Jurnal E-Komtek
Vol. 9 No. 1. (2025) pp. 137-149
https://jurnal.politeknik-kebumen.ac.id/index.php/E-KOMTEK
p-ISSN : 2580-3719 e-ISSN : 2622-3066

Clustering of Earthquakes on The Island of Java Using K-Means Algorithm Based on Magnitude and Depth

Viki Flendiansyah[1], Dimara Kusuma Hakim[2], Feri Wibowo[3], Agung Purwo Wicaksono[4]

[1-4]Department of Informatics Engineering, Universitas Muhammadiyah Purwokerto, Indonesia, 153174

dimriset@gmail.com

https://doi.org/10.37339/e-komtek.v9i1.2397

Published by Politeknik Piksi Ganesha Indonesia

Artikel Info
Submitted:
22-04-2025
Revised:
06-06-2025
Accepted:
25-06-2025
Online first :
30-06-2025

Abstract

Indonesia is one of the countries with a high level of earthquake vulnerability because it is located in the Pacific Ring of Fire. Java Island, as the most populous region and the center of the national economy, has a great risk of earthquake impacts. This study aims to analyze earthquakes in Java Island during the 2019-2024 period using the K-Means algorithm. Clustering the data based on magnitude, depth, location, and time of occurrence resulted in three clusters that reflect the characteristics of earthquakes in the region. This clustering provides important insights into the distribution and intensity of earthquakes in Java. The information obtained can be used to support disaster mitigation efforts more strategically. The government and community are expected to be able to increase preparedness for disaster risks and design effective mitigation policies to minimize the impact of future earthquakes. This research shows the great potential of applying data-driven technology as a basis for decision-making in disaster mitigation in Indonesia.

Keywords: Earthquake, Java Island, Clustering, K-Means, Disaster Mitigation.

Abstrak

Indonesia merupakan salah satu negara dengan tingkat kerawanan gempa bumi yang tinggi karena berada di wilayah Cincin Api Pasifik. Pulau Jawa, sebagai wilayah dengan populasi terpadat dan pusat perekonomian nasional, memiliki risiko besar terhadap dampak gempa bumi. Penelitian ini bertujuan menganalisis gempa bumi di Pulau Jawa selama periode 2019–2024 menggunakan algoritma K-Means. Pengelompokan data dilakukan berdasarkan magnitudo, kedalaman, lokasi, serta waktu kejadian, sehingga menghasilkan tiga klaster yang mencerminkan karakteristik gempa bumi di wilayah tersebut. Klasterisasi ini memberikan wawasan penting mengenai distribusi dan intensitas gempa bumi di Pulau Jawa. Informasi yang diperoleh dapat digunakan dalam mendukung upaya mitigasi bencana secara lebih strategis. Pemerintah dan masyarakat diharapkan mampu meningkatkan kesiapsiagaan terhadap risiko bencana serta merancang kebijakan mitigasi yang efektif guna meminimalkan dampak gempa bumi di masa mendatang. Penelitian ini menunjukkan potensi besar penerapan teknologi berbasis data sebagai dasar pengambilan keputusan dalam mitigasi bencana di Indonesia.

***Kata-kata kunci**: Gempa Bumi, Pulau Jawa, Klasterisasi, K-Means, Mitigasi Bencana.*

1. Introduction

Indonesia often attracts world attention due to the high frequency of natural disasters that hit the country, earning it the nickname "disaster museum" **[1]**. One of the most frequent natural disasters is earthquakes, which have a wide impact on various aspects of life, including major damage in terms of economy, social, business, to important infrastructure such as public facilities, roads, and buildings. Java Island, as the most populous island in Indonesia with a population of around 141 million people, has an important role as the center of government and the heart of the national economy, so that the population density and vital role make it very vulnerable to the impact of earthquakes **[2]**. When an earthquake occurs, the losses incurred are not only material, such as damage to buildings and infrastructure, but also immaterial, such as psychological trauma, loss of jobs, and disruption of social and economic activities that can last a long time. Mitigation and preparedness in dealing with natural disasters in Indonesia, especially in vulnerable areas such as Java Island, are still relatively low **[3]**. This can be seen from the lack of structured efforts to build public awareness, strengthen earthquake-resistant infrastructure, and integrate technology to support disaster management. Furthermore, Indonesia does not yet have a clear and systematic mitigation roadmap to map earthquake risk zones based on their severity, which can help in planning a more effective response.

The K-Medoids method is one of the clustering algorithms that has advantages over K-Means in producing more stable clusters and is resistant to outliers, making it suitable for data with high variability, such as earthquake damage data **[4]**. However, K-Medoids requires a longer computation time than K-Means. Therefore, integrating the K-Means method to speed up the clustering process can provide a more efficient solution without sacrificing the quality of the cluster results.

Based on these problems, a data-based approach becomes very important. One method that can be used to cluster disaster data is K-Means, a clustering algorithm that is known for its simplicity and strong ability to cluster data efficiently **[5]**. Previous research has shown that the K-Means algorithm is faster than other clustering algorithms and is able to produce quality clusters on large datasets **[6].** K-Means provides an adequate solution for various types of datasets, including earthquake data, by dividing the data into several clusters based on the similarity between the data **[7].** By utilizing the K-Means method, it is expected to be able to

group the level of damage caused by earthquakes into several more organized clusters, thus facilitating analysis and decision-making in disaster mitigation and preparedness efforts.

2. Method

This study was conducted with a quantitative approach using the K-Means Clustering method to group the level of damage caused by the earthquake. The research methodology includes the stages of data collection, Implementation of the K-Means Algorithm, Elbow Method, Comparison of Distance Metrics.

2.1 Data Collection

The data collection technique in this study uses the documentation method, namely by taking secondary data from trusted sources, namely the official USGS website. The data obtained includes time, location (latitude and longitude), depth, and magnitude of the earthquake, which are then processed further.

2.2 Implementation of the K-Means Algorithm

Time series is a quantitative method used to analyze data based on patterns that form over time. By utilizing previous data history, this method allows accurate prediction of future information **[8].** Identification of patterns and trends in ordered data makes time series effective for various applications, such as temperature forecasting using the Double Exponential Smoothing approach.

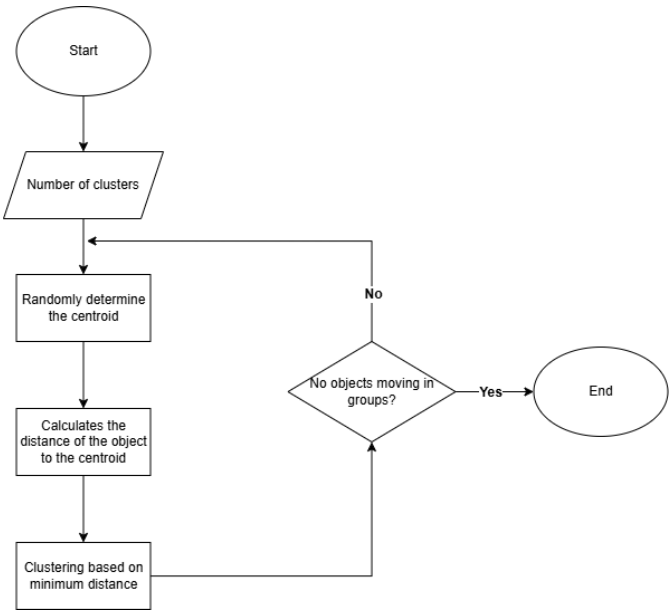

Figure 1. Flowchart K-Means

Figure 1 shows the flowchart of the K-Means calculation process. The steps of the

calculation process in the flowchart are as follows:

1. Start: Determine the initial parameters.
2. Determine the number of clusters (K): Determine the desired number of clusters.
3. Initialize centroids: Select the initial centroid randomly or using a specific method.
4. Calculate the distance of objects to centroids: Calculate the distance of each data to all centroids, generally using Euclidean distance.
5. Clustering: Group data based on the closest distance to the centroid.
6. Centroid update: Update the centroid by calculating the average position of the data in the cluster.
7. Check data movement: Check if any data has moved clusters. If so, repeat step 4.
8. Finish: The process is stopped when no data has moved clusters.

The K-Means algorithm is an unsupervised data clustering method that divides data into clusters based on similarity. The process begins by randomly selecting cluster centers, then grouping the data into clusters based on the Euclidean distance to the nearest center. Each cluster is then updated by calculating the average of the data within it. This process is repeated until the cluster centers converge. The goal of the algorithm is to maximize the uniformity within clusters and minimize the similarity between clusters. This algorithm minimizes the objective function E which is calculated by the sum of the squared errors within the cluster **[8][9].** This distance is calculated using the following Euclidean Distance formula:

$$D_e = \sqrt{(x_i + s_i)^2 + (y_i - t_i)^2} \quad (1)$$

Explanation:

D_e : Euclidean Distance

i : Number of objects

(x, y) : Object coordinates

(s, t) : Centroid coordinates

2.3 Elbow Method

The Elbow method is used to determine the optimal number of clusters in the K-Means algorithm by analyzing the relationship graph between the number of clusters and the Sum of Squared Errors (SSE). The optimal point is marked by an elbow curve, which shows the decrease in SSE begins to slow down, indicating the best number of clusters **[10][11].** The SSE calculation formula in the context of K-Means is as follows:

$$SSE = \sum_{k-1}^{k=1} \sum_{xi \in Sk} \|xi - ck\|^2 \quad (2)$$

Explanation:

x_i is the data point to i

c_j is the centroid for the k cluster

$\|...\|$ is the Euclidean norm that measures the distance between a point and the centroid

2.4 Comparison of Distance Metrics

In the K-Means algorithm, the selection of distance metrics affects the cluster accuracy. Three common metrics are Euclidean Distance, Minkowski Distance, and Manhattan Distance. The Comparative Study of the Accuracy of Euclidean Distance, Minkowski Distance, and Manhattan Distance in the K-Means Clustering Algorithm based on Chi-Square shows that Euclidean Distance is more optimal because it measures geometric distance directly in multidimensional space and is consistent across datasets **[12].** Minkowski and Manhattan Distance also provide good results, but depend on the characteristics of the dataset. Another study, Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering, compares Manhattan with Euclidean Distance and states that Manhattan is more suitable for data with grid or asymmetric distributions, while Euclidean is effective for data with natural distributions **[13].**

3. Results and Discussion

In the Double Exponential Smoothing (DES) method, an analysis is conducted on the secondary data obtained from the NASA POWER platform, which includes three main variables: surface solar radiation (ALLSKY_SFC_SW_DWN), earth surface temperature (TS), and maximum wind speed at 10 meters height (WS10M_MAX). The analysis begins with the visualization of these three variables to identify seasonal patterns and trends over the observation period. The focus is then directed toward forecasting surface temperature using the Double Exponential Smoothing (DES) method, which has undergone annual differencing. The predicted surface temperatures for the upcoming year are presented in tabular form and compared with historical patterns to assess the accuracy of the applied model.

3.1 Dataset Clustering Results

The experimental analysis was conducted using data from the earthquake data set in Java. The data has gone through a pre-processing and transformation stage to ensure its quality and feasibility before being used in further analysis using the K-Means algorithm. This dataset consists of a total of 293 records covering earthquake events in Java Island during the period 2019 to 2024. After the data mining processing stage using the K-Means algorithm, the data was successfully grouped into three clusters. The clustering results show that Cluster 0 includes 205 items, Cluster 1 consists of 18 items, and Cluster 2 includes 70 items. The results of this analysis provide a clear picture of the distribution and characteristics of earthquakes in Java Island, which can be used to support further research or decision making related to disaster mitigation. **Table 1** is the result of the Earthquake Dataset clustering, which presents the initial dataset before the clustering process. 3.2 Data Processing Using RapidMiner.

Table 1. Earthquake Dataset

Time	Latitude	Longitude	Depth	Magnitude	Place
2019-10-14T11:34:02.214Z	-82.152	1.093.923	89.44.00	4.5	66 km SSE of Kroya, Indonesia
2019-10-31T01:56:15.542Z	-78.484	1.084.412	87.49.00	4.4	55 km SW of Sidareja, Indonesia
2019-11-01T08:38:10.866Z	-77.631	108.062	91.71	4.8	45 km SSW of Kawalu, Indonesia
2019-11-09T01:19:44.806Z	-76.339	1.081.636	93.05.00	4.5	28 km S of Kawalu, Indonesia
…					
2024-11-13T01:41:08.664Z	-82.407	1.078.641	88.635	4.9	102 km SSW of Singaparna, Indonesia

3.2 Data Processing Using RapidMiner

Data mining process to group data based on earthquake magnitude and depth variables. This process aims to identify or significant relationships between the two variables. The dataset includes various parameters related to earthquake events, which have been prepared through the pre-processing stage. This study was conducted using RapidMiner software. The analysis process involves three main operators shown in **Figure 2.**

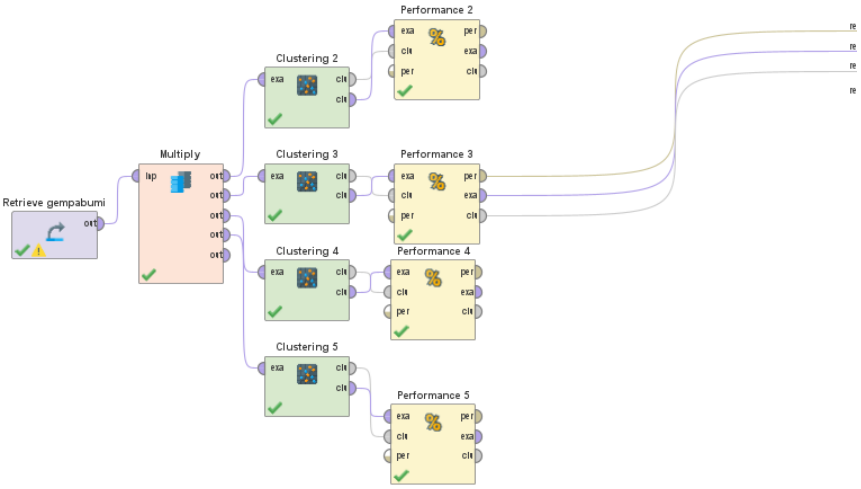

Figure 2. Data Processing Process Using RapidMiner

Figure 2 shows the design model created in the RapidMiner application consisting of data trying, set role apply model and performance. In this model, the Multiply operator is used to integrate data from Excel files with various clustering operators, ensuring that the data can be further processed using the appropriate grouping method. The Clustering operator acts as the main algorithm in the data mining process, where the processed data will be grouped based on or certain characteristics. Meanwhile, the Performance operator is used to calculate the average distance value, which is the basis for determining the optimal number of clusters using the elbow method approach.

3.3 Optimal Cluster Results with the Elbow Method

Determining the optimal number of clusters (k) is done using the Elbow method. This method uses a comparison graph of the Sum of Squared Errors (SSE) value against the number of clusters to help determine the optimal number of clusters. The optimal k value is determined at the "elbow" point, which is when the SSE decline begins to slow down significantly.

Table 2. Optimal Cluster Results with the Elbow Method

Number of Clusters (k)	Sum of Squared Errors (SSE)
2	329.60
3	177.46
4	138.24
5	105.70
6	81.17

Table 2 shows that the SSE value decreases sharply from k = 2 to k = 3, then the SSE decrease begins to slow down after k = 3. Therefore, the optimal number of clusters selected for further

analysis is 3 clusters. This number of clusters will be used as a reference in the next K-Means Clustering process.

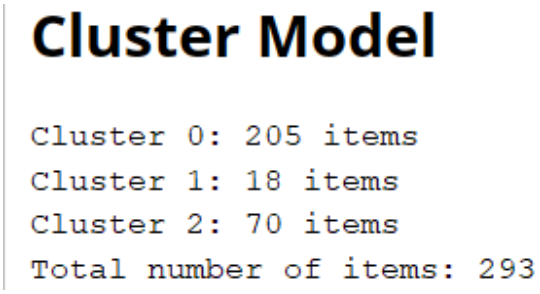

Figure 3. Cluster Model Results

Figure 3 shows the results of data analysis carried out using the RapidMiner application with the K-Means method producing three clusters from a total of 293 records. This process is carried out by setting the parameter k = 3, which means that the data is grouped into three clusters based on or certain characteristics. Cluster 0 includes 205 items, cluster 1 consists of 18 items, and cluster 2 includes 70 items. The data analyzed includes 293 earthquake events recorded on Java Island during the period 2019 to 2024.

3.4 Clustering of Earthquakes in Java Island and Its Implications in Disaster Mitigation

In **Table 3**, the results of earthquake data are grouped into three clusters based on the characteristics of the magnitude and depth of the earthquake. Cluster 0, which includes 205 items, consists of earthquakes with small magnitudes (<4.5) and moderate depths (10–277 km), which are generally not felt or only cause minimal damage. Cluster 1, consisting of 18 items, includes earthquakes with large magnitudes (4.0–7.0) but very deep depths (500–600 km), so that their impact on the surface is relatively small. Meanwhile, Cluster 2, which includes 70 items, contains earthquakes with medium magnitudes (4.5–6.0) and shallow depths (10–150 km), which have the potential to cause more damage on the surface. These results indicate that most earthquakes in Java during this period were small magnitude earthquakes with medium depths.

Table 3. Earthquake Grouping in Java (2019–2024)

Cluster	Magnitude Range (Mag)	Depth Range (km)	Description
Cluster 0	< 4.5	10 - 277	Earthquakes with small magnitudes and moderate depths, often not felt or causing minimal damage.
Cluster 1	4.0 - 7.0	500 - 600	Earthquakes with large magnitudes but occurring at great depths, typically causing less impact on the surface despite the high magnitude.
Cluster 2	4.5 - 6.0	10 - 150	Earthquakes with moderate magnitudes and shallow depths, potentially causing more noticeable damage at the surface.

This grouping is based on the variables of magnitude and depth of the earthquake, which allows the identification and main characteristics of each cluster. The cluster starts from 0 because in programming languages 0 is the first number of the numbering sequence. The scatter plot showing the results of the earthquake grouping based on magnitude and depth can be seen in Figure 5.

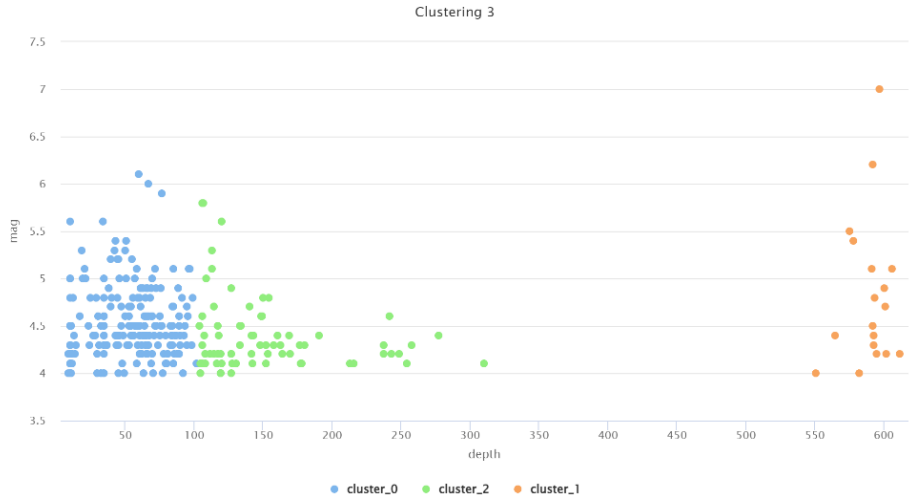

Figure 4. Earthquake Scutter Plot Based on Magnitude and Depth

Grouping earthquake data using the K-Means method provides a clear picture of the pattern of earthquake occurrences in Java. Figure 5 shows that most earthquakes are included in Cluster 0, which has characteristics of small magnitude and moderate depth and tends to be

concentrated in certain areas. Although the impact is relatively small, routine monitoring of these small earthquakes is still needed because it can be an indication of greater tectonic activity in the future.

Meanwhile, Cluster 1 includes large magnitude deep earthquakes, which indicates that these earthquakes generally occur in subduction zones and rarely cause major impacts on the surface. However, further understanding of the dynamics of deep earthquakes is still needed, especially in relation to the possibility of aftershocks or other effects on the surface. Further studies on subduction zones can help improve understanding of the characteristics of these deep earthquakes. In addition, education to the public about the phenomenon of deep earthquakes that may be felt but are not dangerous needs to be carried out so as not to cause panic.

Cluster 2 is the main focus in mitigation strategies because it covers shallow earthquakes with moderate magnitudes, which are more potentially damaging. Areas that frequently experience earthquakes in this category need to be given more attention in the form of implementing earthquake-resistant building standards, especially for high-rise buildings and public facilities. Buildings in earthquake-prone areas need to be designed to withstand shocks by paying attention to appropriate construction regulations. In addition, community preparedness training in dealing with shallow earthquakes that can have a direct impact on settlements needs to be encouraged. Regular evacuation simulations can also help communities be more prepared in dealing with potentially damaging earthquakes.

3.5 Mapping the Distribution of Earthquakes in Java Island

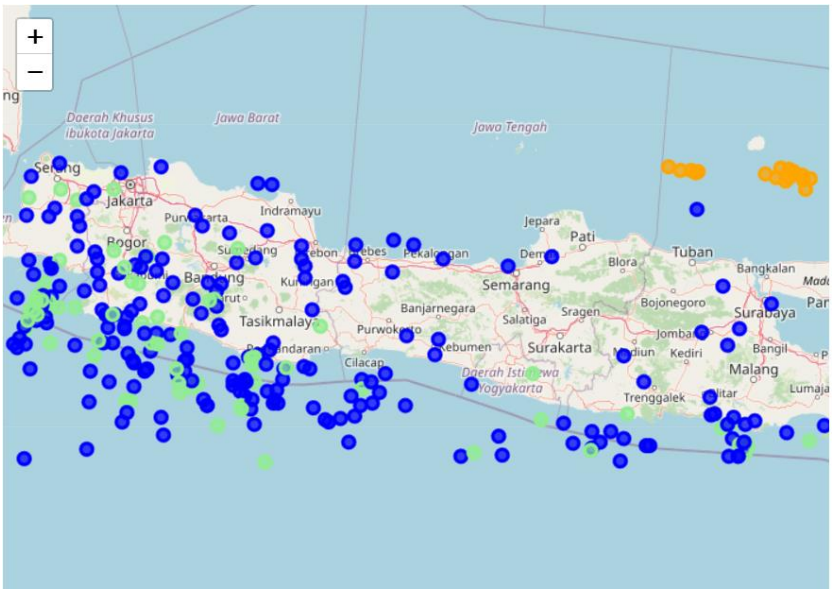

Figure 5. Mapping of earthquake distribution on Java Island

Figure 5 shows the results of earthquake distribution mapping on Java Island based on the results of clustering using the K-Means method. The map shows three clusters with different colors to facilitate the identification of earthquake characteristics in each area. Cluster 0, marked in blue, represents areas with small magnitude and moderate depth earthquakes. Earthquakes in this category generally do not cause significant impacts, but still need to be monitored to identify seismic activity patterns that may change in the future.

Cluster 1, colored orange, shows earthquakes with large magnitudes but occur at very deep depths. Earthquakes in this category usually occur in subduction zones and despite their high magnitudes, their impact on the surface is relatively small. Further understanding of the distribution of these deep earthquakes can help in long-term tectonic activity research, as well as improve community preparedness in dealing with shocks that may be felt but are not dangerous. Meanwhile, Cluster 2, marked in green, describes earthquakes with moderate magnitudes and shallow depths. Earthquakes in this cluster have the potential to have a greater impact on infrastructure and people's lives, because they occur close to the earth's surface. Areas included in this cluster need to be prioritized in implementing mitigation strategies, such as strengthening building structures, planning evacuation routes, and educating local communities about preparedness.

The results of this mapping provide important spatial information for the government in designing risk-based disaster mitigation policies. Cluster data can be used to identify areas that require priority in the development of earthquake-resistant infrastructure, emergency budget allocation, and determination of disaster-prone zones that are strengthened in regional spatial planning. The government can also use this cluster information to prepare evacuation simulations, provide outreach to the community, and strengthen early warning systems that are adjusted to the characteristics of earthquakes in each region.

With this data-based mapping, mitigation steps can be more targeted, help reduce the impact of disasters, and increase community preparedness in facing earthquake risks in Java.

4. Conclusion

The conclusion of this study shows that earthquakes in Java Island can be grouped into three clusters based on earthquake magnitude and depth using the K-Means algorithm. The results of this grouping reveal that the majority of earthquake events are concentrated in Cluster

0, which has a small magnitude and moderate depth, so it tends not to cause significant damage. Cluster 1 includes earthquakes with large magnitudes but occurs at very deep depths, which reduces their impact on the surface. Meanwhile, Cluster 2 contains earthquakes with moderate magnitudes and shallow depths, which have the greatest potential to cause infrastructure damage and social disruption. This grouping provides important insights into earthquake disaster mitigation strategies. Areas that frequently experience earthquakes in Cluster 2 should be prioritized in implementing earthquake-resistant building standards, preparing evacuation routes, and educating the community about preparedness. Meanwhile, Cluster 1 can be used for further studies related to seismic dynamics in the depths of the earth, and Cluster 0 still needs to be monitored to detect seismic activity patterns that may change in the future. By using the K-Means method and a data-based approach, the results of this study are expected to assist in more effective mitigation planning and improve disaster preparedness in Java.

References

[1] H. Artatia and R. B. F. Hakim, "Pengelompokan Dampak Gempa Bumi Dari Segi Kerusakan Fasilitas pada Provinsi yang Berpotensi Gempa di Indonesia Menggunakan K-Means-Clustering," *Pros. Semin. Nas. Mat. dan Pendidik. Mat. UMS*, no. 3, pp. 28–47, 2015, [Online]. Available: http://publikasiilmiah.ums.ac.id/handle/11617/5765

[2] 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.

[3] P. Novianti, D. Setyorini, and U. Rafflesia, "K-means cluster analysis in earthquake epicenter clustering," *Int. J. Adv. Intell. Informatics*, vol. 3, no. 2, pp. 81–89, 2017, doi: 10.26555/ijain.v3i2.100.

[4] D. P. Sari, M. Rosha, and D. Rosadi, "Disaster Mitigation Efforts Using K-Medoids Algorithm and Bayesian Network," *EKSAKTA Berk. Ilm. Bid. MIPA*, vol. 23, no. 03, pp. 231–241, 2022, doi: 10.24036/eksakta/vol23-iss03/304.

[5] D. H. Fisher, "Knowledge Acquisition Via Incremental Conceptual Clustering," *Mach. Learn.*, vol. 2, no. 2, pp. 139–172, 1987, doi: 10.1023/A:1022852608280.

[6] S. Saraswathi and M. I. Sheela, "Croydon's water supply: Safeguarding measures," *Br. Med. J.*, vol. 1, no. 4037, p. 1119, 1938, doi: 10.1136/bmj.1.4037.1119.

[7] P. Prihandoko and B. Bertalya, "a Data Analysis of the Impact of Natural Disaster Using K-Means Clustering Algorithm," *Kursor*, vol. 8, no. 4, p. 169, 2017, doi: 10.28961/kursor.v8i4.109.

[8] A. Likas, N. Vlassis, and J. Verbeek, "The global k-means clustering algorithm Intelligent Autonomous Systems," *ISA Tech. Rep. Ser.*, pp. 1–11, 2011.

[9] W. Jianguo and X. Linyao, "Application of K-means Algorithm in Geological Disaster Monitoring System," *Int. J. Adv. Network, Monit. Control.*, vol. 3, no. 3, pp. 16–22, 2018, doi: 10.21307/ijanmc-2019-002.

[10] E. Umargono, J. E. Suseno, and V. G. S. K., "K-Means Clustering Optimization using the Elbow Method and Early Centroid Determination Based-on Mean and Median," no.

Conrist 2019, pp. 234–240, 2020, doi: 10.5220/0009908402340240.

[11] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.

[12] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 20–24, 2019, doi: 10.30591/jpit.v4i1.1253.

[13] D. Sinwar and R. Kaushik, "An Dengineering Technnology (I J R A S E T) Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering," no. May, 2014.