IMPLEMENTATION OF THE DBSCAN METHOD FOR CLUSTER MAPPING OF EARTHQUAKE SPREAD LOCATION

Muhammad Bariklana [1], Achmad Fauzan[2*]

[1,2] *Statistics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia*
Kaliurang Street KM 14,5, Sleman, Yogyakarta, 55584, Indonesia

*Corresponding author's e-mail: * achmadfauzan@uii.ac.id*

ABSTRACT

West Java area is located on the Pacific Circum and Mediterranean Circum routes, and this causes West Java area to be an unstable area that is characterized by many active working volcanoes and frequent earthquakes. An analysis of the grouping of earthquake data in West Java Province area is urgently needed. The purpose of this study was to classify areas based on the density of earthquake occurrence areas in West Java using Density-Based Spatial Clustering of Application with Noise (DBSCAN). The population in this study is all earthquake events that occurred in 2021. While the sample used in this study is data on the location of the distribution of earthquakes in West Java Province in 2021 taken from the BMKG online data website at dataonline.bmkg.go.id. This research began with nearest-neighbor analysis to see patterns of data distribution. If the data distribution pattern is grouped, then DBSCAN analysis can be continued. The DBSCAN algorithm uses a combination of parameters, namely minimum points (MinPts) and epsilon (Eps). Cluster results are evaluated using the silhouette coefficient. Then, in this study, deeper data exploration was carried out in three ways, namely: (1) Clustering based on the highest silhouette value, (2) clustering by lowering the MinPts value, and (3) clustering based on the smallest upper limit (supremum) value of the silhouette coefficient. The data exploration here aimed to form more clusters while still considering the silhouette coefficient value limits so that there are more areas prone to earthquakes but also maintaining the validity of the results obtained. Next, determine the best cluster results by comparing the cluster results obtained. The best cluster results were obtained at Eps=10000 and MinPts=3, which formed 12 clusters with a silhouette coefficient value of 0.713, which means that the clusters have a strong structure. It is hoped that the information regarding the grouping of areas where earthquakes frequently occur can be used as a form of earthquake disaster mitigation and minimize the impact of losses due to the earthquake.

1. INTRODUCTION

Geographically, most areas in Indonesia are located in disaster-prone areas [1]. Natural disasters in Indonesia are caused by Indonesia's location, which is between three plate meetings, namely the Indo-Australian plate which moves north, the Eurasian plate which moves south, and the Pacific plate which moves from east to west. Indonesia is also traversed by two active mountain ranges in the world, namely the Pacific Circum and the Mediterranean Circumference. Astronomically, West Java province is located between $5°50′ − 7°50′$ South Latitude and $104°48′ − 108°48′$ East Longitude.

West Java region is in the Pacific Circum and Mediterranean Circum pathways, so the area is an unstable area that is marked by many volcanoes that are still actively working and earthquakes often occurring. Earthquakes have distinctive characteristics, namely unavoidable events, very sudden and surprising events, and the time of the earthquake; the location of its central and its strength cannot be predicted appropriately or accurately by anyone, including earthquake experts [2].

According to the Indonesian Central Statistics Agency (BPS), the frequency of earthquakes in Indonesia in 2021 occurred 10519 times [3]. The frequency of the earthquake rose dramatically by 2151 compared to the previous year. In addition, according to the Center for Volcanology and Geological Disaster Mitigation, in Indonesia, there are 127 active volcanoes [4]. Indonesia is the country with the largest number of active volcanoes in the world and is the first rank in the world with the highest number of fatalities. The island of Java is the island with the most active volcanoes in Indonesia, which are 34 active volcanoes.

Of the 34 active volcanoes on Java Island, 15 active volcanoes are found in West Java Province. According to the National Disaster Management Agency, from all-natural disaster events, there were 5402 total disasters in Indonesia in 2021 [5]. Of all natural disasters in Indonesia in 2021, West Java Province is the province with the highest number of natural disasters, namely 1358 events. Therefore, West Java Province is a province with several volcanoes and a high number of natural disasters.

Due to many earthquakes in West Java Province, one of the disaster mitigation efforts is by determining the cluster in the area. One of the suitable clustering methods for spatial data is the *Density-Based Spatial Clustering of Application with Noise* (DBSCAN). DBSCAN is an algorithm that is included in the category of density-based clustering, which is the process of forming the cluster based on the density of distance between objects in the dataset. DBScan has the advantage of being able to detect noise [6]. Harini *et al.* [7] applied DBSCAN methods for disaster, namely by mapping earthquake data in Bali and Nusa Tenggara. Anwar *et al.* [8] applied this method for disaster, namely by mapping the risk of forest fires to identify areas at risk of forest fires.

Several studies related to earthquake clustering include the Reviantika [9] study, which used the K-Means algorithm in earthquake analysis on the island of Java which produced the two best clusters, and Wahyu [10] used the same method and object and produced 4 clusters. Kurmiati [11] used the K-Medoid algorithm to classify earthquake areas in Indonesia, resulting in 3 cluster areas. Febriani [12] conducted a cluster analysis using the Kohonen Self-Organizing Maps (SOMs) algorithm in earthquake clustering in Indonesia, resulting in 3 clusters from the analysis. Sari [13] conducted clustering using K-Affinity Propagation (K-AP) in earthquake clustering in Indonesia, which resulted in 4 clusters. Meanwhile, Arista [14] uses the Fuzzy C-Means algorithm for grouping earthquake events into three groups. However, the clustering algorithm above does not use the spatial concept (in this case, area density). In contrast, the DBSCAN algorithm is a cluster analysis that looks at it from a spatial density point of view. The density calculation uses the location/ Universal Transverse Mercator (UTM) coordinate system so that it is easier to identify which areas are frequently subject to earthquakes and their densities.

This study aims to classify earthquake data in West Java Province based on density. The data used in this study is earthquake data in West Java Province in 2021 taken from the BMKG's online data website at *dataonline.bmkg.go.id*. From the data obtained, the latitude and longitude variables are used for analysis using an algorithm [15]. Furthermore, this research is expected to be used as a form of disaster mitigation to minimize the impact of losses due to earthquakes in West Java Province.

2. RESEARCH METHODS

2.1 Data Sources and Research Stages

The population in this study is all earthquake events that occurred in 2021. While the sample used in this study is data on the location of the distribution of earthquakes in West Java Province in 2021 taken from the BMKG online data website at dataonline.bmkg.go.id. **Table 1** is an illustration of data in research.

Table 1. Earthquake Data for West Java Province in 2021

Date	Time	Latitude (°)	Longitude (°)	Depth (Km)	Magnitude (Rs)
1/5/2021	23:42	-7.47	106.71362	36.2	3.15
1/7/2021	7:19	-7.14	107.01774	81.1	2.74
⋮	⋮	⋮	⋮	⋮	⋮
12/24/2021	0:36	-7.72	107.51556	11.6	3.14

The latitude and longitude variables were used in the DBSCAN algorithm, while the depth and magnitude variables were used to see the earthquake characteristics of each group. The stages of the research conducted are described in **Figure 1**.
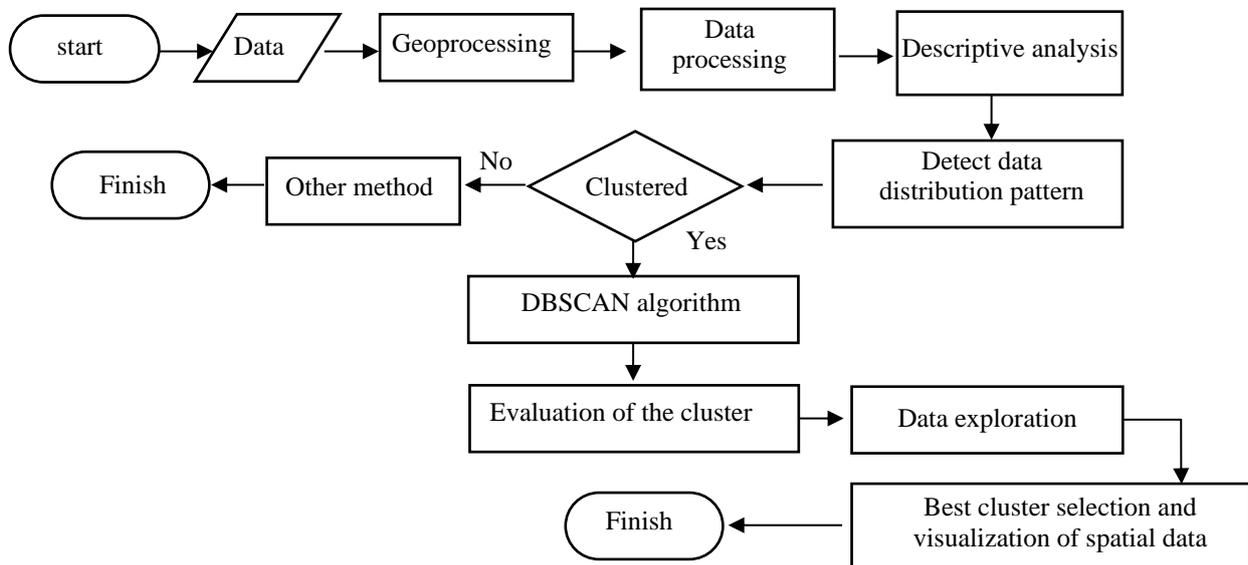
Figure 1. Research Flowchart

The research began by entering data on the location of the distribution of earthquakes in West Java Province in 2021. Then, the second stage was carried out by geoprocessing, georeferencing, and digitization in making maps of the West Java Province area. The third stage was data preprocessing, namely by eliminating duplicate data and data that was outside the territory of West Java Province. The fourth stage was descriptive analysis to see an overview of earthquake data in West Java Province in 2021 and detect patterns of data distribution using nearest-neighbor analysis.

The results of the fourth stage were used as the basis for seeing the results of the data distribution patterns, whether they are grouped or not. If the data distribution pattern is clustered, then it is continued to the DBSCAN algorithm, while if the data distribution pattern is not clustered, then another method is used. The fifth stage was to perform cluster analysis using the DBSCAN algorithm, followed by evaluating the cluster results with the silhouette coefficient. The next stage was deeper data exploration in three ways: (1) clustering based on the highest silhouette value, (2) clustering by lowering the MinPts value, and (3) clustering based on the supremum value of the silhouette coefficient. The last step was to choose the best cluster by comparing the results of the clusters from the optimum silhouette with the results from data exploration and visualization of the best cluster results.

2.2 Nearest-Neighbor Analysis

The nearest-neighbor analysis is a method designed to measure patterns in a data point in two or three dimensions. This method involves calculating the average point distance between all points and their nearest neighbors [16]. The nearest-neighbor index is expressed as the ratio of the observed distance divided by the expected distance. The distance used is the Euclidean distance presented in **Equation (1)**. The nearest-neighbor index is expressed by R and the calculation is written in **Equation (2)** [17].

$$d(i,j) = \sqrt{\left(x_{i1} - x_{j1}\right)^2 + \left(x_{i2} - x_{j2}\right)^2 + \cdots + \left(x_{ip} - x_{jp}\right)^2} \qquad (1)$$

$$R = \frac{\bar{d}_{obs}}{\bar{d}_{ran}} \qquad (2)$$

$\bar{d}_{obs}$ is the observed mean nearest-neighbor or the average distance of observations between each point and the nearest neighbor. $\bar{d}_{ran}$ is the expected mean nearest-neighbor or the expected average distance for the given points in a random pattern. The calculation of $\bar{d}_{obs}$ and $\bar{d}_{ran}$ is presented in **Equation (3)**.

$$\bar{d}_{obs} = \frac{\sum_{i=1}^{n} d_i}{n} \text{ and } \bar{d}_{ran} = \frac{1}{2\sqrt{p}} = \frac{1}{2\sqrt{n/A}} \qquad (3)$$

d_i is the distance between i and its nearest neighbor, n is the number of points, and A is the minimum rectangular area around all points, or a given area value. $SE_{\bar{d}}$ is standard error of the mean nearest neighbor and c is the test statistics, with the calculation formula written in **Equation (4)** and **(5)**.

$$c = \frac{\bar{d}_{obs} - \bar{d}_{ran}}{SE_{\bar{d}}} \qquad (4)$$

$$SE_{\bar{d}} = \frac{0.26136}{\sqrt{np}} = \frac{0.26136}{\sqrt{n^2/A}} \qquad (5)$$

The nearest-neighbor index has values ranging from 0 to 2.15. The nearest-neighbor index with a value of 0 indicates a fully clustered pattern. Meanwhile, the nearest-neighbor index with a value of 2.15 shows a complete dispersion pattern. The random pattern is indicated by the nearest-neighbor index value of 1. The sampling distribution of test statistics c is a normal distribution. In other words, c is a normal standard deviation. Hypothesis test on test statistics c (z_{score}) with null hypothesis (H_0) is randomly distributed data and alternative hypothesis (H_1) is clustered distributed data, in making decisions the z_{tabel} value is used in normal distribution and the absolute value of z_{score}, if $z_{score} > z_{tabel}$, then reject H_0 [18].

2.3 Density-Based Spatial Clustering of Application with Noise (DBSCAN)

DBSCAN is an algorithm in which the process of forming clusters is carried out based on the level of density of distances between objects in the dataset. DBSCAN has the advantage of being able to detect noise [6]. If the number of neighbors and themselves within the Eps radius is less than MinPts and no neighbors become the core because of their presence, the data is categorized as noise [19]. In DBSCAN, 2 parameters are required, namely minimum pointsThe nearest-neighbor (MinPts) and epsilon (Eps) which values are determined by the researchers. MinPts is the minimum number of items in a cluster and Eps is the value for the distance (radius) between items which forms the basis for forming a neighborhood from an item point. Neighborhoods that lie within the radius ($\in$) are called $\in$-neighborhoods of the data object [20]. An object is a core object if the $\in$-neighborhood of the object contains at least MinPts of objects. Core objects are pillars of density areas [17].

There are 2 types of points in a cluster, namely at the core points and at the edges of the border points, where the neighborhood of the border points contains far fewer items than the neighborhood of the core points [21]. A border point may belong to more than 1 cluster. If given a set of D objects, all core objects given the parameters $\in$ and MinPts can be identified. The clustering within is reduced by using core objects and neighborhoods to form densities, where dense areas are clusters [17].

There are several terms in DBSCAN: (1) directly density-reachable: for core object q and object p, it says that p is directly density-reachable of q (with $\in$ and MinPts) if p is in the $\in$-neighborhood of q, (2) density-reachable: if p is density-reachable of q (with $\in$ and MinPts in D) if there is a chain of objects $p_1, ..., p_n$ so that $p_1 = q$, $p_n = p$, and p_{i+1} are directly density-reachable from p_i to $\in$ and MinPts, for $1 \leq i \leq n$, $p_i \in D$, (3) density-connected: two objects $p_1, p_2 \in D$ are density-connected (with $\in$ and MinPts) if

there are objects $q \in D$ such that so that p_1 and p_2 are density-reachable from q to $\in$ and MinPts. **Figure 2** is an example of density-reachability and density-connectivity for a certain $\in$ which is represented by the circle radius and by assuming MinPts=3.
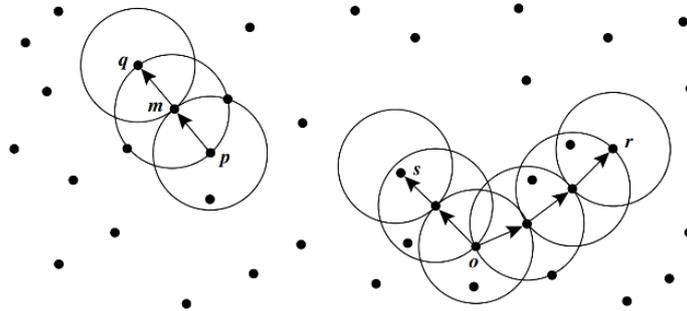
Figure 2. Density-reachability and density-connectivity (**Han** *et al.* **[17]**)

Points m, p, o, r are core objects because each one is in a $\in$-neighborhood that contains at least three points. Object q directly density-reachable from m. Object m is directly density-reachable from p and vice versa. Object q is (indirectly) density-reachable of p because q is directly density-reachable of m and m is directly density-reachable of p. However, p is not density-reachable from q because q is not a core object. Similarly, r and s are density-reachable of o and o is density-reachable of r. Therefore, o, r, and s are all density-connected.

DBSCAN finds clusters by tracing the clusters, which is by examining the $\in$-neighborhood (Eps-neighborhood) of each point in the database. If the $\in$-neighborhood of point p contains more than MinPts, a new cluster with p as the core object is created. Then, DBSCAN iteratively collects density-reachable objects directly from the core object, which may involve merging several density-reachable clusters **[20]**. The sequence of the DBSCAN algorithm is as follows: (1) choose the initial point p randomly, (2) determine Eps and MinPts to take all points that are density reachable to point p, (3) if p is the core point then a cluster is formed, (4) if p is a border point, there is no density-reachable relation of p, and DBSCAN will visit the next point in the database, (5) continue processing until all points have been processed, (6) the result obtained does not depend on the order of the processed points taken.

2.4 Model Evaluation

The quality of a cluster is obtained from the value of the silhouette coefficient **[22]**. The distance between data in the same cluster is called intra-cluster. Meanwhile, the distance between data in one cluster to data in another cluster is called inter-cluster **[23]**. If the difference between intra-cluster and inter-cluster is greater, then the resulting silhouette coefficient value is getting better. For example, if you have data that is clustered into Cluster A, Cluster B, and Cluster C. Therefore, the silhouette coefficient calculation begins by calculating the value $a(i)$ (intra-cluster), which is finding the average distance of the data i with all data in the same cluster, it is assumed here that the data i is in Cluster A **[24]**. The formula for $a(i)$ is written in **Equation (6)**, where A is the amount of data in Cluster A.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq 1} d(i,j) \tag{6}$$

Next, calculate the value of $b(i)$ (inter-cluster), which is the minimum value of the data i mean with all data in different clusters. Then, assume different clusters besides A and cluster C. Then, the calculation of the data i average distance with all data in cluster C ($d(i,C)$) is written in **Equation (7)**, where C is the amount of data in Cluster C. After calculating $d(i,C)$ for all clusters $C \neq A$, then choose the minimum distance as the value of $b(i)$ with **Equation (8)**.

$$d(i,C) = \frac{1}{|C|} \sum_{j \in C} d(i,j) \tag{7}$$

$$b(i) = \underset{C \neq A}{min} \ d(i,j) \tag{8}$$

If Cluster B has a minimum distance, then $d(i, B) = b(i)$, which is called the neighbor of the data i and is the second-best cluster for the data i after Cluster A. After $a(i)$ and $b(i)$ is known, then the next step is to calculate the silhouette value of object i with **Equation (9)**.

$$s(i) = \frac{b(i) - a(i)}{max(a(i), b(i))}$$
(9)

The value of $s(i)$ is between -1 and 1, where each value is interpreted as follows: (1) $s(i) \approx 1$, then the data i is classified properly (in A), (2) $s(i) \approx 0$, then the data i is in the middle between the two clusters well (A dan B), (3) $s(i) \approx -1$, then the data i is classified as weak (closer to Cluster B than A) [25]. The average silhouette $s(i)$ of all objects i in a cluster is called the average silhouette width of that cluster. The average silhouette $s(i)$ for $i = 1, 2, …, n$ is called the average silhouette width for the entire data set. Then, the maximum value of the average silhouette width for the entire data set is called the silhouette coefficient [26].

3.   RESULTS AND DISCUSSION

3.1.   Data Overview

In 2021, earthquakes in West Java Province occurred in 110 locations. Before grouping the data, it will be seen first how the distribution pattern of the data. As the basis for whether the points of the earthquake spread or form clusters, the nearest neighbor analysis is used. Based on the results of calculations using the nearest-neighbor analysis, the nearest-neighbor index value is 0.716320, which indicates a cluster pattern. Besides, the absolute value of the z_{score} is 5.691877 which is worth more than the z_{table}, which is 1.96. Based on this, the location data for the distribution of earthquakes in West Java Province in 2021 has a clustered data distribution pattern.

3.2.   Clustering using DBSCAN Algorithm

In this study, it was carried out on Eps and MinPts values. The MinPts interval used is from 3 to 6, while the Eps value is between 9000 and 23000. From the combination of these two parameters, the best cluster is taken based on the silhouette coefficient. The results of the combination of the two parameters are presented in **Table 2**.

Table 2. Silhouette coefficient evaluation results

No	MinPts	Eps	Silhouette coefficient	Amount of Cluster	Multiple Points on Noise	Multiple Points in Cluster
1	3	9000	0.756	8	53	57
2	3	9200	0.756	8	53	57
⋮	⋮	⋮	⋮	⋮	⋮	⋮
71	3	23000	0.282	2	13	97
72	4	9000	0.808	4	67	43
⋮	⋮	⋮	⋮	⋮	⋮	⋮
142	4	23000	0.282	2	13	97
⋮	⋮	⋮	⋮	⋮	⋮	⋮
284	6	23000	0.121	2	21	89

Based on **Table 2**, the maximum silhouette coefficient value is 0.859, namely at MinPts=5 and Eps=9000, with 3 clusters formed and 71 noise. A silhouette coefficient value of 0.859 means that the cluster has a strong structure [24]. The silhouette coefficient plot for each MinPts is shown in **Figure 3.**
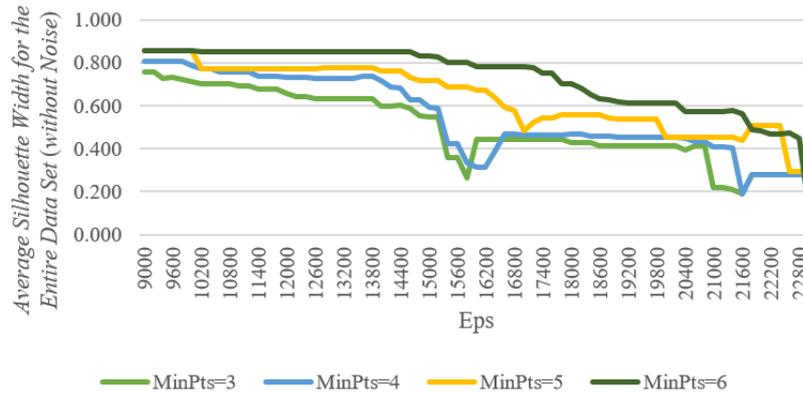
Figure 3. Average silhouette width for the entire data set every MinPts

In **Figure 3**, it is found that the silhouette coefficient is found at MinPts = 9000 and Eps = 5. In addition, slowly the value of the average silhouette width for the entire data set decreases to a MinPts value of 23000 for all Eps.

3.3. Data Exploration

In addition to clustering based on the highest silhouette value, according to the explanation at the research stage, data exploration was carried out. Exploration of seismic clustering data aims at disaster mitigation. Disaster mitigation is a study that aims to minimize the impact of losses due to disaster events, both material losses and moral losses [1]. In this disaster mitigation, what is being done is to create clusters of earthquake-prone areas. Researchers tried to explore this by lowering the MinPts value so that more and more clusters are formed so that there are also more and more areas that are prone to earthquakes.

3.3.1. Clustering by lowering the MinPts value

In this section, the MinPts value was reduced to 3 and 4 from the optimum silhouette value. The visualization of the number of clusters from the combination of the two MinPts is presented in **Figure 4 (a)**. From **Figure 4 (a)** for MinPts=3, it reaches the optimum point with a silhouette coefficient value of 0.756, 8 clusters are obtained, the number of points in the cluster is 57, and the number of points in the noise is 53. Meanwhile, in MinPts=4, it reaches the optimum with a silhouette coefficient value of 0.808 with the number of clusters formed is 4, the number of points in this cluster is 43, and the number of points in the noise is 67.

3.3.2. Clustering based on the value of the smallest (supremum) value of the silhouette coefficient

The results of clustering from the supremum value of a cluster are said to be good ($\approx$0.71), which are visualized in **Figure 4 (b)**. Based on this analysis, MinPts=3 and Eps=10000 are obtained. From the results of this second exploration, there are 12 clusters, 70 points in the cluster and 40 points in the noise.
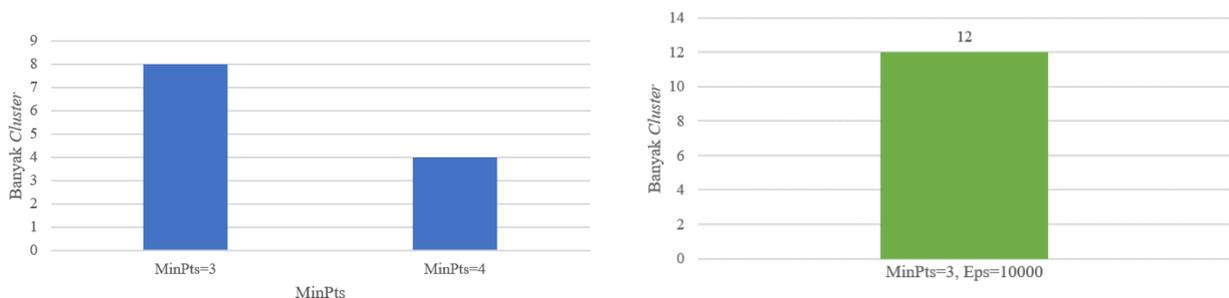
Figure 4. (a) The number of clusters by decreasing the value of MinPts

(b) The number of clusters based on the supremum value of the silhouette coefficient

3.4.   Selection of the Best Cluster Results

In selecting the best cluster, it is determined by comparing the cluster results on the optimum silhouette coefficient value with the cluster results on data exploration. The comparison of cluster results is shown in **Table 3**.

Table 3. Comparison of cluster results

Eps	MinPts	Silhouette Coefficient	Amount of Cluster	Multiple Points in Cluster	Multiple Points on Noise	Region
9000	5	0.859	3	39	71	Bandung and Garut Regencies, and South Sea of West Java (south of Sukabumi and Cianjur Regencies)
9000	3	0.756	8	57	53	Cianjur, Sukabumi, Bandung, and Garut Regencies, and South Sea of West Java (south of Sukabumi and Cianjur Regencies)
9000	4	0.808	4	43	67	Bandung and Garut Regencies, and South Sea of West Java (south of Sukabumi and Cianjur Regencies)
10000	3	0.713	12	70	40	Cianjur, Sukabumi, Karawang, Bandung, and Garut Regencies, and South Sea of West Java (south of Sukabumi, Cianjur and Garut Regencies)

From **Table 3**, it is found that the optimum silhouette obtained is 0.859 (Eps=9000, MinPts=3) with 3 clusters. Meanwhile, the clusters that have the highest number are found in the 3rd data exploration results (Eps=10000, MinPts=3, and silhouette coefficient = 0.713), that is, with a total of 12 clusters. In addition, the results of clusters that have more points in the cluster than the number of points in the noise are found in the cluster results in the 1st data exploration (Eps=9000, MinPts=3, and silhouette coefficient = 0.756) and cluster results on the 3rd data exploration (Eps=10000, MinPts=3, and silhouette coefficient = 0.713).

From these results, the researcher determined the best cluster results based on the number of clusters obtained, namely the cluster with the highest number of clusters. This is because the larger the number of clusters that are formed, the more areas that are prone to earthquakes are detected, the hope is that this can be used as information for the public to be aware of areas that are prone to earthquakes. In addition, the best cluster results are taken from clusters that have many points in the cluster that are larger than the many points in the noise. Therefore, it is determined that the best cluster results are found in the 3rd data exploration results (Eps=10000, MinPts=3, and silhouette coefficient=0.713), namely with 12 clusters, 70 points in the cluster, and 40 points in the noise. The visualization of the cluster results can be seen in **Figure 5**.
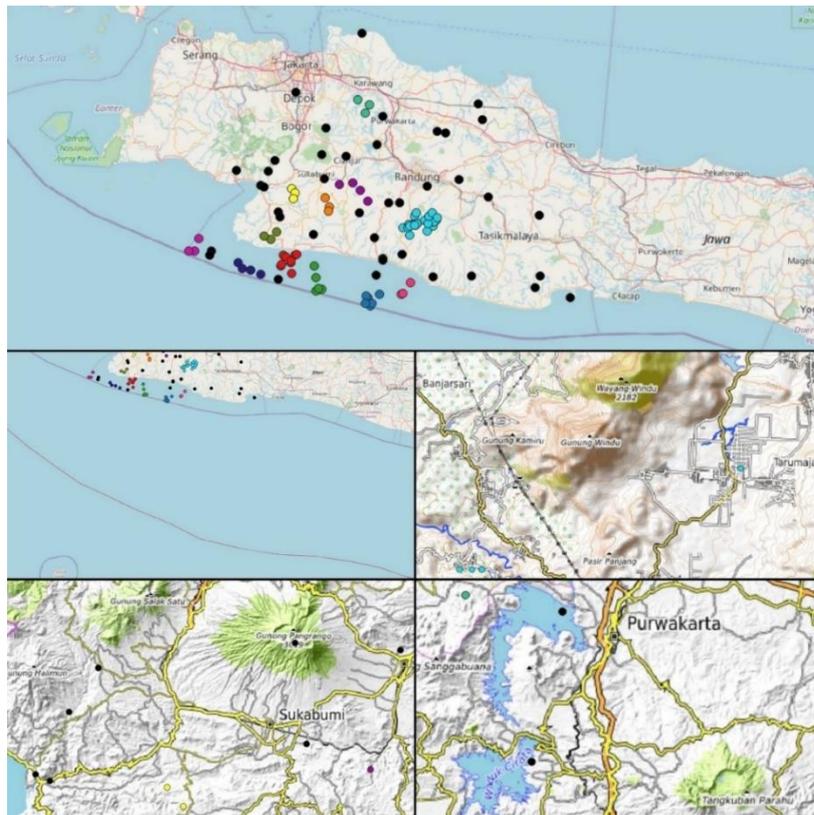
Figure 5. Visualization of best cluster results

From **Figure 5**, there are 6 clusters of earthquakes located in the South Sea area of West Java, which is close to the meeting line of the Eurasian and Indo-Australian plates (shown in the red line in the South Sea of West Java). Then, there is 1 cluster located in Bandung and Garut Regencies, which is close to Mount Wayang (inactive volcano). In addition, there are 5 clusters located in Sukabumi, Cianjur, West Bandung, and Karawang regencies, which are close to Mount Salak, Mount Gede Pangrango, and Mount Tangkuban Perahu (active volcanoes). The cluster areas in **Figure 5** are presented in **Table 4**.

Table 4. Regions in each cluster at the best cluster results

Cluster	Region	Amount of Earthquake Points
1	South Sea of West Java (South of Sukabumi Regency)	9
2	Cianjur Regency	3
3	Sukabumi Regency	3
4	Sukabumi Regency	3
5	South Sea of West Java (South of Cianjur Regency)	5
6	Karawang Regency	3
7	Bandung and Garut Regencies	24
8	South Sea of West Java (South of Cianjur Regency)	6
9	South Sea of West Java (South of Sukabumi Regency)	4
10	Cianjur and Bandung Regencies	4
11	South Sea of West Java (South of Sukabumi Regency)	3
12	South Sea of West Java (South of Garut Regency)	3

The characteristics of the depth and strength of the earthquake in the silhouette coefficient supremum for each cluster are shown in **Table 5**.

Table 5. Characteristics of earthquakes in the best cluster results

Cluster	Variable	Statistics				
		Mean	Median	Standard Deviation	Minimum	Maximum
1	Depth (Km)	42.4	34.5	25.18	27.4	108.3
	Magnitude (Rs)	3.6	3.41	0.76	2.82	4.84
2	Depth (Km)	84.13	81.1	75.7	10	161.3
	Magnitude (Rs)	3	2.9	0.32	2.74	3.35
3	Depth (Km)	49.63	10	68.65	10	128.9

Cluster	Variable	Statistics				
		Mean	Median	Standard Deviation	Minimum	Maximum
4	Magnitude (Rs)	2.61	2.41	0.51	2.24	3.19
	Depth (Km)	89.3	90.7	8.09	80.6	96.6
5	Magnitude (Rs)	3.17	3.05	0.35	2.9	3.57
	Depth (Km)	41.78	28.3	33.07	10	95.6
6	Magnitude (Rs)	2.97	2.95	0.49	2.37	3.56
	Depth (Km)	106.47	10	167.09	10	299.4
7	Magnitude (Rs)	2.89	2.93	0.09	2.79	2.95
	Depth (Km)	11.42	10	4.52	10	31.9
8	Magnitude (Rs)	2.61	2.54	0.42	1.92	3.66
	Depth (Km)	27.88	27.95	6.09	16.8	34.4
9	Magnitude (Rs)	3.17	3.16	0.5	2.44	3.88
	Depth (Km)	52.35	50.3	26.14	27.3	81.5
10	Magnitude (Rs)	3.75	3.8	0.88	2.76	4.64
	Depth (Km)	55.48	13.65	86.15	10	184.6
11	Magnitude (Rs)	2.7	2.67	0.55	2.18	3.29
	Depth (Km)	29.27	28.1	3.5	26.5	33.2
12	Magnitude (Rs)	3.58	3.66	0.25	3.3	3.77
	Depth (Km)	13.5	11.6	4.74	10	18.9
	Magnitude (Rs)	3.1	3.11	0.05	3.04	3.14

From **Table 5**, it is found that the average shallowest earthquake depth is in Cluster 7 (Bandung and Garut Regencies), while the highest average earthquake strength is in Cluster 9 (the southern sea area of West Java (south of Sukabumi Regency)). In addition, for most parts in each cluster, the standard deviation values were smaller than the average value, meaning that the data on the depth of the earthquake and the strength of the earthquake in each cluster varied less. Furthermore, the characteristics of the depth and strength of the earthquake on the noise data in the supremum silhouette are shown in **Table 6**.

Table 6. **Characteristics of earthquakes in best Cluster Results**

Noise	Variable	Statistics				
		Mean	Median	Standard Deviation	Minimum	Maximum
	Depth (Km)	83.66	50.2	92.73	10	324.2
	Magnitude (Rs)	3.1	3.09	0.51	1.89	4.65

From **Table 6**, it is found that the noise data in the silhouette coefficient supremum has varying earthquake depths because the standard deviation value is greater than the average value. Meanwhile, the strength of the earthquake varies less because the standard deviation value is smaller than the average value. Furthermore, researchers compared the best cluster results obtained with the earthquake hazard map in West Java Province in 2020 by the Regional Disaster Management Agency (BPBD) of West Java Province which is shown in **Figure 6**. In **Figure 6**, the earthquake hazard index is shown based on colors. If the color is closer to green, then it indicates a lower earthquake hazard index; if the color is closer to yellow then it indicates a medium earthquake hazard index; and if the color is closer to red then it indicates a higher earthquake hazard index.

Figure 6. **Earthquake hazard map (Source: BPBD-West Java Province [27])**

If the best cluster results obtained by researchers are compared to the earthquake hazard map from the BPBD of West Java Province in 2020, the results tend to be the same, namely, some areas are prone to earthquakes in Sukabumi, Cianjur, Bandung, Garut regencies, and some areas in the southern West Java that have the earthquake index is quite high by BNPB. In addition, there are also earthquake-prone areas in Karawang and Garut Regencies which have an earthquake index that tends to be moderate by the BPBD of West Java Province. Therefore, the best cluster results obtained by researchers with the earthquake hazard map from the BPBD of West Java Province in 2020 show the same results. Overall, West Java is an unstable area which is characterized by many volcanoes because it is located on the Pacific Circum and Mediterranean Circum pathways, so in the mainland area of West Java Province, earthquakes (volcanic earthquakes) often occur. In addition, in the South Seas area, West Java is located close to the meeting line of the Eurasian and Indo-Australian plates, so earthquakes are also common (tectonic earthquakes).

Therefore, possible disaster mitigation is to conduct outreach and increase awareness in the community. In addition, conducting basic disaster training for officials and the community is also necessary. For the government, it can supervise the implementation of various regulations regarding spatial planning, building permits (IMB), and other regulations related to disaster prevention that aim to minimize the impact of disasters. In addition, it can conduct research and study the characteristics of disasters as well as disaster risk analysis [1].

4. CONCLUSIONS

Earthquakes in West Java Province have a clustered pattern of data distribution. 12 clusters were formed, which were obtained using DBSCAN. The evaluation results obtained are 0.713 using the silhouette coefficient, which means that the cluster has a strong structure. Each cluster has its earthquake characteristics. This can be a form of disaster mitigation to minimize the impact of losses due to earthquakes in the West Java Province area and become information for the public to be aware of areas that are prone to earthquakes.
.

ACKNOWLEDGMENT

REFERENCES

[1]   D. Hermon, *Geografi Bencana Alam - Rajawali Pers*. Jakarta: PT RajaGrafindo Persada, 2015.
[2]   A. M. Nur, "Gempa Bumi, Tsunami dan Mitigasinya," *Gempa Bumi, Tsunami dan Mitigasinya*, vol. 7, no. 1, 2010.
[3]   BPS-Statistics Indonesia, *Statistical Yearbook of Indonesia in Infografis*. Jakarta: BPS-Statistics Indonesia, 2022.
[4]   PVMBG, "Tipe Gunung Api di Indonesia," *Magma Indonesia*, 2021. [Online]. Available: https://magma.esdm.go.id/v1/edukasi/tipe-gunung-api-di-indonesia-a-b-dan-c.
[5]   BNPB, "Bencana Indonesia 2021," 2022. [Online]. Available: https://gis.bnpb.go.id/infografis/bencana/tahun 2021.png.
[6]   M. T. Furqon and L. Muflikhah, "Clustering the Potential risk of Tsunami using Density-Based Spatial Clustering of Application with Noise (DBSCAN)," *J. Enviromental Eng. Sustain. Technol.*, vol. 3, no. 1, pp. 1–8, 2016.
[7]   S. Harini, H. Fahmi, A. D. Mulyanto, and M. Khudzaifah, "The earthquake events and impacts mapping in Bali and Nusa Tenggara using a clustering method," in *IOP Conference Series: Earth and Environmental Science*, 2020.
[8]   M. T. Anwar, W. Hadikurniawati, E. Winarno, and A. Supriyanto, "Wildfire Risk Map Based on DBSCAN Clustering and Cluster Density Evaluation," *Adv. Sustain. Sci. Eng. Technol.*, vol. 1, no. 1, pp. 1–7, 2019.
[9]   F. Reviantika, C. N. Harahap, and Y. Azhar, "Analisis Gempa Bumi pada Pulau Jawa Menggunakan Clustering Algoritma K-Means," *J. Din. Inform.*, vol. 9, no. 1, pp. 51–60, 2020.
[10]   A. Wahyu, and Rushendra, "Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa," *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 1, pp. 175–179, 2022.
[11]   D. Kurmiati, M. Zakiy Fauzi, Ripangi, A. Falegas, and Indria, "Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm Klasterisasi Daerah Rawan Gempa Bumi di Indonesia Menggunakan Algoritma K-Medoids," *Indones. J. Mach. Learn. Comput. Sci.* , vol. 1, no. 1, pp. 47–57, 2021.
[12]   B. S. Febriani and R. F. Hakim, "Analisis Clustering Gempa Bumi Selama Satu Bulan Terakhir dengan Menggunakan Algoritma Self-Organizing Maps (SOMs) Kohonen," in *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS*, 2015, pp. 715–722.
[13]   N. N. Sari, "K-AFFINITY PROPAGATION (K-AP) CLUSTERING UNTUK KLASIFIKASI GEMPA BUMI (Studi Kasus: Gempa Bumi di Indonesia Tahun 2017)," Universitas Islam Indonesai, 2018.

[14]   R. R. Arista, R. A. Asmara, and D. Puspitasari, "Pengelompokan Kejadian Gempa Bumi Menggunakan Fuzzy C-Means Clustering," *J. Teknol. Inform. dan Terap.*, vol. 04, no. 02, pp. 3–10, 2017.

[15]   M. Amiruzzaman, R. Rahman, M. R. Islam, and R. M. Nor, "Logical Analysis of Built-in DBSCAN Functions in Popular Data Science Programming Languages," *Mist Int. J. Sci. Technol.*, vol. 10, no. June, pp. 25–32, 2022.

[16]   D. Ebdon, *Statistics in Geography*. Blackwell Ppublishing, 1985.

[17]   J. Han, M. Kamber, and J. Pei, *Data mining: Data Mining Concepts and Techniques*. USA: Morgan Kaufman, 2014.

[18]   B. N. Boots and A. Getls, *Point Pattern Analysis*. Sage Publication, 2020.

[19]   E. Prasetyo, *Data Mining: Konsep dan Aplikasi menggunakan Matlab*. Jakarta: CV Andi Offset, 2012.

[20]   I. M. S. Putra, "Algoritma DBSCAN (Density-Based Spatial Clustering of Application with Noise) dan contoh perhitungannya.," Bali, 2018.

[21]   M. Ester, H. P. Kriege, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Satial Databases with Noise," in *KDD-96*, 1996, pp. 226–231.

[22]   P. J. Rousseeuw, "Silhouettes: A graphical Aid to the Iinterpretation and Validation of Cluster Analysis," *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, 1987.

[23]   N. Kaoungku, S. Keerachart, R. Chanklan, K. Kerdprasop, and N. Kerdprasop, "The silhouette width Criterion for Clustering and Association Mining to Select Image Features," *Int. J. Mach. Learn. Comput.*, vol. 8, no. 1, pp. 69–73, 2018.

[24]   A. Struyf, M. Hubert, and P. J. Rousseeuw, "Clustering in an Object-Oriented Environment," *J. Stat. Softw.*, vol. 1, no. 4, pp. 1–30, 1997.

[25]   R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering," *Techno.COM*, vol. 20, no. 2, pp. 186–197, 2021.

[26]   L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley Series in Probability and Statistics, 2005.

[27]   BPBD Provinsi Jawa Barat, "Peta Bahaya Gempa Bumi di Provinsi Jawa Barat," *Badan Nasional Penanggulangan Bencana Daerah Provinsi Jawa Barat*, 2022. [Online]. Available: https://bpbd.jabarprov.go.id/peta-krb-2/.