population growth implies an increase in the number of child population, which has an impact on increasing the potential for insecurity. As stated by Cutter [7] and Rufat [11], the high proportion of children and aged population contributes to an increase in the additional responsibilities that must be shouldered by the productive age in the effort to save and recover from disasters.

Furthermore, on the macroeconomic side, the average poverty rate of regions in cluster 1 is the second highest and only slightly different from the highest average poverty rate in Java and its surroundings. By interpreting poverty as limited access [9], the high poverty rate indicates that the regions in cluster 1 have high vulnerability when experiencing disasters. The poor cannot afford to buy disaster preparedness equipment or prepare emergency funds so they are more vulnerable to recovering longer when a disaster occurs. In terms of health, represented by access to proper drinking water sources, districts in cluster 1 have the lowest average household access. Proper drinking water plays an important role in supporting health. When natural disasters occur, areas with lower access require special attention so that it is not more difficult to obtain clean water due to damage to water supply infrastructure.

The government, both central and regional, should pay more attention to the areas in cluster 1. Moreover, the majority of cluster 1 areas are located on the southern coastline which is one of the subduction locations of the earth's tectonic plates that have a greater tendency to experience natural disasters, especially earthquakes and tsunami waves. Various integrated mitigation efforts can be prepared, for example by building adequate evacuation route infrastructure, activating tsunami wave early warning systems, and providing adequate education to residents about disaster preparedness.

Cluster 2 consists of urban areas in Java with the highest social vulnerability characteristics in terms of open unemployment and the proportion of rental housing. Members of cluster 2 consist of 37 districts/cities detailed in figure 3. Property prices have increased dramatically along with the increasing need for housing due to urbanization that occurs in major cities in Indonesia, making low-income people prefer to rent housing [30]. High open unemployment rates indicate strong competition for jobs. The majority of jobs in urban areas come from the industrial and service sectors that require high skills and expertise levels, followed by a high level of population density, causing the labor market to be unable to accommodate the entire available labor force, resulting in inefficiencies in labor absorption [31].

The recommended remedies for vulnerabilities in cluster 2 are pre-employment training and business capital inclusiveness. Table 5 provides information that cluster 2 members have the best level of education so it is expected that these methods can provide a more sustainable resolution to vulnerabilities. Then, vulnerability in terms of housing ownership can be supported by the government through a simple flat rental program in low-income areas as currently implemented.

## 4. CONCLUSIONS

The best *clustering* method to group districts in Java and surrounding areas is K-Medoids with Manhattan distance based on Davies Bouldin Index. Using criteria, such as Silhouette and Elbow, resulted in 2 as the optimal number of clusters. Cluster 1 consists of 82 districts/cities in Java. This cluster comprises almost 69% of the districts/municipalities in Java. Cluster 2 consists of urban areas in Java with the highest social vulnerability characteristics in terms of open unemployment and the proportion of rental housing. Members of cluster 2 consist of 37 districts/municipalities.

The problems faced by each district in Java and its surroundings are different. This is reflected in the different characteristics of the clusters formed. Policy implications should be in line with key vulnerabilities. In addition, the problems most experienced by the regions are poverty and low education. This means that the government should preserve job training and financial inclusion programs, so as to help people to avoid vulnerability. In addition, disaster education is needed to help people make good strategies in dealing with disasters.

## REFERENCES

[1] BNPB. Indonesia Disaster Data 2022. 2022 [Accessed: December, 4 2023]. https://bnpb.go.id/.
[2] DLH. Official Press Release Regarding Potential 8.8 SR Earthquake and 20 m Tsunami in the South of Java Island. 2022 [Accessed: December 4, 2023]. https://dlh.kulonprogokab.go.id/detil/483/press-release-resmi-bmkg-terkait-potensi-gempa-bumi-88-sr-dan-tsunami-20-m-di-selatan-pulau-jawa.
[3] UNISDR. Implementation of the Hyogo Framework for Action: Summary of Reports 2007-2013. *United Nations Secretariat of the International Strategy of Disaster Reduction* (UNISDR), Geneva, Switzerland. 2014 [Accessed: December 4, 2023]. https://www.preventionweb.net/files/32916_implementationofthehyogoframeworkfo.pdf
[4] Cutter SL. Vulnerability to Environmental Hazards. *Progress in Human Geography*. 1996; 20(4): 529-539. DOI: https://doi.org/10.1177/030913259602000407
[5] Wisner B, Blaikie P, Cannon T, Davis I. At Risk: Natural Hazards, People's Vulnerability, and Disaster. *Routledge*. 2023; (2). https://www.researchgate.net/publication/245532449_At_Risk_Natural_Hazards
[6] Flanagan BE, Gregory EW, Hallisey EJ, Heitgerd JL, Lewis B. A Social Vulnerability Index for Disaster Management. *Journal of Homeland Security and Emergency Management*. 2008; 8(1): 1-22, Available at: https://atsdr.cdc.gov/placeandhealth/svi/img/pdf/Flanagan_2011_SVIforDisasterManagement-508.pdf
[7] Cutter SL, Boruff BJ, Shirley WL. Social vulnerability to environmental hazards. *Social Science Quarterly*. 2003; 84(2): 242-261. DOI: https://doi.or/10.1111/1540-6237.8402002
[8] Birkmann J, Setiadi NJ, Gebert N. *Socio-economic Vulnerability at the Local Level in the Context of Tsunami Early War Evacuation Planning in the City of Padang, West Sumatra*. Proceedings of the International Conference on Tsunami Warning (ICTW). Bali, Indonesia. 2008; 1-8.
[9] Siagian TH, Purhadi P, Suhartono S, Ritonga H. Social vulnerability to natural hazards in Indonesia: driving factors and policy implications, Nat. Hazards. *Journal of the International Society for the Prevention and Mitigation of Natural Hazards*. 2014; 70(2): 1603-1617. DOI: https://doi.org/10.1007/s11069-013-0888-3
[10] Nasution BI, Kurniawan R, Siagian TH, Fudholi A. Revisiting Social Vulnerability Analysis in Indonesia: An Optimized Spatial Fuzzy *Clustering* Approach. *International Journal of Disaster Risk Reduction.* 2020; 51: 1-11. DOI: https://doi.org/10.1016/j.ijdrr.2020.101801
[11] Rufat S. Spectroscopy of Urban Vulnerability, *Annals of the Association of American Geographers.* 2014; 103(3): 505-525. DOI: https://doi.or/10.1080/00045608.2012.702485
[12] Fadlurohman A, Nur IM. Pengelompokkan Provinsi di Indonesia Berdasarkan Indikator Perumahan dan Kesehatan Lingkungan Menggunakan Metode KMedoids. *Prosiding Seminar Nasional Unimus*. 2023; 6.
[13] Aurora P, Deepali, Varshney S. Analysis of K-Means and K-Medoids Algorithm for Big Data. *Procedia Computer Science*. 2016; 78: 7-12.
[14] Marlina D, Fernando A, Ramadhan A. Impelemntasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*. 2018; 4(2): 64-71.
[15] Riyanto B. Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokkan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*. 2019; 3(1): 562–568.
[16] Rofiqi AY. Clustering Berita Olahraga Berbahasa Indonesia Menggunakan Metode K-Medoid Bersyarat. *Jurnal SimanteC*. 2017; 6(1): 25–32.
[17] Sangga, VAP. Perbandingan algoritma K-Means dan algoritma K-Medoids dalam pengelompokan komoditas peternakan di provinsi Jawa Tengah tahun 2015. *Tugas Akhir Universitas Islam Indonesia, Yogyakarta*. 2018.

[18] Sinatrya NS, Wardhani LK. Analysis of K-Means and K- Medoids's Performance Using Big Data Technology. *6th International Conference on Cyber and IT Service Management (CITSM)*. 2018; 1–5.

[19] Damanik IIP, Solikhun S, Saragih IS, Parlina I, Suhendro D, Wanto A. Algoritma K-Medoids untuk Mengelompokkan Desa yang Memiliki Fasilitas Sekolah di Indonesia. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*. 2019; 1: 520–527.

[20] Kurniati D, Fauzi MZ, Ripangi, Falegas A, Indria. Klasterisasi Daerah Rawan Gempa Bumi di Indonesia Menggunakan Algoritma K-Medoids: Clustering Of Earthquake Prone Areas In Indonesia Using K-Medoids Algorithm. *Indonesian Journal of Machine Learning and Computer Science*. 2021; 1(1): 47-57.

[21] Kurniawan R, Nasution BI, Agustina N, Yuniarto B. Revisiting social vulnerability analysis in Indonesia data. *Data in* Brief. 2022; 40: 1-7. DOI: https://doi.or/10.1016/j.dib.2021.107743

[22] Farissa RA, Mayasari R, Umaidah Y. Comparison of K-Means and K-Medoids Algorithms for Drug Data Clustering with Sillhouette Coefficient. *Journal of Applied Informatics and Computing (JAIC).* 2021; 5(2): 109-116.

[23] Vercillis C. Business Intelligence, Data Mining and Optimization for Decision making. Milan: Wiley. 2009.

[24] Han J, Kamber M. Data Mining, Concepts and Techniques. Waltham: Morgan Kauffman Publisher. 2017.

[25] Setyawati AW. Implementation of Partitioning Around Medoid (PAM) Algorithm for Clustering High Schools in Yogyakarta Based on National Exam Absorption Score. Thesis. Yogyakarta: Faculty of Science and Technology, Sanata Darma University. 2017

[26] Wira B, Budianto AE, Wiguna AS. Implementation of the K-Medoids *Clustering* Method to Find Patterns of New Student Study Program Selection in 2018 at Kanjuruhan University Malang. *Journal of Applied Science & Technology.* 2019; 1(3): 54-69.

[27] H. Anton. Elementary Linear Algebra. Seventh Edition. New Jersey: Wiley. 1993

[28] Ningrat FR, Maruddani DAI, Wuryandari T. Cluster Analysis with K-Means and Fuzzy C-Means *Clustering* Algorithms for Grouping Corporate Bond Data. *Gaussian Journal.* 2016; 5(4): 641-650.

[29] Anggara M, Sujiani H, Nasution H. Selection of Distance Measure in K-Means *Clustering* for Member Grouping at Alvaro Fitness. *Journal of Information Systems and Technology (JUSTIN).* 2016; 1(1): 1-6.

[30] Nelson KW. Choices and Opportunities: Low-Income Rental Housing in Indonesia. *Review of Urban & Regional Development Studies*. 1989; 1: 49-63. https://doi.org/10.1111/j.1467-940X.1989.tb00012.x

[31] Mankiw GN. Macroeconomics. Seventh Edition. New York: Worth Publishers. 2010.


## BIBLIOGRAPHY OF AUTHORS

Indah Manfaati Nur, currently a student in Doctoral Program of the Mathematics Department, Universitas Gadjah Mada, Yogyakarta. I have a specific interest in Statistics, Data Mining and Machine Learning.