Volume 4 Issue 2 (2024) Pages 189-198 P-ISSN:2798-2696 E-ISSIN:2798-0057
JOSTECH
Journal of Science and Technology
Website: https://ejournal.uinib.ac.id/jurnal/index.php/jostech Email: jostech@uinib.ac.id

Recognition of Voronoi Cell Distribution in Earthquake Epicenter Data in the Sunda Strait Region, Indonesia

Triyana Muliawati[1], Fuji Lestari[✉2], Danni Gathot Harbowo[3], Mika Alvionita S[4]
Department of Mathematics, Institut Teknologi Sumatera, Indonesia [1]
Department of Actuarial Science, Institut Teknologi Sumatera, Indonesia [2]
Department of Geological Engineering, Institut Teknologi Sumatera, Indonesia [3]
Department of Data Science, Institut Teknologi Sumatera, Indonesia [4]
email: triyana.muliawati@ma.itera.ac.id [1], fuji.lestari@at.itera.ac.id[2], danni.gathot@gl.itera.ac.id [3]mika.alvionita@sd.itera.ac.id[4]

Received 31 Augustus 2024, Accepted 27 September 2024, Published 30 September 2024

Abstrak

Selat Sunda merupakan salah satu jalur transportasi tersibuk di Indonesia dan dikenal dengan sejarah bencana geologi yang signifikan yang disebabkan oleh aktivitas tektonik dinamis dari lempeng tektonik Eurasia dan Indo-Australia. Bencana-bencana tersebut termasuk letusan super gunung berapi Krakatau pada tahun 1883, tsunami Selat Sunda pada tahun 2018, dan puluhan tahun gempa bumi yang sering terjadi. Untuk menghadapi tantangan ini, penelitian ini melakukan analisis statistik terhadap frekuensi dan distribusi aktivitas seismik di region Selat Sunda berdasarkan data episenter yang tercatat dalam katalog Gempa Bumi dari United States Geological Survey (USGS). Kami mengumpulkan 440 data gempa bumi multivariat antara tahun 1990 dan 2023 (lebih dari tiga dekade). Hasil penelitian ini menunjukkan bahwa pendekatan pembelajaran mesin secara tepat mengidentifikasi empat parameter relevan untuk pengelompokan *k-means*, diikuti oleh analisis evaluasi pengelompokan dengan menggunakan nilai *silhouette* untuk merekognisi distribusi sel Voronoi yang digunakan untuk membagi ruang dalam beberapa wilayah. Berdasarkan data gempa Selat Sunda tahun 1990 hingga 2023, terdapat dua nilai analisis silhouette terbesar yaitu 0.40 dan 0.39 yang terletak pada $k = 3$ dan $k = 5$. Ini mengindikasikan bahwa pengelompokan ini memiliki validitas yang baik dalam mengidentifikasi area rawan gempa. Pendekatan ini telah mengenali dan mengidentifikasi *cell areas* aktivitas gempa di Selat Sunda, khususnya di sekitar Anak Krakatau. Studi ini memberi pemahaman baru tentang karakteristik spasial-temporal dan mengidentifikasi kelompok daerah rawan gempa. Informasi yang dihasilkan dalam studi ini mempermudah mengevaluasi risiko bencana yang diakibatkan oleh gempa di masa depan khususnya yang memiliki episenter di region Selat Sunda.

Kata Kunci: gempa, *k-means clustering*, selat sunda, metode silhouette, voronoi

Abstract

The Sunda Strait is one of the busiest transportation routes in Indonesia and is known for its significant geological disaster history caused by the dynamic tectonic activity of the Eurasian and Indo-Australian tectonic plates. These disasters include the supervolcanic eruption of Krakatoa in 1883, the Sunda Strait tsunami in 2018, and decades of frequent earthquakes. To address these challenges, this study analyzes the frequency and distribution of seismic activity *in the Sunda Strait region based on epicenter data recorded in the United States Geological Survey (USGS) Earthquake Catalog. We collected 440 multivariate earthquake data points*

between 1990 and 2023 (over three decades). The results of this study show that a machine learning approach accurately identified four relevant parameters for k-means clustering, followed by a silhouette value analysis to recognize the distribution of Voronoi cells. Based on earthquake data from the Sunda Strait between 1990 and 2023, there are two highest silhouette analysis values, 0.40 and 0.39, found at $k = 3$ and $k = 5$. This indicates that the clustering has good validity in identifying earthquake-prone areas. This approach has recognized and identified the cell areas of earthquake activity in the Sunda Strait, particularly around Anak Krakatoa. This study provides new insights into the spatiotemporal characteristics and identifies clusters of earthquake-prone areas. The information generated in this study facilitates the evaluation of future earthquake disaster risks, especially those with epicenters in the Sunda Strait region.

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

✉ Corresponding author

INTRODUCTION

Indonesia is included in the Pacific Ring of Fire, which contains around 75% of all volcanoes on Earth[1]. The formation of volcanic arcs in Indonesian territory will impact lithospheric activity conditions in Indonesia, such as earthquakes, tsunamis, and eruptions [2][3]. The Sunda Strait, a vital transportation route between Java and Sumatra, has a significant geological history due to the interaction of the Eurasian and Indo-Australian plates. Several major disasters, such as the Krakatoa eruption and tsunami in 1883 and the Sunda Strait earthquake and tsunami in 2018, demonstrate the global impact of this tectonic activity [4][5]. Earthquakes occur when seismic energy is released from beneath the Earth's surface, causing widespread damage, large numbers of casualties, and even impacting global climate change [6].

Figure 1. Indonesia's geographical location, especially the location of the Pacific Ring of Fire, which resulted in earthquakes in the Sunda Strait (red box); Map courtesy: Google Earth

Determining an earthquake's location is an important thing to consider. 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. The goal of k-means clustering is to divide a set of data into groups called clusters based on similar data characteristics. k-means can be used on various types of data that have various dimensions, both data in two dimensions and data with higher dimensions [7]. This makes k-means suitable for many applications across various fields, one of which is well-suited for earthquake data clustering. Silhouette analysis measures how well the earthquake data has been grouped by k-means clustering into relevant clusters. In spatial-temporal computational geometry, a Voronoi diagram can be used to determine the location of a group of interest. A Voronoi diagram divides a region into several parts called cells, with one location point (site) at the center of each cell. 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 area has been paired with the nearest site [8]. The main aim of this research is to determine earthquake-prone locations in the hope of helping understand the recognition of the distribution of Voronoi cells from seismic activity and evaluating the risk of geohazards in the Sunda Strait region [9].

METHOD

Data

The review of earthquake dynamics phenomena in this study covers the earthquake events in Indonesia from 1 January 1990 to 31 July 2023 (over three 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 are 440 earthquake data collected, including the time of the earthquake, epicentre coordinates (latitude and longitude), earthquake magnitude, and depth of the earthquake epicentre (Figure 2).

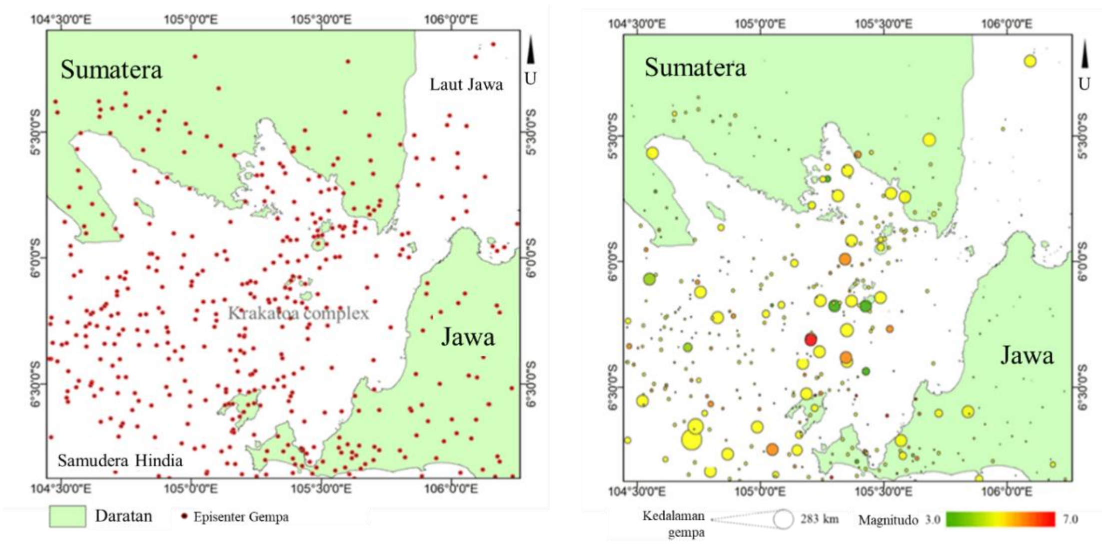

Figure 1. The location of the earthquake epicentre was detected in the Sunda Strait region from January 1990 to July 2023 (left). Distribution of earthquake epicentre depth (km) and magnitude (mag) for this dataset from January 1990 to June 2023 (right)

Figure 2 shows the distribution location of earthquake data around the Sunda Strait area, between the islands of Sumatra and Java, where the 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. From Figure 2 (right), the distribution of the earthquake can be seen in terms of depth (km) and magnitude (mag). In the figure, the magnitude occurred between 5.7 km – and 283 km (The larger the circle at the depth of the earthquake, the deeper the quake occurred), and the magnitude that occurred was between 3.0 and 7.0 (the more critical the mag value, the greater the earthquake vibration that happened). 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.

k-Means Clustering

One of the machine learning algorithms used to analyze data grouping into several different groups is k-means clustering [10]. This grouping is formed based on each group's data similarity level by minimizing variation within clusters and maximizing variation between clusters[11]. Following are the steps of the k-means clustering algorithm:

1. Random initial selection of the number of clusters (k) and centroids to be created [12].
2. Calculate the Euclidean distance between data points and all centroids (Equation 1). 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 ith feature, and p represents the data dimension.

4. Steps 2 and 3 are repeated 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 their quality. So, the silhouette method was used to analyze the data to determine the appropriate and optimal k value to be used in this research.

Silhouette Method

The silhouette method in k-means clustering is intended to determine the optimal number of clusters (k) [13]. 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:

1. Select the number of clusters (k) to be created. In this study, the k chosen were $k = 3, 5, 10, 15,$ dan 20.
2. Apply the k -means clustering algorithm to each selected k value in the data.
3. Silhouette value calculation using Equation 3.

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

   Where a_i is the average distance from a data point to another in the same cluster, and b is the smallest average distance from a data point to a data point in a different cluster.
4. Calculation of the average value of all silhouette values using Equation 4.

$$S = \frac{1}{n}\sum_{i=1}^{n} s_i \tag{4}$$

   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.

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 plane into several areas in a Voronoi Diagram is based on a set of seed points. Each region in a Voronoi diagram corresponds to a location closer to a particular seed point than others. Voronoi diagrams have the following characteristics:Titik seed (p_i) merupakan pusat dari diagram.

1. Voronoi areas are polygonal or irregular in shape. The Voronoi area is defined by each seed point, consisting of all points in the plane closer to the seed point.

2. Voronoi edge (e) is the boundary between Voronoi areas formed by a set of central points between neighbouring seed points.
3. Voronoi vertex (v) is the intersection point of three or more Voronoi edges, which means the Voronoi vertex neighbours three or more seed points.

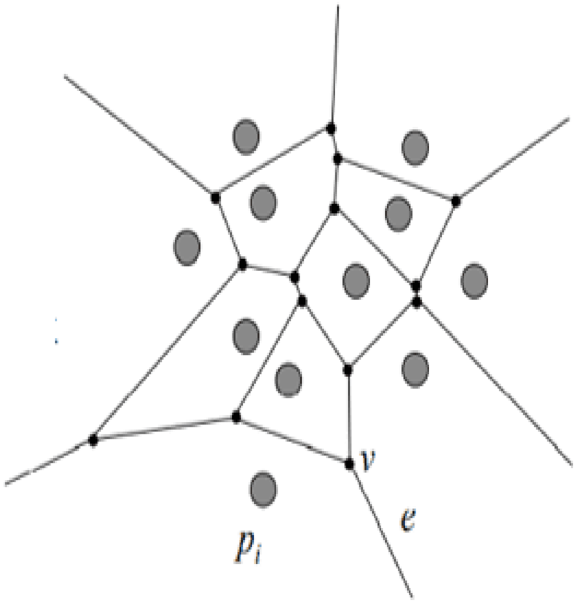

Figure 2. Voronoi diagram illustration

The tools for processing this research data are Google Collaboration with the Python programming language, and spatial data visualization uses QGIS software.

RESULT AND DISCUSSION

Data from earthquake location points were formed into clusters using silhouette analysis with $k = 3, 5, 10, 15,$ dan 20 clusters. Furthermore, spatial mapping of the cluster results was carried out using the Voronoi cell distribution, and the data distribution was obtained and plotted using QGIS software. From Figure 5, it can be seen that there are several variations of k in the grouping in k-means clustering. The k-means clustering results divide k areas into the Voronoi cell distribution.

The smaller the k value used, the more extensive and less detailed the Voronoi cell areas formed will be because the data will be grouped into a few large clusters.

Meanwhile, the more k, the smaller and more detailed the Voronoi area formed because the data will be grouped into more small clusters.

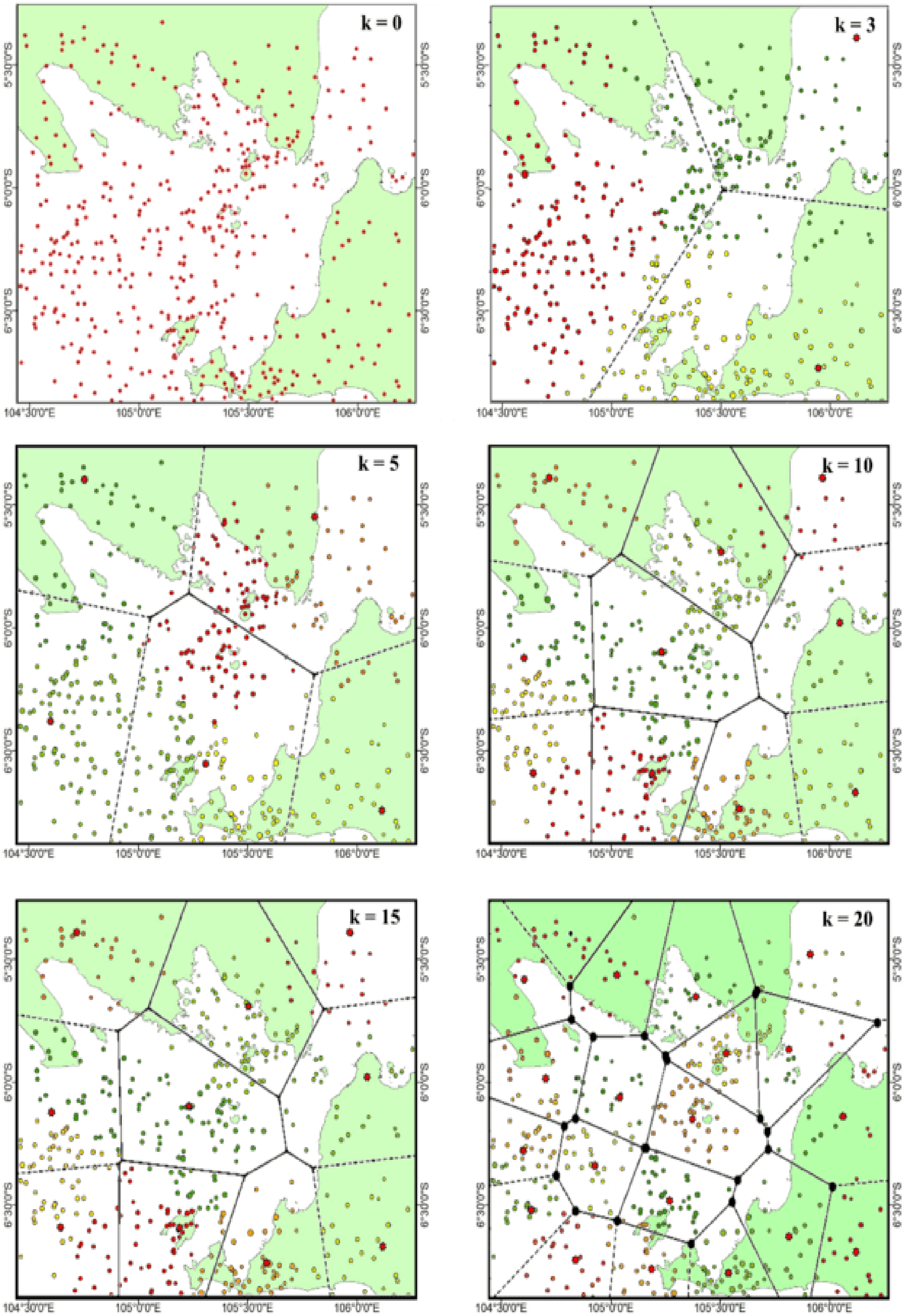

Figure 3. Comparison of k value variations of k-means clustering (k = 0, 3, 5, 10, 15, 20).

In this case, having more Voronoi cell 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 Krakatau Volcano area. This detailed location information can help us recognise the distribution of Voronoi cells in earthquake epicentre data and mitigate risks in areas that are the centre of Voronoi in the Sunda Strait region.

This information helps find/recognize and identify volcanic areas whose activity can trigger or trigger earthquakes, namely the Anak Krakatau volcano [14]. The successful identification of this study led to field studies and the development of an early warning system perspective. The scientific facts in this research serve as a reference to enrich the perspective and considerations in evaluating geohazard risks in the Sunda Strait region, considering the increasing potential for human activity in the Sunda Strait region and the growing tourism value in this region [15][16].

The findings of this study need to be followed up on for comprehensive research regarding the potential and spatio-temporal prediction of earthquake events. A comparison of data grouping with a higher k value accompanied by a hierarchical clustering approach can be used to compare the output of the Voronoi cells that will be formed. It is hoped that a more detailed spatiotemporal study in the following research will be able to reveal long-term periodic patterns that are important for mitigating earthquake disasters in the Sunda Strait region, along with the integration of evidence from geophysical and geological data such as the types of components that make up the crust and the continuity of faults in each area formed by Voronoi cells.

CONCLUSION

Based on Sunda Strait earthquake data from 1990 to 2023, there are two most significant silhouette analysis values, 0.40 and 0.39, located at $k = 3$ and $k = 5$ in k-means clustering. The higher the silhouette value and closer to the value of 1, which means that the earthquake clusters in that cluster have a pretty good distance from each other and are 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 have implications for regional clusters and provide new information regarding spatial-temporal characteristics.

ACKNOWLEDGEMENT

We would like to express our gratitude to our colleagues for their valuable discussions, which played a significant role in contributing to this research.

REFERENCES

[1] T. Muliawati and D. G. Harbowo, “A Statistical review of the dates and patterns of volcanic activity of Lewotolo Volcano, East Nusa Tenggara, Indonesia,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1245, no. 1, 2023, doi: 10.1088/1755-1315/1245/1/012006.

[2] P. L. Whelley, C. G. Newhall, and K. E. Bradley, “The frequency of explosive volcanic eruptions in Southeast Asia,” *Bull. Volcanol.*, vol. 77, no. 1, 2015, doi:

10.1007/s00445-014-0893-8.
[3] R. Hall, "Southeast Asia: New Views of the Geology of the Malay Archipelago," *Annu. Rev. Earth Planet. Sci.*, vol. 45, pp. 331–358, 2017, doi: 10.1146/annurev-earth-063016-020633.
[4] N. Wilson, V. Valler, M. Cassidy, M. Boyd, L. Mani, and S. Brönnimann, "Impact of the Tambora volcanic eruption of 1815 on islands and relevance to future sunlight-blocking catastrophes," *Sci. Rep.*, vol. 13, no. 1, pp. 1–13, 2023, doi: 10.1038/s41598-023-30729-2.
[5] S. Murphy *et al.*, "Tsunamigenic earthquake simulations using experimentally derived friction laws," *Earth Planet. Sci. Lett.*, vol. 486, no. March, pp. 155–165, 2018, doi: 10.1016/j.epsl.2018.01.011.
[6] D. G. Harbowo and T. Muliawati, "State transition matrix and Markov-chain diagram for frequent volcanic eruptions: Krakatoa, Indonesia," *E3S Web Conf.*, vol. 479, pp. 1–12, 2024, doi: 10.1051/e3sconf/202447902005.
[7] T. Muliawati *et al.*, "k-Means Clustering to Enhance the Petrified Wood Composition Data Analyses and Its Interpretation," *Indones. J. Appl. Math.*, vol. 3, no. 1, p. 26, 2023, doi: 10.35472/indojam.v3i1.1288.
[8] O. G. Berg and P. H. von Hippel, "Selection of DNA binding sites by regulatory proteins," *Trends Biochem. Sci.*, vol. 13, no. 6, pp. 207–211, 1988, doi: 10.1016/0968-0004(88)90085-0.
[9] K. F. Afifah, F. Lestari, and F. Brigitta, "Estimation of Earthquake Insurance Premium By Considering Earthquake Probability and the Damage Ratio of Buildings in Banda Aceh and Sibolga City," *J. Sci. Appl. Technol.*, vol. 8, no. 1, p. 45, 2024, doi: 10.35472/jsat.v8i1.1719.
[10] S. R. Taylor, "Geochemical comparison of Australian and North American tektites," *Geochim. Cosmochim. Acta*, vol. 29, no. 2, pp. 65–70, 1965, doi: 10.1016/0016-7037(65)90117-1.
[11] T. Muliawati and H. Murfi, "Eigenspace-based fuzzy c-means for sensing trending topics in Twitter," *AIP Conf. Proc.*, vol. 1862, 2017, doi: 10.1063/1.4991244.
[12] T. M. Kodinariya and P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering," *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 1, no. 6, pp. 2321–7782, 2013.
[13] S. Paembonan and H. Abduh, "Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat," *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt_jiit.v6i2.659.
[14] P. Cunningham and S. J. Delany, "k-Nearest Neighbour Classifiers: 2nd Edition (with Python examples)," Apr. 2020, doi: 10.1145/3459665.
[15] S. Petrosino, P. Cusano, and P. Madonia, "Tidal and hydrological periodicities of seismicity reveal new risk scenarios at Campi Flegrei caldera," *Sci. Rep.*, vol. 8, no. 1, pp. 1–12, 2018, doi: 10.1038/s41598-018-31760-4.
[16] H. C. Natalia, D. G. Harbowo, and R. Ikhram, "Potensi Geodiversity di Sekitar Kawasan Anak Krakatau-Way Kambas, Provinsi Lampung, Sebagai Kandidat Geopark Indonesia," *J. Sci. Appl. Technol.*, vol. 5, no. 1, p. 47, 2021, doi: 10.35472/jsat.v5i1.318.