NAICE Journal of Next Applied Informatics in Computer Engineering
Vol. 1, No. 1, April 2026, pp. 1~7
ISSN: , DOI:

Spatial Clustering of Seismic Events in Bali and West Nusa Tenggara

I Made Andi Darma Kesuma[1], I Putu Yudi Prabhadika[2]
[1] Geomatics Engineering Master Program, Faculty of Engineering, Universitas Gadjah Mada
[2] Computer Engineering Study Programs, Faculty of Engineering and Planning, Universitas Warmadewa

Article Info

Article history:

Received February 13, 2026
Revised March 29, 2026
Accepted March 31, 2026

Keywords:

Clustering
DBSCAN
K-Means
Bali-NTB
Seismic

ABSTRACT

The precise delineation of seismic source zones is fundamental to seismic hazard assessment, particularly in complex tectonic settings like around Bali and West Nusa Tenggara (NTB). Traditional zoning methods often struggle to capture the irregular geometries of active faults, leading to potential inaccuracies in hazard estimation. This study presents a comparative analysis of partition-based (K-Means) and density-based (DBSCAN) clustering algorithms to identify active fault structures using earthquake data recorded from 2015 to 2025. The dataset, comprising standardized hypocentral coordinates, was processed to evaluate the ability of each algorithm to segregate active ruptures from background seismicity. Results indicate that K-Means (k=3) imposed artificial, spherical boundaries that fragmented continuous geological features and misclassified distant outliers. Conversely, DBSCAN, optimized with an Epsilon (eps) of 0.25 and Minimum Points (MinPts) of 5, successfully delineated the linear continuity of the Flores Back-Arc Thrust and isolated distinct volcanic earthquake swarms while effectively filtering noise. The study concludes that DBSCAN offers superior performance for seismotectonic mapping in the Bali-NTB region, providing a more geologically realistic framework for defining seismic hazard zones than traditional partitioning methods.

Corresponding Author:

I Putu Yudi Prabhadika
Computer Engineering Study Programs, Faculty of Engineering and Planning, Universitas Warmadewa
Email: yudipdika@warmadewa.ac.id

1. INTRODUCTION

Indonesia is a country with some of the highest seismic activity in the world due to its location at the meeting point of four major tectonic plates [1]. Indonesia also located in the “Ring of Fire” [2], as a result of the plate collision and the ring of fire, Indonesia is prone to disasters especially earthquakes [3], [4]. Sabtaji [5] stated that during the period from 2009 to 2019, there were 71,628 earthquake occurrences in Indonesia. This includes Bali and West Nusa Tenggara (NTB) region, which are areas with relatively high earthquake vulnerability [6]. Seismic hazard mitigation requires precise delineation of these active earthquake zones. Spatial clustering has proven to be a powerful tool and fundamental step for analyzing seismic events [7].

Despite the high seismicity, traditional seismic zoning often relies on manual delineation or grid-based binning, which can oversimplify complex fault geometries. Machine Learning (ML) has emerged as a powerful tool for performing automatic clustering, which is part of unsupervised learning [8]. The choice of algorithm used for clustering significantly dictates the resulting seismic zoning. K-means is one of the most popular clustering methods [9], [10], [11], [12]. This method has the advantage of fast clustering speed, but less effective for clustering non-spherical data [12]. This presents a critical problem in the Bali-NTB context, where fault lines are linear and not spherical. Misapplying an algorithm could lead to artificially partitioned fault zones, misrepresenting the true continuity of the hazard.

Recent studies have explored unsupervised learning for seismic analysis. Malik [13] demonstrated the utility of K-Means algorithm for earthquake distribution mapping in Indonesia, explaining its effectiveness for broad regional zoning. However, their approach does not consider the irregular geometry of the fault. On the other hand, Hafid [14] applied DBSCAN (Density-Based Spatial Clustering of Applications with Noise) to identify potential damage zones, highlighting its ability to handle noise and arbitrary shapes. Karavak [15] further utilized clustering for seismic micro zonation. However, there remains a gap in the literature specifically comparing these two algorithms within the complex fault of Bali-NTB. Most studies focus on one method in isolation without empirically demonstrating why density-based clustering is geologically superior for this specific tectonic setting.

This study proposes a comparative analysis of K-Means and DBSCAN algorithms to determine the optimal method for spatially clustering seismic data in the Bali-NTB region. The research utilizes coordinate data (Latitude, Longitude). The study specifically investigates the sensitivity of the DBSCAN epsilon parameter and minimum samples, identifying an optimal configuration that preserves fault continuity while filtering background noise, contrasting this with the rigid partitioning of K-Means.

The primary innovation of this research lies in the empirical demonstration of algorithm suitability for linear fault systems versus volcanic swarms. Unlike previous general studies, this research provides a tuned parameter set specifically optimized for the Bali-NTB region. The value of this research is a more accurate, automated framework for defining seismic source zones, which directly contributes to improved seismic hazard assessment (SHA) and disaster risk reduction strategies in the region.

2. METHOD

This study focuses on the seismically active region encompassing Bali and West Nusa Tenggara (NTB), defined by the coordinates of 8.0° S to 10.0° S and longitude 114.0° E to 119.0° E. This location was selected due to its complex tectonic setting, characterized by the convergence of the Indo-Australian Plate subduction in the south and the Flores Back-Arc Thrust in the north, which generates frequent and geometrically distinct seismic events. Earthquake point data is obtained from USGS Earthquakes, which provides data based on historical events. The dataset comprises earthquake events recorded from January 2015 to December 2025. Data consists of 22 attributes, which are filtered into latitude, longitude, depth, and magnitude attributes. To ensure spatial accuracy during clustering, the data were standardized using Z-score normalization to mitigate the disparity between latitudinal and longitudinal distances. Figure 1 shows how the data plotted and visualized.

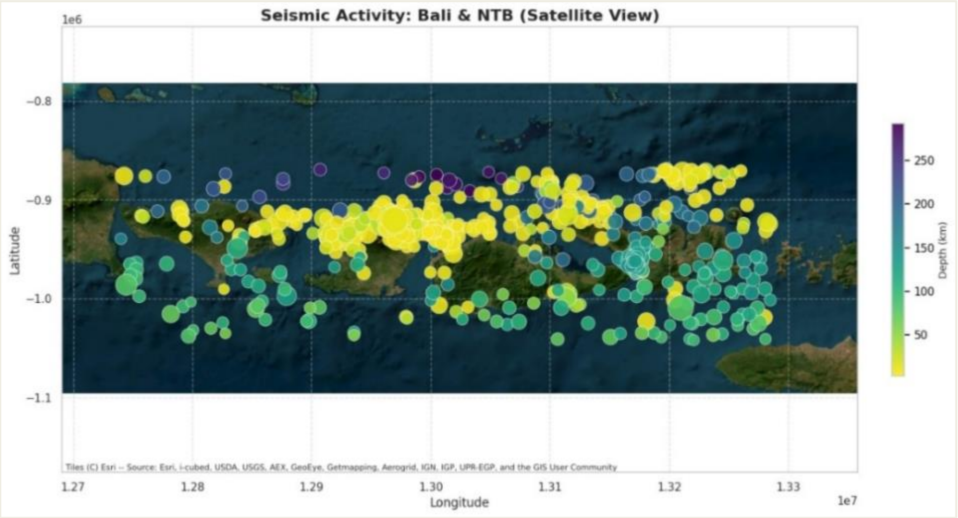

Figure 1. Seismic activity in Bali and NTB

To automate the identification of seismic source zones, this study utilizes unsupervised machine learning, a paradigm designed to discover hidden structures within unlabeled datasets. Specifically, spatial clustering algorithms are employed to partition the earthquake catalog into geologically meaningful groups based on hypocentral proximity [16]. In seismological applications, clustering serves as a critical data-driven technique to segregate active fault ruptures and aftershock sequences from random background seismicity. This research contrasts two fundamental clustering methodologies: partition-based clustering, which enforces a specific number of geometric zones, and density-based clustering, which defines clusters based on the concentration of events in the spatial domain [17].

The first clustering approach implemented was the K-Means algorithm, a simple partitional clustering algorithm that attempts to find **k** non-overlapping clusters that represented by centroids [18]. This algorithm operates by repeatedly assigning points (in this case, earthquake points) to the nearest center of mass to form "k" non-overlapping clusters, minimizing the sum of squares within the clusters [16]. For this study, the number of clusters (**k**) was determined to represent broad regional zones, partitioning the seismic data into distinct geometric territories. While effective for general categorization, K-Means assumes that clusters are spherical and of similar size, a theoretical limitation when applied to the linear and irregular fault structures found in the Bali-NTB region.

To address the geometric limitations of partition-based methods, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was applied. DBSCAN is a density-based, non-parametric algorithm designed to identify clusters and noise in spatial databases [19], [20]. Unlike K-Means, DBSCAN groups data points based on density connectivity, requiring two key parameters: the radius of the neighborhood (**Epsilon**) and the minimum number of points required to form a dense region (**MinPts**) [16]. This method is particularly suited for seismicity analysis as it can identify clusters of arbitrary shapes, such as elongated fault ruptures, and automatically classify sparse, isolated events as noise [17]. In this study, the Epsilon parameter was iteratively tuned to bridge significant fault segments without merging distinct, localized earthquake swarms. MinPts parameter is also tuned to ensure that clusters are formed properly.

The comparative analysis of the two algorithms was conducted by evaluating their ability to delineate known geological structures and handle background seismicity. The performance was assessed qualitatively by overlaying the resulting clusters onto the tectonic map of the Flores Back-Arc Thrust and the subduction zone. The primary criteria for success were the algorithm's capacity to maintain the continuity of linear fault ruptures, its sensitivity in distinguishing separate aftershock sequences (swarms) from the main fault, and its effectiveness in filtering out random background noise. This structural comparison determines which method offers a more geologically realistic representation of the seismic hazard sources in the region.

3. RESULTS AND DISCUSSION (10 PT)

3.1. Spatial Partitioning using K-Means Algorithm

The application of the K-Means algorithm with k = 3 resulted in a distinct partition of the Bali-NTB region into three geometrically contiguous zones: a western cluster covering Bali-Lombok, a northwest cluster around Pulau Mojo, and a southeast cluster around Bajo. As visualized in Figure 2, where the clustering results are superimposed onto a satellite map, the algorithm successfully minimized intra-cluster variance to create compact groupings. The result of this algorithm forces all points to be grouped into clusters, with no points remaining ungrouped or becoming noise.

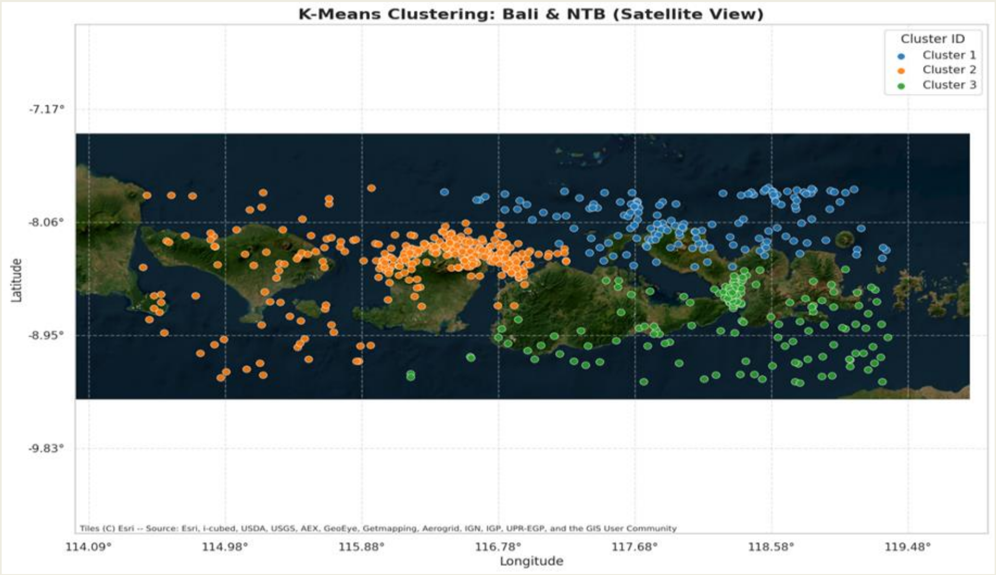

Figure 2. K-Means Clustering Result

However, the resulting boundaries exhibited significant discordance with the known tectonic setting. K-Means algorithm demonstrated high sensitivity to outliers; scattered background seismicity occurring far south of the subduction trench was forcibly assigned to the nearest cluster centroid (Cluster 3, Green). This "forced assignment" artificially expands the spatial footprint of the hazard zones, encompassing areas of low seismic density into high-density clusters.

3.2. Structural Identification using DBSCAN Algorithm

The implementation of the DBSCAN algorithm revealed a fundamentally different spatial organization of seismicity. Through iterative parameter tuning, the optimal structural delineation was achieved with an **Epsilon** (eps) of 0.25 (standardized distance) and a minimum points threshold (**MinPts**) of 5. As shown in Figure 3, this configuration identified approximately 10 distinct seismic clusters while classifying sparse events as noise.

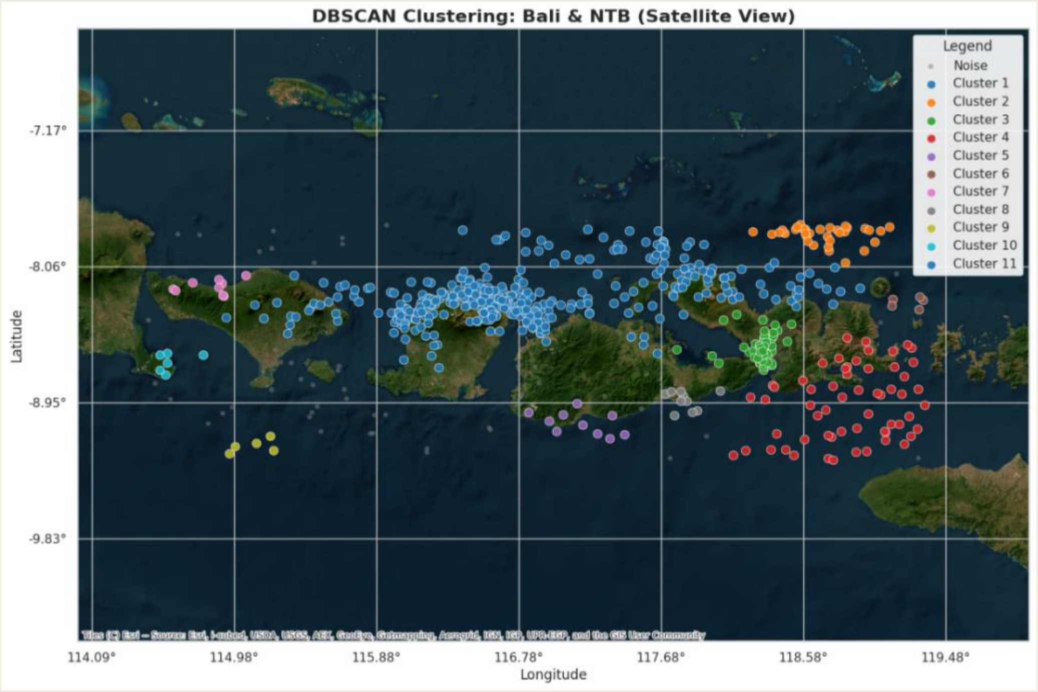

Figure 3. DBSCAN Clustering Result

The selection of the DBSCAN hyperparameters, specifically an Epsilon (eps) of 0.25 and a Minimum Points (MinPts) threshold of 5, was critical for balancing structural continuity with swarm preservation. Initial sensitivity testing with a stricter radius (eps=0.20) resulted in the artificial fragmentation into disjointed segments, failing to capture the fault's linear continuity. Conversely, increasing the radius beyond 0.25 risked merging the distinct volcanic swarms in the southeast into the main tectonic system, which would obscure the separation between different seismic sources. The chosen eps=0.25 (in standardized Euclidean distance) effectively bridged the minor spatial gaps characteristic of a continuous rupture zone while maintaining the separation of independent fault systems. Furthermore, the MinPts parameter was held at 5 to ensure that smaller, high-density clusters, typical of localized aftershock sequences or volcanic tremors, were preserved. Increasing this threshold (e.g., to MinPts=10) was observed to aggressively filter these geologically significant micro-clusters, reclassifying them as noise and resulting in a loss of valuable hazard data for localized risk assessment.

The most significant finding was the identification of Cluster 2 (Orange), a continuous, elongated structure running East-West along the northern variability of the islands. This cluster spatially correlates with the geometry of the Flores Back-Arc Thrust, a major reverse fault system responsible for significant uplift and seismic activity in the region. Unlike K-Means, DBSCAN preserved the continuity of this linear fault system without artificial fragmentation. Additionally, the algorithm successfully distinguished localized high-density swarms in the southeast from the primary rupture zone. These smaller clusters likely correspond to localized aftershock sequences or volcanic tectonic activity near Mount Tambora and Mount Rinjani.

3.3. Comparative Analysis

The comparison of the two algorithms highlights a critical trade-off between zonal simplification and structural fidelity. K-Means proved effective for broad administrative zoning, providing clear, easily interpretable boundaries suitable for general regional categorization (e.g., Province-level risk levels). However, it failed to capture the non-convex nature of the tectonic faults.

Conversely, DBSCAN demonstrated superior performance in seismotectonic mapping. By utilizing density reachability, it overcame the "spherical assumption" limitation of K-Means. The ability of DBSCAN to bridge the minor spatial gaps in the Flores Thrust while simultaneously isolating distinct earthquake swarms (like Clusters 2 and 3) aligns more closely with the geological reality. Furthermore, the noise-filtering capability of DBSCAN ensured that the statistical properties of the identified clusters were not polluted by random background seismicity, a feature that is critical for the accurate calculation of subsequent seismic parameters such as the b-value. Therefore, for the specific purpose of delineating active fault geometries in the complex Bali-NTB subduction and back-arc setting, DBSCAN provides a more scientifically robust result than K-Means.

4. CONCLUSIONS AND SUGGESTIONS

This study presented a comparative analysis of the K-Means and DBSCAN algorithms for the spatial clustering of seismic data in the geologically complex region of Bali and West Nusa Tenggara. By evaluating the performance of both partition-based and density-based approaches, the research aimed to determine the most effective method for automating the delineation of active fault zones and earthquake swarms.

The results demonstrate that while K-Means provides a computationally efficient method for broad regional zoning, it is fundamentally limited by its assumption of spherical cluster geometry. The algorithm forced continuous fault lines into arbitrary partitions and failed to distinguish significant tectonic structures from background noise, making it suitable primarily for administrative hazard categorization rather than detailed tectonic analysis.

In contrast, DBSCAN proved to be the superior algorithm for seismotectonic mapping in this region. With parameters optimized to an Epsilon (eps) of 0.25 and Minimum Points (MinPts) of 5, DBSCAN successfully identified the linear continuity of the Flores Back-Arc Thrust without artificial fragmentation. Furthermore, it demonstrated high sensitivity in isolating localized earthquake swarms, attributed to volcanic and aftershock activity, while effectively filtering out random background events as noise. Consequently, this study concludes that density-based clustering is the preferred methodological framework for defining precise seismic source geometries in the Bali-NTB region, providing a more rigorous basis for seismic hazard assessment than traditional partitioning methods.

Based on the findings of this study, future research should expand the scope of seismic clustering by incorporating focal depth as a weighted variable to enable 3D structural analysis, allowing for a clearer distinction between shallow crustal faults and deeper subduction zone events. Additionally, integrating temporal dimensions through Space-Time DBSCAN (ST-DBSCAN) is recommended to better distinguish transient aftershock sequences from stable background seismicity. It is also suggested that the specific fault geometries identified in this study be utilized as input source zones for Probabilistic Seismic Hazard Assessment (PSHA) to empirically validate whether density-based zoning yields more accurate Gutenberg-Richter b-value statistics compared to traditional grid-based models, while further comparative studies using hierarchical algorithms like OPTICS could address potential limitations related to varying cluster densities.

REFERENCES

[1] R. Munir, L. Nurahmah, V. Friesky, M. Imran, and A. Haryono, “Kajian Nilai B-Value untuk Menganalisis Aktivitas Seismik di Wilayah Pulau Jawa, Bali dan Nusa Tenggara Barat,” *GEOSAINS KUTAI BASIN*, vol. 7, no. 2, pp. 57–63, Oct. 2024, doi: 10.30872/GEOFISUNMUL.V7I2.1436.

[2] N. A. Pambudi, “Geothermal power generation in Indonesia, a country within the ring of fire: Current status, future development and policy,” *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 2893–2901, Jan. 2018, doi: 10.1016/J.RSER.2017.06.096.

[3] A. M. Utomo, A. Poniman, Y. Prihanto, and R. A. G. Gultom, “Utilization of Synthetic Aperture Radar (SAR) Imagery to Identify Potential for Earthquake Disasters to Improve Disaster Preparedness,” *IOP Conf Ser Earth Environ Sci*, vol. 950, no. 1, Jan. 2022, doi: 10.1088/1755-1315/950/1/012095.

[4] M. Muhajir and N. N. Sari, “K-Affinity Propagation (K-AP) and K-Means Clustering for Classification of Earthquakes in Indonesia,” *Proceeding - 2018 International Symposium on Advanced Intelligent Informatics: Revolutionize Intelligent Informatics Spectrum for Humanity, SAIN 2018*, pp. 6–10, Jul. 2018, doi: 10.1109/SAIN.2018.8673344.

[5] A. Sabtaji, “Statistik Kejadian Gempa Bumi Tektonik Tiap Provinsi di Wilayah Indonesia Selama 11 Tahun Pengamatan (2009-2019),” *BULETIN METEOROLOGI, KLIMATOLOGI, DAN GEOFISIKA*, vol. 1, no. 7, pp. 31–46, 2020.

[6] J. Nugraha, G. Pasau, B. Sunardi, and S. Widiyantoro, “ANALISIS HAZARD GEMPA DAN ISOSEISMAL UNTUK WILAYAH JAWA-BALI-NTB,” *Jurnal Meteorologi dan Geofisika*, vol. 15, no. 1, Apr. 2014, doi: 10.31172/JMG.V15I1.168.

[7] M. Sinambela, H. H. Arrizal, E. Darnila, Munawar, M. R. Nugraha, and A. Widodo, “Clustering Analysis of Earthquake Based on K-Means, DBSCAN, and Fuzzy C-Means in North Sumatra,” *Proceedings - 2024 International Conference on Information Technology and Computing, ICITCOM 2024*, pp. 242–247, 2024, doi: 10.1109/ICITCOM62788.2024.10762334.

[8] G. Fung, *A Comprehensive Overview of Basic Clustering Algorithms*. 2001.

[9] M. M. T. Chiang and B. Mirkin, “Intelligent choice of the number of clusters in k-means clustering: An experimental study with different cluster spreads,” *J Classif*, vol. 27, no. 1, pp. 3–40, Mar. 2010, doi: 10.1007/s00357-010-9049-5.

[10] I. K. Khan *et al.*, “Addressing limitations of the K-means clustering algorithm: outliers, non-spherical data, and optimal cluster selection,” *AIMS Mathematics*, vol. 9, no. 9, pp. 25070–25097, 2024, doi: 10.3934/math.20241222.

[11] Y. Su, “Improvement and Empirical Study of K-Means Clustering Algorithm Based on Chinese Retrieval,” *Advances in Intelligent Systems and Computing*, vol. 1017, pp. 182–188, 2020, doi: 10.1007/978-3-030-25128-4_24.

[12] H. He, Y. He, F. Wang, and W. Zhu, “Improved K-means algorithm for clustering non-spherical data,” *Expert Syst*, vol. 39, Nov. 2022, doi: https://doi.org/10.1111/exsy.13062.

[13] D. Malik, I. M. Artha Agastya, A. Y. Anjaya, Kusrini, and S. Hartantyo, “Earthquake Distribution Mapping in Indonesia Using K-Means Clustering Algorithm,” *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2024*, pp. 499–504, 2024, doi: 10.1109/ICITISEE63424.2024.10730706.

[14] H. Hafid, R. Hidayat, and R. H. S., “Implementation of DBSCAN for Earthquake Clustering in Indonesia with Potential Surface Damage,” *Journal of Mathematics, Computations and Statistics*, vol. 8, no. 1, pp. 124–133, Apr. 2025, doi: 10.35580/JMATHCOS.V8I1.7318.

[15] H. N. Karakavak, H. O. Cekim, G. Ozel, and S. Tekin, “Seismic Microzonation and Earthquake Forecasting in Western Anatolia Using K-Means Clustering and Entropy-Based Magnitude Volatility Analysis,” *Geotechnical and Geological Engineering 2025 43:2*, vol. 43, no. 2, pp. 99-, Jan. 2025, doi: 10.1007/S10706-024-03062-2.

[16] J. Han, J. Pei, and H. Tong, “Data Mining: Concepts and Techniques, Fourth Edition,” *Data Mining: Concepts and Techniques, Fourth Edition*, pp. 1–752, Jan. 2022, doi: 10.1016/C2013-0-18660-6.

[17] H. Hafid, R. Hidayat, and R. H. S., “Implementation of DBSCAN for Earthquake Clustering in Indonesia with Potential Surface Damage,” *Journal of Mathematics, Computations and Statistics*, vol. 8, no. 1, pp. 124–133, Apr. 2025, doi: 10.35580/JMATHCOS.V8I1.7318.

[18] J. Wu, *Advances in K-means Clustering: A Data Mining Thinking*. Springer Science & Business Media, 2012. Accessed: Jan. 06, 2026. [Online]. Available: https://books.google.co.id/books?id=pI2_F8SqWcQC&printsec=frontcover&hl=id#v=onepage&q&f=false

[19] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” 1996, Accessed: Jan. 08, 2026. [Online]. Available: www.aaai.org

[20] I. Taheri Sarteshnizi, M. Sarvi, S. A. Bagloee, and N. Nassir, “Temporal pattern mining of urban traffic volume data: a pairwise hybrid clustering method,” *Transportmetrica B*, vol. 11, no. 1, Dec. 2023, doi: https://doi.org/10.1080/21680566.2023.2185496.