JOURNAL OF FUNDAMENTAL MATHEMATICS
AND APPLICATIONS (JFMA) VOL. 7 NO. 2 (NOV 2024)
Available online at www.jfma.math.fsm.undip.ac.id

RECOGNIZING THE SPATIAL DISTRIBUTION AND VORONOI PATTERNS OF THE RECORDED EARTHQUAKE EPICENTERS IN SUNDA STRAIT, INDONESIA

Triyana Muliawati[1], Fuji Lestari[2]*, Mika Alvionita[3], Ardika Satria[4], DG Harbowo[5]

[1]Department of Mathematics, Institut Teknologi Sumatera, 35365, Wayhui, Lampung, Indonesia
[2,4]Department of Actuarial Science, Institut Teknologi Sumatera, 35365, Wayhui, Lampung, Indonesia
[3]Department of Science Data, Institut Teknologi Sumatera, 35365, Wayhui, Lampung, Indonesia
[5]Center for Sustainable Development Goals Studies (ITERA SDGs Center), Institut Teknologi Sumatera, 35365, Wayhui, Lampung, Indonesia and Department of Geology, Institut Teknologi Sumatera, 35365, Wayhui, Lampung, Indonesia

Email : [1]triyana.muliawati@ma.itera.ac.id, [2]fuji.lestari@at.itera.ac.id, [3]mika.alvionita@sd.itera.ac.id, [4]ardika.satria@sd.itera.ac.id, [5]danni.gathot@gl.itera.ac.id
[*]Corresponding author

Abstract. Currently, Sunda Strait is one of the most active transportation hubs. However, this region also bears a notable history of geohazards associated with the dynamics of tectonic activity of the Eurasian and Indo-Australian tectonic plates, such as the super-eruption of Krakatoa volcano in 1883, the Sunda Strait tsunami in 2018, and decades of frequent earthquakes. To address these challenges, this study conducted a statistical analysis of the frequency and distribution of seismic activities in the Sunda Strait region based on recorded epicenter data in the United States Geological Survey's (USGS) Earthquake catalog. We assembled 440 multivariate earthquake data points between 1990 and 2023 (over three decades). The results of this study indicate that the machine learning approach precisely identifies four relevant parameters for k-means clustering, followed by an analysis of silhouette values to recognize Voronoi patterns ($k = 3$, $k = 5$, $k = 10$, $k = 15$, and $k = 20$). These statistical patterns also have significant implications for the number of epicenter clusters and recognizing its spatial distribution. It provides a new understanding of the spatial-temporal characteristics and locates the list of frequent earthquake regions. Having all the necessary information would help to comprehensively evaluate geohazard risks in Sunda Strait region.

Keywords: earthquake, k-means clustering, silhouette, sunda strait, voronoi.

I. INTRODUCTION

Indonesia is included in the Pacific Ring of Fire, which contains around 75% of all the volcanoes on Earth. The formation of volcanic arcs in Indonesian territory will impact conditions of lithospheric activity in Indonesia, such as earthquakes, tsunamis and eruptions [1,2]. The Sunda Strait is an active transportation centre that connects Java and Sumatra. Geohistorical records show that there was a global impact due to the tectonic activity of the Eurasian and Indo-Australian tectonic plates, such as the large eruption and tsunami of Krakatoa (1883), the Sunda Strait earthquake and tsunami (2018), and the natural disasters of earthquakes that have occurred over the last few decades. An earthquake is a natural event that occurs when energy is released through seismic waves beneath the Earth's surface. These activities cause extensive damage, causing massive numbers of victims, and impact global climate change [3-5].

Fig 1. Indonesia's geographical position, especially the location of the Pacific Ring of Fire which resulted in earthquakes in Sunda Strait; Map courtesy: Google Earth.

Determining an earthquake's location is an important thing to consider [6]. Using a machine learning approach, we analysed the frequency and distribution of seismic data in the Sunda Strait region. k-means clustering is a clustering algorithm used in machine learning and data analysis [7-8]. The goal of k-means clustering is to divide a set of data into groups called clusters based on similar data characteristics. Silhouette analysis measures how well the data has been grouped by k-means clustering into relevant clusters. Search for a location in a group you want to know in spatial-temporal computational geometry via a Voronoi diagram. The Voronoi diagram shows the division of a particular area into several parts called cells, where each part contains one location point (site). Each point in a cell is closer to the site in that cell compared to other sites in that region. Thus, each point in the region has been paired with the closest site [9]. The main aim of this research is to determine earthquake-prone locations in the hope of helping understand seismic patterns and activity and evaluating geohazard risks in the Sunda Strait.

II. DATA AND METHOD

3.1 Dataset

The review of earthquake dynamics phenomena in this study covers the earthquake events in Indonesia from 1 January 1990 to 31 July 2023 (for 3 decades). Tectonic earthquake activity data is compiled from the United States Geological Survey's (USGS) Earthquake Catalog global earthquake database at earthquake.usgs.gov; in this case, the data extracted is only for the Sunda Strait region. There were 440 earthquake data collected, including the time of the earthquake, epicentre coordinates (latitude and longitude), earthquake magnitude, and depth of the earthquake epicentre (**Fig. 2** and **Fig. 3**).

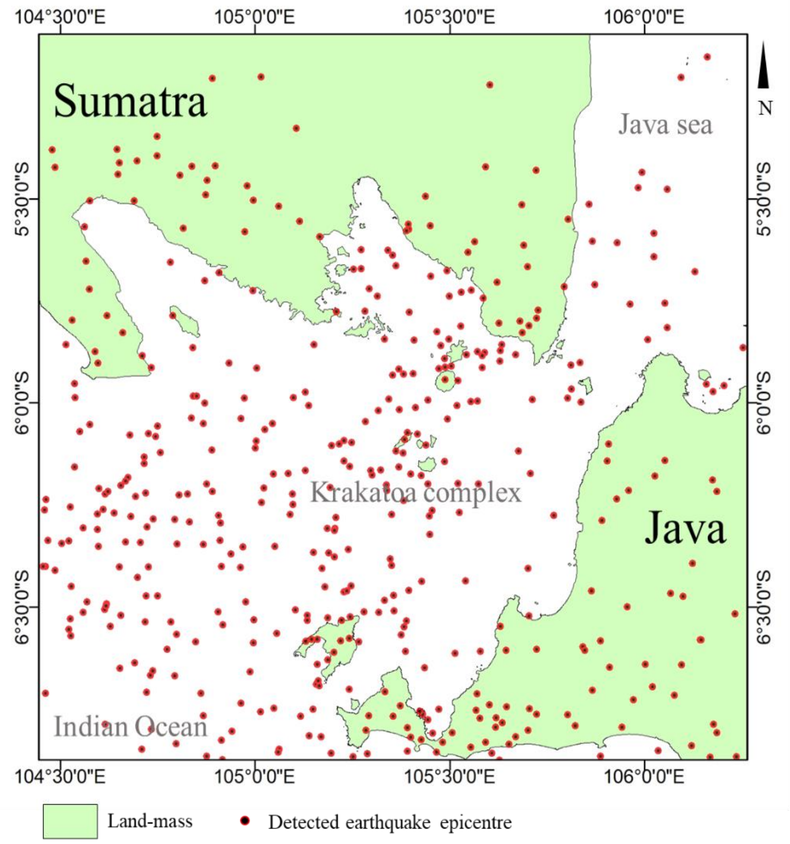

Fig 2. Detected earthquake epicenter locations in Sunda Strait region from Januari 1990 to July 2023

Fig 2. is the distribution location of earthquake data that occurred around the Sunda Strait area, between the islands of Sumatra and Java, where depth and magnitude values can be measured. Earthquakes are measured using the moment magnitude scale. This scale measures the strength of an earthquake, with a higher magnitude value indicating a stronger earthquake. An earthquake's depth can vary from very shallow (a few kilometres below the surface) to very deep (hundreds of kilometres below the surface). Earthquake depth is the vertical distance from the surface to the epicentre.

JOURNAL OF FUNDAMENTAL MATHEMATICS
AND APPLICATIONS (JFMA) VOL. 7 NO. 2 (NOV 2024)
Available online at www.jfma.math.fsm.undip.ac.id

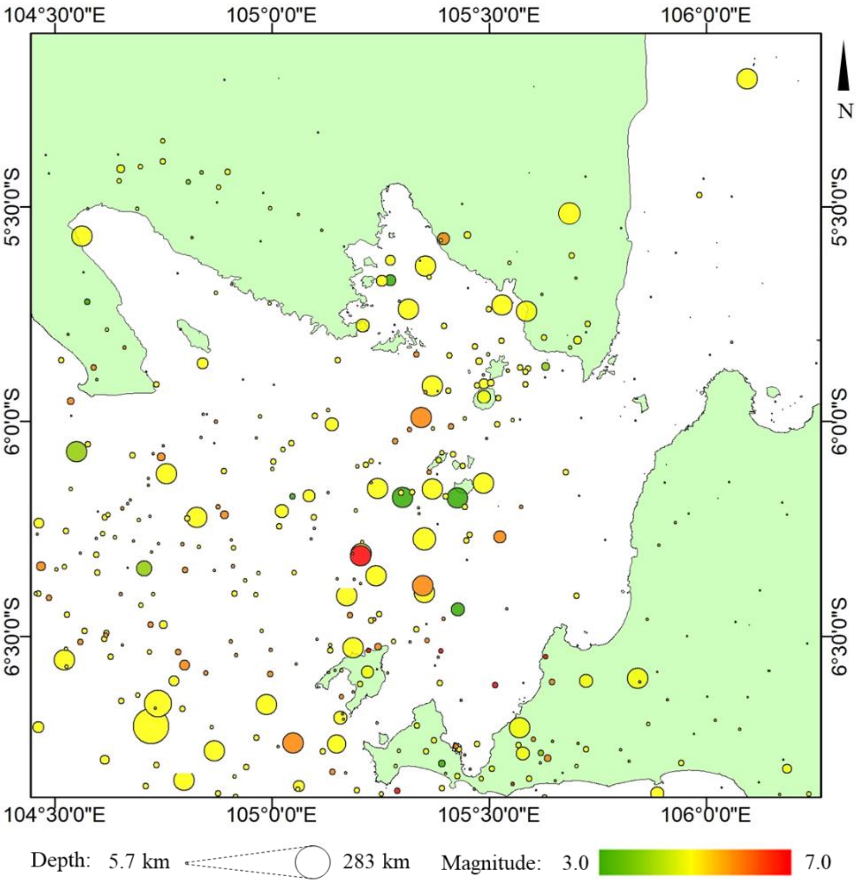

Fig 3. Distribution of earthquake epicenter depth (km) and magnitude (mag) for this dataset from January 1990 to June 2023.

The distribution of the earthquake can be seen in terms of depth (km) and magnitude (mag) in **Fig. 3**. The depth that occurred was between 5.7 km – 283 km (the bigger the circle, the deeper the earthquake occurred) and the magnitude that occurred was between 3.0 – 7.0 (the greater the mag value, the greater the earthquake vibration that occurs). The Krakatoa volcano area is an area that quite often experiences earthquakes with magnitudes of up to 7.0. This high magnitude has a destructive impact and can even trigger a tsunami if it occurs under the sea.

3.2 k-Means clustering

One of the machine learning algorithms used to analyze data grouping into several different groups is k-means clustering. This grouping is formed based on each group's data similarity level by minimizing variation within clusters and maximizing variation between clusters. Following are the steps of the k-means clustering algorithm [10-12]:

1. Random initial selection of the number of clusters (k) and centroids to be created.
2. Calculate the Euclidean distance between data points and all centroids (**Equation 1**) [7-8]. The value of the closest distance between the data point and the centroid causes the data to be similar to that cluster.

$$D(x_i, C_i) = \|x_{ij} - C_{ij}\|_2 = \sqrt{\sum_{j=1}^{p} \left(x_{ij} - C_{ij}\right)^2} \quad (1)$$

3. Recalculate the new centroid by taking the average value of all data points in each cluster (**Equation 2**).

$$C_i = \frac{1}{M}\sum_{j=1}^{M} x_j \quad (2)$$

where M represents the amount of data in a cluster, i represents the i th feature, and p represents the data dimension.
4. Repetition of steps 2 and 3 is carried out until the stopping criteria are met, namely, the centroid no longer changes, the data points in each cluster no longer change, and the number of iterations has been reached.

The choice of the number of clusters (k) significantly affects the quality of the clusters. So, an analysis technique using the silhouette method was carried out to determine the appropriate and optimal k value for this research.

3.3 Silhouette Method

The silhouette method in k-means clustering is intended to determine the optimal number of clusters (k). Silhouette values are between -1.00 to 1.00. If the silhouette value is close to 0, then the data is on the boundary between two clusters, or the cluster is unclear. If the silhouette value is closer to 1, the data is similar to its cluster and not to other clusters.
Here are the steps for the silhouette method [13]:
1. Select the number of clusters (k) to be created. In this research, the k chosen is $k = 3$, $k = 5$, $k = 10$, $k = 15$, and $k = 20$.
2. Apply the k-means clustering algorithm to each selected k value in the data.
3. Calculation of silhouette values using

$$s_i = \frac{b_i - a_i}{(a_i, b_i)} \quad (3)$$

wheres a_i is the average distance from the data point to the other data points in the same cluster and b is the smallest average distance from the data point to the data points in a different cluster.
4. Calculation of the average value of all silhouette values with **Equation 4**.

$$S = \frac{1}{n}\sum_{i=1}^{n} s_i \quad (4)$$

with n representing the amount of data.
5. Selecting the highest average silhouette (S) value in k-means clustering shows better grouping with the optimal k value.

3.4 Voronoi Diagram

Voronoi diagrams are a mathematical concept used in geographic information systems (GIS) to analyze spatial data, such as finding the nearest location of an earthquake. The division of a field into several regions in a Voronoi Diagram is based on a set of seed points. Each region in a Voronoi diagram corresponds to a region closer to a particular seed point than other seed points. Voronoi diagrams have the following characteristics:

1. The seed point (p_i) is the center of the diagram.
2. Voronoi regions are polygonal or irregular in shape. The Voronoi region is defined by each seed point, consisting of all points in the plane closer to the seed point.
3. Voronoi edge (e) is a boundary between Voronoi regions formed by a set of points that are equidistant between two neighboring seed points.
4. Voronoi vertex (v) is the intersection point of three or more Voronoi edges, which means the Voronoi vertex neighbors three or more seed points.

The following is an illustration of a Voronoi diagram

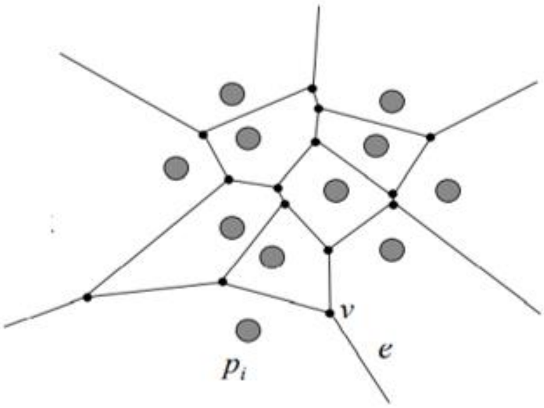

Fig 4. Voronoi diagram illustration

This research data processing uses Google Collaboratory with the Python programming language, and visualization uses ArcGIS software [14-15].

III. RESULT AND DISCUSSION

Data from earthquake location points are formed into clusters with the help of silhouette analysis $k = 3$, $k = 5$, $k = 10$, $k = 15$, and $k = 20$ cluster. Furthermore, spatial mapping of the cluster results was carried out using a Voronoi diagram; the data distribution was obtained, which was plotted using ArcGIS software.

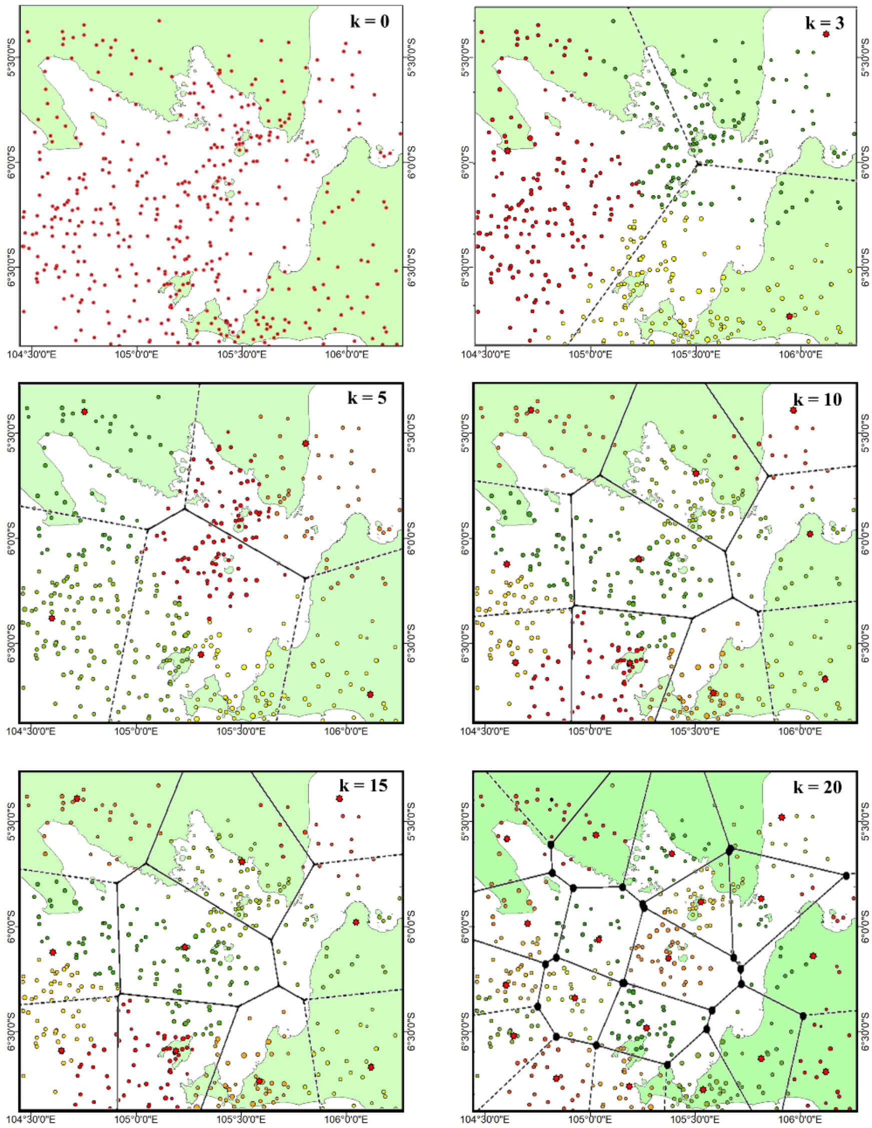

Fig 5. Comparison of variation k value of k-Means clustering (k = 0, 3, 5, 15, 20).

It can be seen in **Fig. 5** that there are several variations of k in the grouping in k-means clustering. The k-means clustering results divide k regions on the Voronoi diagram. The smaller the k value used, the more extensive and less detailed the Voronoi regions formed because the data will be grouped into a few large clusters. Meanwhile, the more k, the smaller and more detailed the Voronoi regions formed because the data will be grouped into more small clusters. In this case, having more Voronoi areas can provide information on which areas significantly influence the occurrence of earthquakes in the Sunda Strait region. This earthquake-prone area ranges from the Anak Krakataoa area [16-17]. This detailed location information can help us understand seismic patterns and activity and mitigate risks in areas that are the centre of Voronoi in the Sunda Strait region.

IV. CONCLUSION

Based on Sunda Strait earthquake data from 1990 to 2023, there are two largest silhouette analysis values, namely 0.40 and 0.39, which are located at $k = 3$ and $k = 5$ in k-means clustering. The higher the silhouette value and closer to the value 1, which means that the earthquake clusters in the cluster have a fairly good distance from each other and far from other earthquake clusters, the better the clustering performance. In this case, the statistical pattern of earthquake data distribution formed from the Voronoi diagram can imply regional clusters and provide new information regarding spatial-temporal characteristics. This information helps find/recognize and identify areas where earthquake activity is not observed, namely around the Anak Krakatoa. The successful identification of this study led to field studies and the development of an early warning system perspective. The scientific facts in this study serve as a reference to enrich perspectives and considerations for evaluating geohazard risks in the Sunda Strait region.

ACKNOWLEDGEMENT

We express our gratitude to the Faculty of Science of Institut Teknologi Sumatera for research grant assistance. We also thank our colleagues for their fruitful discussions in producing this research.

REFERENCES

[1] Whelley, P.L., Newhall, C.G. and Bradley, K.E., 2015. The frequency of explosive volcanic eruptions in Southeast Asia. *Bulletin of volcanology*, *77*, pp.1-11.

[2] Hall, R., 2017. Southeast Asia: New views of the geology of the Malay Archipelago. *Annual Review of Earth and Planetary Sciences*, *45*, pp.331-358.

[3] Wilson, N., Valler, V., Cassidy, M., Boyd, M., Mani, L. and Brönnimann, S., 2023. Impact of the Tambora volcanic eruption of 1815 on islands and relevance to future sunlight-blocking catastrophes. *Scientific Reports*, *13*(1), p.3649.

[4] Lorito, S., Romano, F. and Lay, T., 2022. Tsunamigenic major and great earthquakes (2004–2013): source processes inverted from seismic, geodetic, and sea-level data. *Complexity in Tsunamis, Volcanoes, and their Hazards*, pp.247-298.

[5] D'Arrigo, R., Carslaw, K.S., Zhu, T., Dessler, A.E., Forster, P. and Polvani, L.M., 2019, December. Northern Hemisphere wintertime continental warming following large, low-latitude volcanic eruptions? No evidence from Pinatubo (1991) and Krakatau (1883). In *AGU Fall Meeting Abstracts* (Vol. 2019, pp. A21B-01).

[6] Dumont, S., Custódio, S., Petrosino, S., Thomas, A.M. and Sottili, G., 2023. Tides, earthquakes, and volcanic eruptions. *A Journey Through Tides*, pp.333-364.

[7] Muliawati, T., Harbowo, D. G., Lubis, A. M. F., Turnip, J. D., Irda, E. R., Azahra, A., & Marito, Y. (2023). k-Means Clustering to Enhance the Petrified Wood Composition Data Analyses and Its Interpretation. *Indonesian Journal of Applied Mathematics*, *3*(1), 26-33.

[8] Muliawati, T., and D. G. Harbowo. "A Statistical review of the dates and patterns of volcanic activity of Lewotolo Volcano, East Nusa Tenggara, Indonesia." *IOP Conference Series: Earth and Environmental Science*. Vol. 1245. No. 1. IOP Publishing, 2023.

[9] Berg, O. G., & von Hippel, P. H. (1987). Selection of DNA binding sites by regulatory proteins: Statistical-mechanical theory and application to operators and promoters. *Journal of molecular biology*, *193*(4), 723-743.
[10] Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern recognition*, *36*(2), 451-461.
[11] Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, *1*(6), 90-95.
[12] Muliawati, T., and Hendri Murfi. "Eigenspace-based fuzzy c-means for sensing trending topics in Twitter." *AIP Conference Proceedings*. Vol. 1862. No. 1. AIP Publishing, 2017.
[13] Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, *6*(2), 48-54.
[14] Rydelek, P. A., Davis, P. M., & Koyanagi, R. Y. (1988). Tidal triggering of earthquake swarms at Kilauea volcano, Hawaii. *Journal of Geophysical Research: Solid Earth*, *93*(B5), 4401-4411.
[15] Petrosino, S., Cusano, P., & Madonia, P. (2018). Tidal and hydrological periodicities of seismicity reveal new risk scenarios at Campi Flegrei caldera. *Scientific reports*, *8*(1), 1-12.
[16] Harbowo, Danni Gathot. "An assessment of the scientific value of Krakatoa, Indonesia from a geoheritage perspective." *Journal of Applied Geoscience and Engineering* 2.1 (2023): 11-25.
[17] Natalia, Happy Christin, Danni Gathot Harbowo, and Rinaldi Ikhram. "Potensi Geodiversity di Sekitar Kawasan Anak Krakatau-Way Kambas, Provinsi Lampung, Sebagai Kandidat Geopark Indonesia." *Journal of Science and Applicative Technology* 5.1 (2021): 47-57.