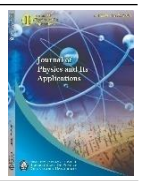
Journal of Physics and Its Applications

Journal homepage :https://ejournal2.undip.ac.id/index.php/jpa/index



Clustering of Seismicity in the Indonesian Region for the 2018-2020 Period using the DBSCAN Algorithm

Akrima Amalia[1*], Udi Harmoko[2], Gatot Yuliyanto[2]

[1]*Physics Undergraduate Study Program, Department of Physics, Diponegoro University, Semarang, Indonesia*
[2]*Department of Physics, Diponegoro University, Semarang, Indonesia*

[*)]*Corresponding author: udiharmoko@lecturer.undip.ac.id*

A R T I C L E I N F O

Article history:
Received: 29 July 2021
Accepted: 6 November 2021
Available online: 29 November 2021

Keywords:
Clustering
DBSCAN
Seismicity
Indonesia

A B S T R A C T

Indonesia is located at the confluence of 3 large, active plates that are constantly moving. Therefore, Indonesia is one of the countries that has a high level of seismicity risk. This study aims to classify seismicity data in the Indonesian region based on coordinate data which contains variable data on frequency of occurrence, depth, and strength of seismicity. Seismicity data was obtained from the BMKG official website using data for the period 2018 to 2020. The clustering technique used was the DBSCAN algorithm. This algorithm requires epsilon and MinPts input parameters. The results of the cluster formed will then be validated using silhouette coefficients. Based on the coordinate data, 4 clusters were formed with 4 disturbances. Based on the characteristic data, 3 clusters were formed with 5 disturbances. The silhouette coefficient obtained was 0.35 for coordinate data and 0.39 for characteristic data. This research is useful for increasing the use value of abundant seismicity information and can be used as an effort to mitigate seismicity natural disasters.

1. Introduction

Seismicity modeling has been done by many researchers. However, seismicity modeling is still a major and open challenge in geoscience. One of the well-known techniques used is the clustering technique [1].

The application of the clustering technique was carried out by [2] to find 3 phases of the eruption process based on seismicity clusters in volcanic areas. Another study conducted by [1] to determine clusters with irregular shapes in seismic areas. This technique is able to identify the fault section of the repeated seismicity events carried out by [3]. As well as being able to identify non-linear site responses carried out in [4].

According to [5], Indonesia's territory is located between the confluence of 3 large active plates, namely the Indo-Australian plate, the Pacific Ocean and the Eurasian plate, resulting in Indonesia having a fairly high level of seismicity even in the world.

The application of the clustering technique was also carried out by researchers using seismicity data in the Indonesian region. The study was conducted using data over a certain period of time. The study was conducted using seismicity data until 2017. There are 2019 seismicity data used but only for a few months [6-11]. Therefore, this study will use a continuation period, namely 2018 to 2020.

The algorithm that will be applied in this research is the DBSCAN (Density Based Spatial Clustering of Application with Noise) algorithm. This algorithm was chosen because based on previous research, DBSCAN is more effective than the CLARANS algorithm [12]. According to [6] its ability is also different from other algorithms such as K-Means and K-Medoids. The DBSCAN algorithm can be used to detect outliers or noise and there is no need to determine the number of clusters at the beginning. Research conducted by [8] states that this algorithm is better at determining parameters than the DMDBSCAN (Dynamic Method Density Based Spatial Clustering of Application with Noise) algorithm.

Based on this explanation, this research will offer a study to participate in implementing the clustering technique using the DBSCAN algorithm and seismicity data in the Indonesian region for the period 2018 to 2020. This study aims to cluster seismicity data based on coordinate data and seismicity characteristics data, as well as to visualize the results of clusters formed. The benefits of this research are to increase the use value of abundant seismic information, to increase efforts to mitigate seismicity natural disasters, and to be used for the development of science related to the clustering of spatial data using the DBSCAN algorithm.

2. Methods

Based on the previous explanation, the following are the research methods used in this study (Fig. 1). The research data was obtained by downloading on the official website of BMKG (Badan Meteorology and Climatology Geophysics). Seismicity data used was a type of tectonic seismicity (caused by a shift in the earth's crust or tectonic events). The data that has been obtained were then be analyzed using descriptive statistics. Preprocessing of data was carried out with data cleaning, data integration, data selection, and data transformation so that the data is ready to be used for clustering.

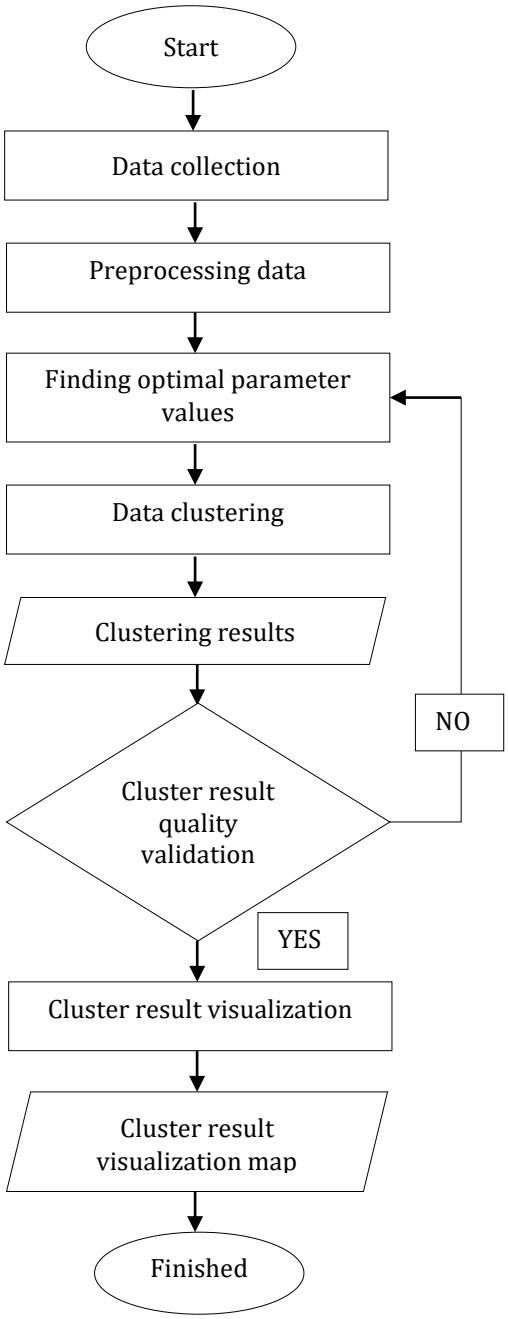


Fig.1: Research flow chart.

Data clustering was done using the DBSCAN algorithm. This algorithm requires input parameters Epsilon (maximum radius between one point to another) and MinPts (minimum point in a cluster that is formed). The input parameters are more effectively searched using the search for optimal values or parameters using the help of a k-dist graph. The clustering will show the number of clusters and the noise formed. As for determining whether or not the cluster is formed by means of validation. Validation is done using the silhouette coefficient value. The best results will then be visualized using the help of map visualization. Fig 1. shows the stages of research carried out in this study.

3. Results and Discussion

Data collection was done by downloading the results obtained as many as 30,302 data with 8 types of variables. The data were then be classified based on the seismicity incident area so as to get the results of 47 seismicity areas. The variables used were the latitude and longitude coordinates, the depth of the seismicity, the strength of the seismicity, and the frequency of the occurrence of seismicity. Figure 2 is a visualization of the results of the distribution of seismicity for the period 2018 to 2020.

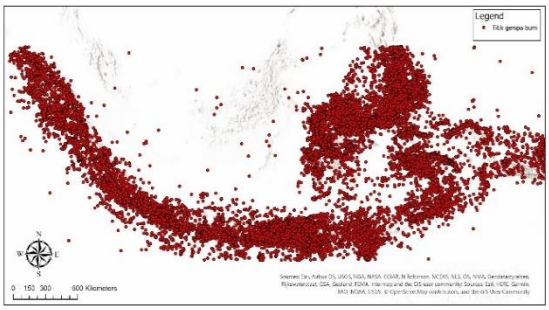


Fig. 2: Distribution of seismicitys for the period 2018-2020.

Based on the seismicity distribution map in Fig 2. which will be explained briefly in Table 1, regarding the descriptive analysis of the data to be used for research consists of the average value, median value, minimum value, and maximum value of the seismicity frequency data variable earth, the depth of the seismicity, and the strength of the seismicity.

It is shown in Table 1 that the average depth value is at a depth of 66 km which is included in the category of shallow seismicity because it is less than 70 km [13]. As for the average seismicity strength of 4.0 on the Richter scale, which if equated with the MMI scale is classified as a seismicity with a low risk level.

Table 1. Descriptive statistics.

Information	Frequency of occurrence	Depth (km)	Seismicity strength (SR)
Average	645	66	4.0
Middle value	250	39	3.8
Minimum	1	10	2.9
Maximum	3823	469	6

Data clustering is done after determining the optimal parameter values. After the seismicity data has been classified based on the area of occurrence and analyzed, parameter determination is carried out using a technique that is more effective than trial and error, namely with the help of a k-dist graph as shown in Fig 3. Determination of the best or optimal epsilon and MinPts parameter values seen from sharp changes that occur in the graph. The epsilon parameter is indicated by a sharp change, while the k input parameter indicates the MinPts input parameter.

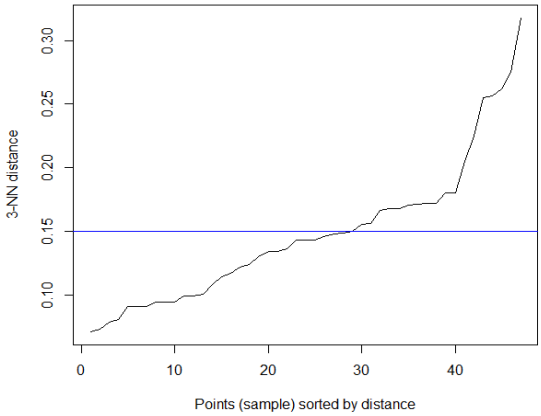


Fig. 3: K-dist graph.

Based on the help of the k-dist graph as shown in Fig 3. the determination of the optimal input parameter values for Epsilon and MinPts is obtained as in Table 2. and Table 3.Fig 3. shows that the input value of k is 3, which is seen on the y-axis (3-NN distance) while the epsilon value obtained from a sharp change is at a value of 0.15. The determination of the input value will then be assessed using the silhouette coefficient for the best validation.

Based on Table 2. the results of clustering based on seismicity coordinates obtained the best silhouette value of 0.35 with an epsilon input value of 0.15 and MinPts of 3. The results obtained show that 4 clusters are formed with 4 noise. The next table, below This is the result of clustering based on data on seismicity characteristics.

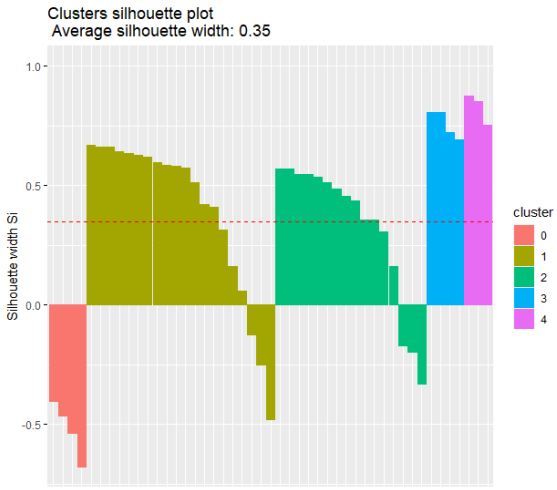
Table 2. Clustering results based on seismicity coordinate data.

Eps	MinPts	Cluster	Noise	Silhouette
0.1	1	20	0	0.31
	2	10	10	0.3
	3	5	20	0.18
	4	2	29	0.15
0.15	1	7	0	0.25
	2	5	2	0.32
	3	4	4	0.35
	4	4	8	0.27
0.2	1	3	0	0.09
	2	3	0	0.09
	3	2	2	0.09
	4	2	2	0.09

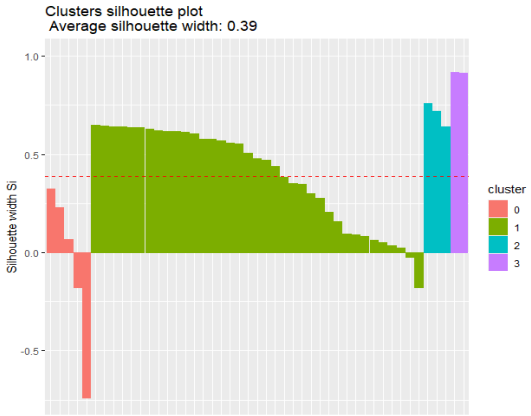
Table 3. Clustering results based on seismicity coordinate data.

Eps	MinPts	Cluster	Noise	Silhouette
0.1	1	18	0	-0.05
	2	4	14	-0.03
0.2	1	8	0	0.26
	2	3	5	0.39
0.3	1	4	0	0.19
	2	2	2	0.28

Based on Table 3. shows the best results based on the largest silhouette coefficient value with input epsilon of 0.2 and MinPts of 2. The results obtained are clusters formed of 3 with noise of 5. Based on research that has been done previously by several references contained in In the introduction section, the results or silhouette coefficient values obtained are almost the same, namely in the number past 0.2 with noise less than 10. The best validation value is obtained if the value is close to 1. However, this value is included in the category that is in line with previous research. Or it can be said to strengthen what is the result of previous research. Fig 4. shows the results of visualizing the average silhouette value.



(a)



(b)

Fig. 4: Silhouette value (a) based on seismicity coordinate data (b) based on seismicity characteristic data.

Based on Fig. 4. shows that each data or category in the cluster has a different color. This is shown for both cluster and noise. As for the average value that is used as the silhouette coefficient or validation, it is taken from the average silhouette width that has been shown at the top of the visual. The results obtained in figure (a) are the best silhouette results based on seismicity coordinate data and figure (b) shows the best silhouette results based on seismicity characteristic data. This difference in value indicates that a number close to 1 or in this case 0.39 is included in the best cluster category.

Determination of the optimal parameter values and their validation have been carried out. The next picture will show how the cluster result plot is formed in the software used. Based on the cluster formed, each cluster will be distinguished by color.

Based on Fig 5. the results of the 5 clusters formed are plotted using existing software, so that each cluster with a different color and shape comes out, where each cluster has a density with one another. On Fig. 5. shows the longitude and latitude on the coordinates of the graph because the data entered using the relationship or input coordinate data that contains latitude and longitude. Subsequent analysis was carried out for seismicity data based on the characteristics shown in Table 4.

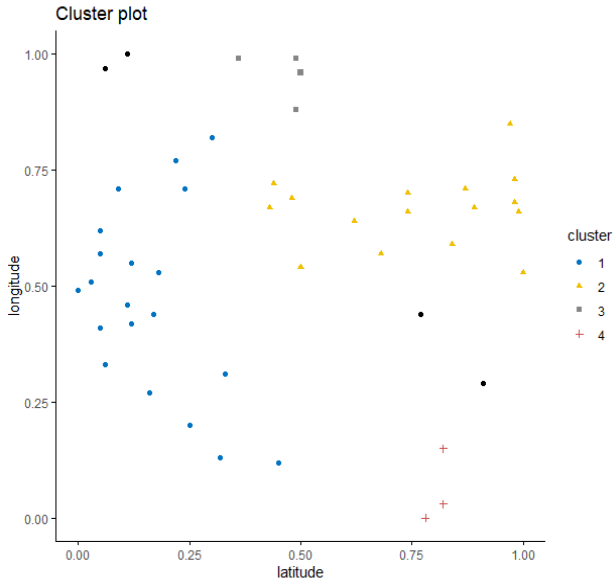


Fig. 5: Cluster result plot is formed based on seismicity coordinate data.

Based on Table 4 3 clusters were formed with 5 noise, having the most members in cluster 1 with 37 cluster members. In addition to having the most members, cluster 1 has a fairly large average seismicity strength that needs attention. However, on average, each cluster has a fairly low depth below 70 km, so it is classified as a shallow seismicity that also needs to be watched out for, including if the location of the settlement is close to the fault zone of the seismicity.

Table 4. Profiling the results of the cluster formed based on data on seismicity characteristics.

Cluster	Variable	Statistic			Range	Number of cluster members
		Average	Minimum	Maximum		
1	Frequency of occurrence	504	1	1883	1 s/d 1883	37
	Depth (km)	44	10	116	10 s/d 116	
	Strength (SR)	3,9	2,9	4,8	2,9 s/d 4,8	
2	Frequency of occurrence	3	1	4	1 s/d 4	3
	Depth (km)	24	10	51	10 s/d 51	
	Strength (SR)	5,6	5,3	6,0	5,3 s/d 6,0	
3	Frequency of occurrence	3711	3599	3823	3599 s/d 3823	2
	Depth (km)	19	13	25	13 s/d 25	
	Strength (SR)	3,2	3,1	3,2	3,1 s/d 3,2	
Noise	Frequency of occurrence	846	1	2796	1 s/d 2796	5
	Depth (km)	274	10	51	10 s/d 51	
	Strength (SR)	4,0	3,2	5,2	3,2 s/d 5,2	

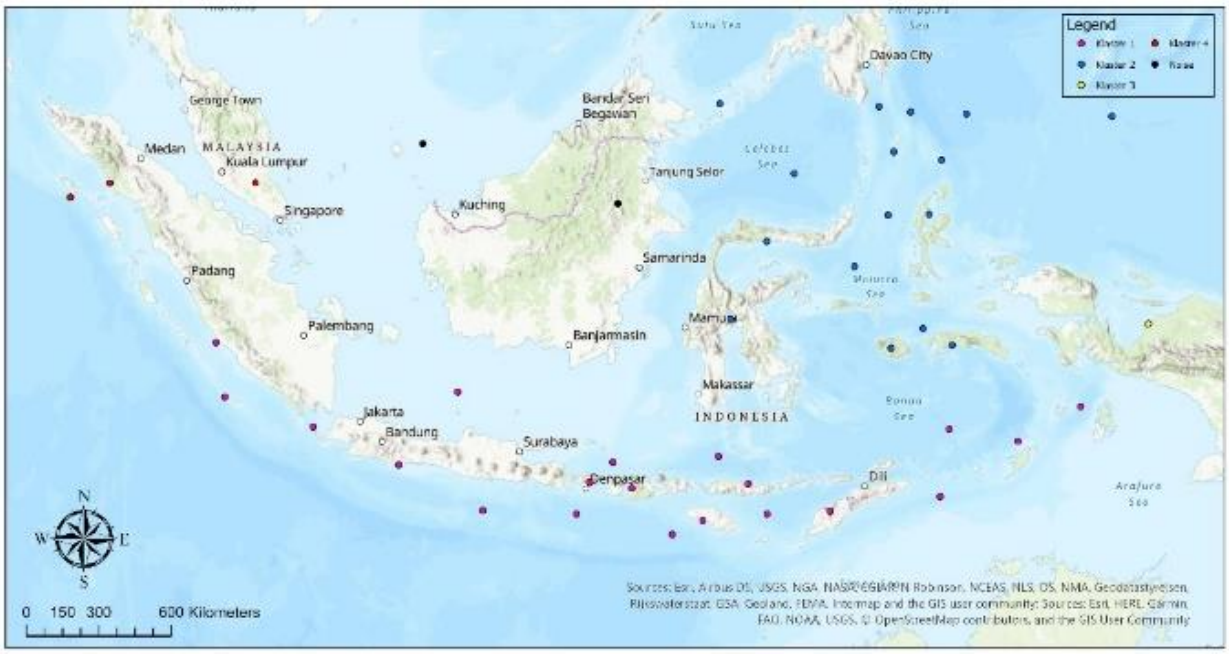


Fig. 6: Visualization of cluster results formed.

Another purpose of this research is to visually show the cluster formed. This is shown in Fig 6. The results of this visualization are shown from the results of the cluster based on the coordinate data or the latitude and longitude of the seismicity.

4. Conclusion

The conclusion from the results of this study is that the clustering of seismicitys carried out using seismicity coordinate data and seismicity characteristics got the best silhouette values of 0.35 and 0.39. Based on the results of the visualization that has been made, the distribution of seismicitys is almost evenly distributed throughout Indonesia, but the western part of Indonesia has seismicity points or the frequency of seismicity occurrences is quite low compared to the central and eastern parts of Indonesia.

5. Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

We are grateful to the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) for access to their 2018-2020 Period of the Indonesian Region data which were used in this study.

References

[1] G. Georgoulas, A. Konstantaras, E. Katsifarakis, C.D. Stylios, E. Marayelakis, and G.J. Vachtsevanos, "Seismic-mass" *Density-Based Algorithm for Spatio-Temporal Clustering, Expert* Systems *with Applications*, **40**, 4183-4189, (2013).

[2] F. Massin, V. Ferrazzini, P. Bachelery, A. Nercessian, Z. Duputel, and T. Staudacher, "Structures and Evolution of the Plumbing System of Piton de la Fournaise Volcano Inferred from Clustering of 2007 Eruption Cycle Sismicity,"J. Volcano. Geotherm. Res.,201, 96-106, (2011).

[3] P. Martinez-Garzon, Y. Ben-Zion, I. Zaliapin, and M. Bohnhoff, "Seismic Clustering in the Sea of Marmara : Implications for Monitoring Seismicity Process,"Tectonophysics, 768, 1-11, (2019).

[4] K. Ji, R. Wen, Y. Ren, and Y. P. Dhakal, "Nonlinear Seismic Site Response Classification using K-means Clustering Algorithm : Case Study of the September 6, 2018 Mw6.6 Hokkaido Iburi-Tobu Seismicity , Japan,"Soil Dynamics and Seismicitys Eng., 128, 1-14, (2020).

[5] E. W. Santoso, "Spatial Planning of Meulaboh City after the Seismicity and Tsunami of 26 December 2004 Proposed Recommendations,"Alami, 10, 2, 13-17, (2005).

[6] M. T. Furqon and L. Muflikhah, "Clustering The Potential Risk of Tsunami using Density-Based Spatial Clustering of Application with Noise (DBSCAN),"*J. Env. Eng. & Sustain. Tech.,***3**, 1, 1-8, (2016).

[7] N. N. Halim and E. Widodo, "Clustering of Seismicity Impacts in Indonesia using Kohonen Self Organizing Maps,"Proceedings of SI MaNIs (National Seminar on Integration of Mathematics and Islamic Values), 1, 1, 188-194, (2017).

[8] K. N. Aulia, "The Effectiveness of DBSCAN Method in Clustering Coordinate Points and Characteristics of Seismicitys in Indonesia,"

Thesis, Statistics Study Program, Faculty of Science and Mathematics, Islamic University of Indonesia, (2018).
[9] E. Rahmi, "Application of the DBSCAN Algorithm for Clustering Seismicity Regions in Indonesia, Thesis, Information Systems Study Program,"*Faculty of Science and Technology, SultasSyarif Kasim State Islamic University Riau,* (2018).
[10] I. H. Rifa, H. Pratiwi, and Respatiwulan, "Implementation of the Clara Algorithm for Seismicity Data in Indonesia,"Nat. Sem. on Mathematics Edu. Res. (SNP2M) 2019 UMT, 161-166, (2019).
[11] Hartatik and A.S.D. Cahya, "Clustering Seismicity Damage on Java Island using SOM,"J. Sci. Intech: Infor. Tech., 2, 2, 25-34, (2020).
[12] M. Ester, H. Kriegel, J. Sander, and X. Xu, A "Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *Proc. 199 Int. conf. Knowledge Discovery and Data Mining (KDD'96)*, 226-231, (1996).
[13] J. A. Katili, and P. Marks, *Geology*, Djakarta: Department Urusan Research Nasional, (1963).