Article Received 4 August 2023. Revised 17 September 2023.

Comparison of K-Means Algorithm and DBSCAN on Aftershock Activity in the Flores Sea: Seismic Activity 2019-2022

Anyela Aprianti, Adi Jufriansah*, Pujianti Bejahida Donuata, Azmi Khusnani, John Ayuba

***Abstract*—This study seeks to determine whether the clustering method can be used to analyze Flores Sea earthquake activity. The BMKG Repo is the source for real earthquake vibration data collection in this investigation. The stages of this research include preparing the data in CSV format and then preparing the data to eliminate useless data by identifying missing data. Based on the research data, it was determined that the K-Means and DBSCAN methods are used to determine the clustering method for analyzing earthquake activity. In addition, the data is depicted using a graphical Elbow method to determine the number of clusters of aftershocks in the Flores Sea. The results of the visualization of aftershocks that followed earthquakes in the Flores Sea between 2019 and 2022 revealed three distinct groups of earthquake source depths: 33 to 70 kilometres, 150 to 300 kilometres, and 500 to 800 kilometres. Regarding the silhouette index parameter, the K-Means algorithm is preferable to the DBSCAN algorithm when clustering results are used to analyze earthquake activity.**

***Index Terms*— cluster method, DBSCAN, earthquake, K-Means**

I. INTRODUCTION

SEISMIC activity is a complex natural phenomenon that frequently causes substantial devastation and social effects [1]. Aftershocks, or what are commonly referred to as aftershocks, are minor earthquakes that occur after the primary earthquake [2]. Aftershocks can cause fear and anxiety in the general public, particularly if they occur long after the primary earthquake [3]. As one of Indonesia's active tectonic plate connection zone regions, the Flores Sea is prone to earthquakes. Seismic events in this region can make it difficult for scientists and researchers to understand the characteristics of aftershocks and seismic activity patterns [4]. A suitable and efficient analytical method is required to analyze and comprehend the aftershock activity in the Flores Sea from 2019 to 2022. K-Means and DBSCAN are two clustering algorithms frequently employed in seismic analysis [5].

The K-Means algorithm is one of the most extensively used clustering algorithms and is utilized in numerous research fields, such as seismology [6], [7]. This algorithm functions by clustering data based on the shortest distance between each data point and the group's center [8]. K-Means can assist in identifying seismic patterns in aftershock activity in the Flores Sea, which can provide valuable insights into the characteristics and behavior of aftershocks in the region [1]. The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm is a clustering technique that emphasizes data density-based grouping [9]. DBSCAN can identify concentrated regions in the seismic distribution and isolate noise or improperly classified data [10], [11]. This makes this algorithm attractive for analyzing the Flores Sea's complex seismic activity.

In this study, we will compare the performance of the K-Means and DBSCAN algorithms in classifying aftershock activity data from 2019 to 2022 in the Flores Sea. This comparative analysis will enhance our understanding of the characteristics of aftershocks in this region and contribute to efforts to reduce the likelihood of future earthquakes. Through a greater comprehension of seismic activity patterns in the Flores Sea, it is anticipated that this study's results will significantly contribute to seismology and our knowledge of this complex natural phenomenon. In addition, the findings of this study can provide authorities and related parties with valuable information for disaster mitigation and the development of earthquake-resistant infrastructure in the region.

II. METHOD

We collected seismic data from the Flores Sea between 2019 and 2022 and preprocessed and normalized the data for this investigation. The K-Means and DBSCAN algorithms were then implemented to classify aftershock activity in the region [11]–[13]. The evaluation was conducted with the clustering evaluation metric, and the outcomes of the two algorithms were contrasted and analyzed [14], [15]. This study's findings shed light on the patterns and characteristics of aftershocks in the Flores Sea and assess the benefits and drawbacks of each algorithm. These findings can significantly contribute to seismology and the development of disaster mitigation in this region, which is prone to earthquakes.

A. Data Collection

Flores Sea seismic data from 2019 to 2022 are acquired from trusted seismological institutions, such as the Indonesian Meteorology, Climatology, and Geophysics Agency (BMKG) and IRIS DMC. The data must include the earthquake's time, location (latitude and longitude), depth, and magnitude of aftershocks.

B. Preprocessing of Data

Collected seismic data must be preprocessed to guarantee its quality and purity. This includes handling missing data, outliers, and duplicates and conducting data transformations as required.

C. Data Normalization

As the clustering algorithm is sensitive to data scale, it is necessary to normalize the data to equalize the range of variables.

D. Parameter Selection

Define The number of clusters (k) is the most important parameter to select in the K-Means algorithm. In the DBSCAN algorithm, the epsilon distance () and the minimum number of points (MinPts) within a radius of are essential parameters for identifying clusters. Testing each algorithm's parameters with various values may be necessary to conduct a fair comparison.

E. K-Means Algorithm Implementation

Define As the clustering algorithm is sensitive to data scale, it is necessary to normalize the data to equalize the range of variables.

F. DBSCAN Algorithm Implementation

The DBSCAN algorithm uses pre-processed seismic data and determines clusters based on the data point density.

G. Outcome Evaluation

The results of both algorithms are evaluated using a clustering evaluation metric such as the Davies-Bouldin index, Dunn index, or another validity index. This metric aids in assessing the quality of the clusters generated by each algorithm.

H. Analysis and Comparison of Results

The results of both the K-Means and DBSCAN algorithms are analyzed and compared to determine the performance of each algorithm in classifying aftershock activity in the Flores Sea. The comparison results provide insight into this region's patterns and characteristics of aftershocks.

I. Interpretation of Results

The findings from the comparative analysis are interpreted to determine the advantages and disadvantages of each algorithm in the analysis of aftershock activity in the Flores Sea. This conclusion is beneficial for shedding additional light on seismology science and disaster mitigation policies.

J. Conclusion

In the conclusion section, summarize the comparison's results and the most significant findings of this study. Please indicate whether the K-Means or DBSCAN algorithm is more appropriate or effective for analyzing aftershock activity in the Flores Sea, and provide recommendations for their use in future research and development.

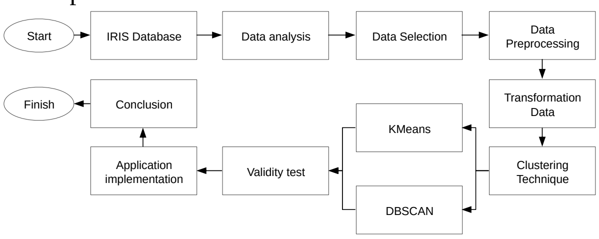

Fig. 1. Research method

III. Result and Discussion

Indonesia's Flores Sea has a complex geological history that generates significant seismic activity [16]. The Pacific Ring of Fire, a fault line that is both seismically and volcanically active, borders this ocean [4]. The Indo-Australian, Eurasian, and Pacific Plate interact and induce earthquakes in this region [5], [17]. Seismicity around the Flores Sea is also a result of subduction activity beneath the Indo-Australian tectonic plate, which is descending beneath the Sunda Plate [18]–[20]. In this region, earthquakes can potentially cause tsunamis, endangering infrastructure and coastal communities. Understanding the background of this seismicity is crucial for raising awareness of potential threats and enhancing mitigation efforts to safeguard Flores Sea residents and ecosystems. From 2019 to 2022, 1,445 seismic data records were obtained from the BMKG database for this investigation. The first five earthquake distribution catalogue data and the last five earthquake distribution catalogue data.

	Date	Time	Latitude	Longitude	Depth	Magnitude
0	12/31/2019	23:03:33.474	5.77	104.99	83	3.3
1	12/31/2019	22:13:21.681	1.62	126.37	10	3.9
2	12/31/2019	21:33:50.385	3.42	99.62	35	4.0
3	12/31/2019	21:25:05.256	10.15	115.94	10	3.7
4	12/31/2019	18:23:13.053	8.09	107.60	18	2.8

Fig. 2. The first five earthquake distribution catalogue data

	Date	Time	Latitude	Longitude	Depth	Magnitude
11655	1/1/2019	03:56:53.187	8.25	114.89	10	2.3
11656	1/1/2019	03:19:36.356	7.77	119.06	10	3.4
11657	1/1/2019	03:00:42.736	5.75	126.93	69	4.7
11658	1/1/2019	02:55:54.190	6.65	105.19	10	3.1
11659	1/1/2019	00:58:45.213	0.24	123.13	69	3.8

Fig. 3. The last five earthquake distribution catalogue data

.

Before commencing earthquake data analysis, it is essential to perform the required data processing. This method begins with collecting data from the first to the last five earthquakes. The following phase was feature selection, in which the coordinates for latitude, longitude, earthquake magnitude, and depth of earthquake occurrence were chosen as essential information for the subsequent study. These data facilitate a more specialised study and play a crucial role in earthquake modelling.

Before conducting any additional analysis, remember that the data collected may be incomplete or error-free. Consequently, data cleansing, or data cleaning utilising imputation techniques, is the next stage in this process. The imputation method approximates missing values [21], [22]. This method seeks to reduce the potential negative effects of missing data on the effectiveness of machine learning models. Figure 4 illustrates imputation in parameter estimation of the data distribution, which is still pertinent to new tests.

By optimising the likelihood or log-likelihood function, the least squares alternative is the imputation method used to analyse earthquake data [5]. This method permits replacing absent values with a more precise estimate based on the probability that the value will occur. By employing this technique, it is possible to maximise the amount of information gleaned from the available data while minimising the danger of bias or analytical distortion. As a result, the conclusions reached after analysing the earthquake data will be more reliable and can be utilised to assist policymakers in coping more effectively with future earthquake hazards.

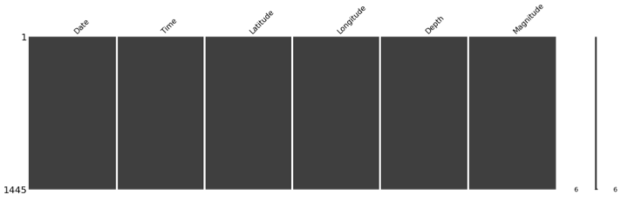

Fig. 4. Data cleaning results

In data analysis that attempts to identify a relationship or relationships between two or more variables, calculating the correlation between attributes is an essential step. To uncover latent patterns and trends in the data, correlation examines the degree to which two variables move in tandem. Statistical methods such as Pearson or Spearman correlation coefficients are typically employed in this process, depending on the type of data encountered [23], [24]. Analysts can enhance their decisions and gain a deeper understanding of the relationships between these factors by examining the strength of the correlation between features.

Typically, several procedures are employed when determining the relationship between qualities. The data must first be properly cleaned and processed to overcome outliers and eliminate missing data, which could skew the correlation's results. In addition, the correlation calculation is performed using a technique appropriate for the data type provided. Pearson's correlation coefficient can be used if the data are normally distributed; however, Spearman's correlation coefficient is preferred if the data are not normally distributed [23], [24]. After calculating the correlation, the data are analysed to determine the strength of the relationship between the characteristics, whether positive (both attributes increase at the same rate) or negative (one attribute increases while the other decreases). The findings of this connection serve as the basis for making decisions during data analysis and can be used to identify significant variables that affect a system.

The phase of establishing the correlation between variables also affects the validity and quality of data analysis. A high correlation between two variables implies a strong association between them, and one can use this relationship as a guide when developing prediction models or making decisions based on these variables. Remember that correlation does not imply causation, as the close relationship between two variables does not necessarily indicate that one variable caused the other to change. As a result, careful consideration and a thorough understanding of the investigated data must always go hand in hand with correlation analysis. Correctly concluding the phases of calculating the correlation between attributes enables analysts to comprehend the complexity of the data better and make significant contributions to spatial data analysis. The correlation between attributes is presented in Figure 5.

Fig. 5. The correlation between attributes

The connection depicted in Figure 5 has a maximum value of 0.1. This demonstrates that the correlation criterion is insufficient, necessitating normalisation of the data. Normalisation is a crucial stage in the data analysis for eliminating or reducing imbalances and scale differences between features. Extreme data values may affect the efficacy of the analytical model, particularly when distance-based or optimisation techniques are employed. Normalisation modifies attribute values so that all data has an analogous scale or a similar distribution. This normalisation procedure can increase the efficacy and precision of data analysis, allowing us to identify more representative patterns, trends, or correlations.

Numerous normalisation methods, such as Min-Max Scaling, Z-Score Normalisation, and Robust Scaling, are frequently employed in data analysis. The attribute values are transformed using min-max scaling into the range [0, 1], where the minimum value is 0 and maximum values are 1. Standardisation, an alternative term for Z-Score Normalisation, modifies the data with a mean of 0 and a standard deviation of 1. This method scales the data quite equitably while preserving the shape of the distribution. Using the median and interquartile range, Robust Scaling reduces the impact of anomalies during normalisation. The appropriate normalisation technique should be selected depending on the characteristics of the used data and the research objectives [25]–[27].

Normalisation is also crucial to the machine learning process. Because most machine learning algorithms are particularly sensitive to variations in attribute scales, some characteristics may dominate the learning process without normalisation, resulting in models that are biassed towards these attributes. Normalisation enables machine learning models to recognise more intricate and diverse patterns in standardised data [28], [29]. In addition, normalisation reduces the likelihood of overfitting and accelerates the convergence of learning algorithms. Normalisation is a crucial step that must be considered in the data analysis and machine learning processes [30], [31]. By properly implementing normalisation, we can extract deeper insights and comprehension from the data and improve the efficacy of subsequent analysis and predictive models. Figure 6 depicts a three-dimensional visualisation of latitude, longitude, magnitude, and depth data. As can be seen, the distribution of earthquake data is separated into three colour categories in this visualisation.

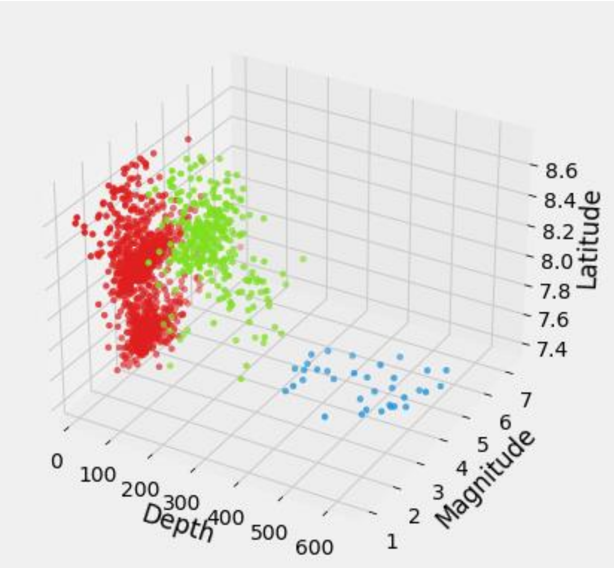

Fig. 6. A three-dimensional visualisation of latitude, seismic magnitude, and depth data

The results of the identified distribution of earthquakes are then determined using random initial centroid values. This is beneficial for calculating the distance distribution matrix so that it can be carried on to the object grouping stage and used to determine cluster members based on the shortest distance from the centroid. In addition, recurrent literacy of the data is performed to generate new, improved centroid distances, as depicted in Figure 7.

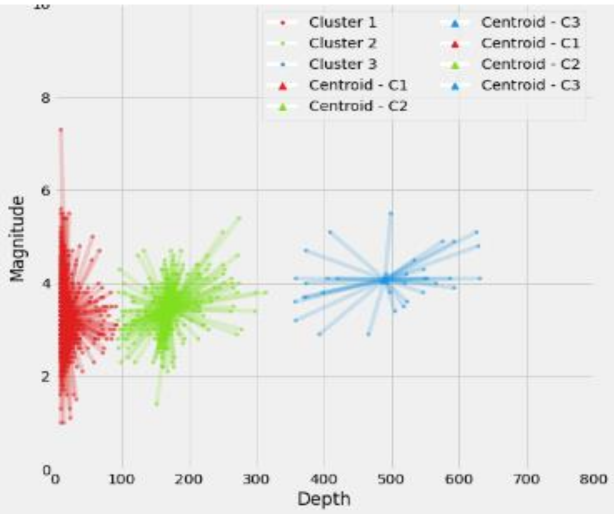

Fig. 7. Centroid point

Figure 7 shows that cluster 1 earthquake activity in Indonesia from 2019 to 2022 is more prevalent at depths between 0 and 90 kilometres. In contrast, cluster 2 earthquake activity is more prevalent at depths between 90 and 300 kilometres. In cluster 3, the depth range from 300 to 700 kilometres had a lower frequency. Comparatively, the DBSCAN algorithm (Figure 8) yielded only two cluster values.

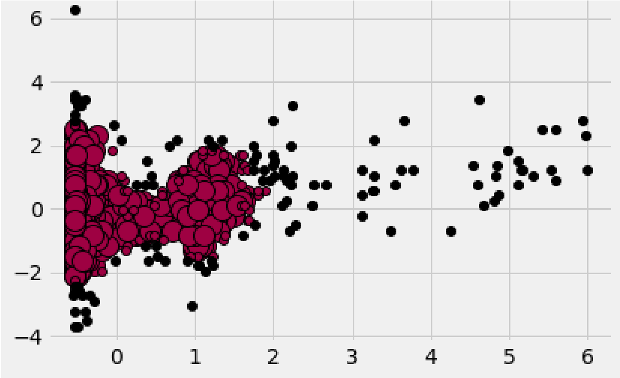

Fig. 7. A three-dimensional visualisation of latitude, seismic magnitude, and depth data

First cluster with position data of latitude 4.2131346, longitude 121.40729405, depth 18.64426488, and magnitude 3.40164563. Latitude position data for the second cluster is 4.73856731, longitude data is 121.82910727, profundity is 146.01655868, and magnitude data is 3.85773938. Latitude position data is 5.7792682, longitude data is 123.2396748, data depth is 461.05691057, and magnitude data is 4.42439024.

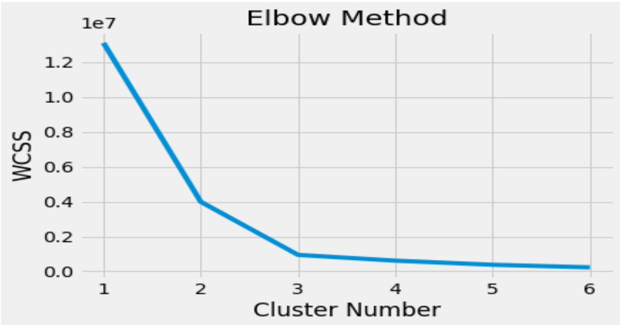

Fig. 8. A three-dimensional visualisation of latitude, seismic magnitude, and depth data

The number of clusters derived using the Elbow method based on the data fracture in Figure 8 is k = 3, or there are three clusters; this number is the result of forming optimal clusters for earthquake distribution data in 2019. Figure 10 depicts the Aftershock visualisation from 2019 to 2022 in its entirety.

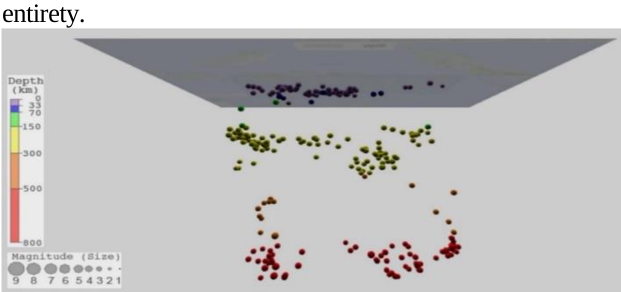

Fig. 10. 3D Aftershock in the Flores Sea

Studying aftershock activity in the Flores Sea from 2019 to 2022 and comparing the K-means and DBSCAN algorithms has substantial implications for seismic comprehension and mitigating the risk of natural disasters. K-Means and DBSCAN are two clustering techniques with different approaches to data processing. In contrast to DBSCAN, which clusters data based on data density, K-Means clusters data based on the distance to the cluster centre, or centroid [1], [5]. The comparison's findings cast light on each system's advantages and disadvantages for identifying trends and connections in aftershock activity.

According to the study's findings, the two algorithms K-Means and DBSCAN provide insightful information regarding aftershock activity in the Flores Sea. Both techniques yield relevant aftershock clusters and provide distinct perspectives on the region's seismic activity. K-Means clusters are primarily based on the distance to the cluster centre, whereas DBSCAN clusters are based on spatial density and detect clusters that do not have a normal geometric shape.

This study also includes a 3D visualisation of the data, which provides a more in-depth view of the aftershock distribution. The understanding of the geographical pattern and profundity of aftershocks is enhanced by three-dimensional visualisation. These findings provide seismologists and other relevant parties with vital information they can use to control and mitigate risks in affected areas.

This study determines the optimal number of clusters using a graphical elbow approach. This method can determine the number of aftershock clusters that best suit the data [32]–[34]. This method is employed in this study to contribute to the body of knowledge concerning the structure and clustering patterns that are valuable and reliable for analysing aftershock activity.

This study substantially impacts the study of natural hazards and seismology as a whole. By comparing the K-means and DBSCAN algorithms for analysing aftershock activity in the Flores Sea, we can understand the characteristics and distribution of aftershocks. These discoveries can aid in earthquake modelling, disaster mitigation planning, and developing risk-reduction measures for aftershocks. By gaining a deeper understanding of seismic activity, we can enhance community preparedness and reduce the impact of potential natural disasters.

V. Conclusion

The clustering method is useful for analysing earthquake activity because it can be used to organise earthquakes with similar characteristics into distinct clusters or groups. Popular clustering techniques like K-Means and DBSCAN were utilised in this study. K-Means is a method for clustering data based on its proximity to a particular cluster centre. DBSCAN, which stands for Density-Based Spatial Clustering of Applications with Noise, clusters data based on the data density encircling the data points. We intend to identify patterns and relationships between earthquakes in the Flores Sea region using these two methodologies.

The Flores Sea region aftershock earthquake clustering results from 2019 to 2022 are then visualised in three dimensions. According to the statistics, three notable groups or concentrations of earthquake source depths exist. The first cluster has a depth of 33 to 70 kilometres; the second cluster has a depth of 150 to 300 kilometres; and the third cluster has a depth of 500 to 800 kilometres. Through 3D visualisation, we can better comprehend the pattern of these earthquakes' depth distribution, thereby gaining valuable insights into the region's seismic activity.

We use the elbow method with 3D visualisation to determine the optimal number of Flores Sea aftershock data clusters. The elbow method graph is a technique for determining the optimal number of clusters based on the rate at which cluster variance decreases as the number of clusters increases. Using the elbow method chart results, we can determine the optimal number of earthquake aftershock clusters and gain a greater understanding of the structure and pattern of data clustering. We intend to determine the most relevant and valuable aftershock clusters for seismic activity research in the Flores marine region by combining 3D visualisation and elbow technique analysis.

An analysis of the aftershock clustering data in the Flores Sea region will greatly assist in comprehending the seismic risk in the area. By identifying the depth groups of the earthquake sources, we can assess the potential hazards that may result from each cluster. Additionally, this knowledge may aid in developing earthquake response and disaster mitigation plans. This research's clustering techniques and data analysis significantly impact seismology and the study of natural disasters in general. By understanding the patterns and characteristics of earthquakes in the Flores Sea, we can enhance community preparedness and resilience in regions prone to earthquakes and other seismic catastrophes.

References

[1] A. Jufriansah *et al.*, "Estimation of Flores Sea Aftershock Rupture Data Based on AI," *Indones. Rev. Phys*, vol. 6, no. 1, pp. 46–56, 2023, doi: 10.12928/irip.v6i1.6705.

[2] A. Jufriansah, A. Khusnani, Y. Pramudya, and M. Afriyanto, "Comparison of aftershock behavior of the flores sea 12 december 1992 and 14 december 2021," *Journal of Physics: Theories and Applications*, vol. 7, no. 1, pp. 65-74., 2023.

[3] Y. Liu, J. Zhuang, and C. Jiang, "Background Seismicity before and after the 1976 Ms 7.8 Tangshan Earthquake: Is Its Aftershock Sequence Still Continuing?," *Seismological Research Letters*, vol. 92, no. 2A, pp. 877–885, Mar. 2021, doi: 10.1785/0220200179.

[4] A. Khusnani, A. Jufriansah, and M. Afriyanto, "Utilization of Seismic Data as a Tsunami Vulnerability Review," *Indonesian Review of Physics*, vol. 5, no. 2, pp. 66–72, 2022.

[5] A. Jufriansah, Y. Pramudya, A. Khusnani, and S. Saputra, "Analysis of Earthquake Activity in Indonesia by Clustering Method," *Journal of Physics: Theories and Applications*, vol. 5, no. 2, p. 92, Sep. 2021, doi: 10.20961/jphystheor-appl.v5i2.59133.

[6] H. S. Kuyuk, E. Yildirim, E. Dogan, and G. Horasan, "Application of k-means and Gaussian mixture model for classification of seismic activities in Istanbul," *Nonlinear Process Geophys*, vol. 19, no. 4, pp. 411–419, Aug. 2012, doi: 10.5194/npg-19-411-2012.

[7] Z. Sheikhhosseini, N. Mirzaei, R. Heidari, and H. Monkaresi, "Delineation of potential seismic sources using weighted K-means cluster analysis and particle swarm optimization (PSO)," *Acta Geophysica*, vol. 69, no. 6, pp. 2161–2172, Dec. 2021, doi: 10.1007/s11600-021-00683-6.

[8] S. J. Nanda and G. Panda, "A survey on nature inspired metaheuristic algorithms for partitional clustering," *Swarm Evol Comput*, vol. 16, pp. 1–18, Jun. 2014, doi: 10.1016/j.swevo.2013.11.003.

[9] A. AlAli and F. Anifowose, "Seismic velocity modeling in the digital transformation era: a review of the role of machine learning," *J Pet Explor Prod Technol*, vol. 12, no. 1, pp. 21–34, Jan. 2022, doi: 10.1007/s13202-021-01304-0.

[10] A. SHİRAZY, A. HEZARKHANİ, A. SHİRAZİ, S. KHAKMARDAN, and R. ROOKİ, "K-Means Clustering and General Regression Neural Network Methods for Copper Mineralization probability in Char-Farsakh, Iran," *Türkiye Jeoloji Bülteni / Geological Bulletin of Turkey*, Dec. 2021, doi: 10.25288/tjb.1010636.

[11] A. E. Ezugwu *et al.*, "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Eng Appl Artif Intell*, vol. 110, p. 104743, Apr. 2022, doi: 10.1016/j.engappai.2022.104743.

[12] F. Ramdani, O. Kettani, and B. Tadili, "Evidence for subduction beneath Gibraltar Arc and Andean regions from k-means earthquake centroids," *J Seismol*, vol. 19, no. 1, pp. 41–53, Jan. 2015, doi: 10.1007/s10950-014-9449-9.

[13] R. K. Vijay and S. J. Nanda, "A Variable ε-DBSCAN Algorithm for Declustering Earthquake Catalogs," 2019, pp. 639–651. doi: 10.1007/978-981-13-1592-3_50.

[14] E. Amigó, J. Gonzalo, J. Artiles, and F. Verdejo, "A comparison of extrinsic clustering evaluation metrics based on formal constraints," *Inf Retr Boston*, vol. 12, no. 4, pp. 461–486, Aug. 2009, doi: 10.1007/s10791-008-9066-8.

[15] A. Fahad *et al.*, "A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis," *IEEE Trans Emerg Top Comput*, vol. 2, no. 3, pp. 267–279, Sep. 2014, doi: 10.1109/TETC.2014.2330519.

[16] R. Maneno, B. J. Sentosa, and G. Rachman, "Relocation Of Earthquake Hypocenter In The Flores Region Using Hypo71," *IPTEK The Journal of Engineering*, vol. 5, no. 2, May 2019, doi: 10.12962/j23378557.v5i2.a5024.

[17] J. Beckers and T. Lay, "Very broadband seismic analysis of the 1992 Flores, Indonesia, earthquake ($M_w = 7.9$)," *J Geophys Res Solid Earth*, vol. 100, no. B9, pp. 18179–18193, Sep. 1995, doi: 10.1029/95JB01689.

[18] I. R. Pranantyo and P. R. Cummins, "Multi-Data-Type Source Estimation for the 1992 Flores Earthquake and Tsunami," *Pure Appl Geophys*, vol. 176, no. 7, pp. 2969–2983, 2019, doi: 10.1007/s00024-018-2078-4.

[19] I. R. Pranantyo, M. Heidarzadeh, and P. R. Cummins, "Complex tsunami hazards in eastern Indonesia from seismic and non-seismic sources: Deterministic modelling based on historical and modern data," *Geosci Lett*, vol. 8, no. 1, 2021, doi: 10.1186/s40562-021-00190-y.

[20] I. R. Pranantyo, M. Heidarzadeh, and P. R. Cummins, "Complex tsunami hazards in eastern Indonesia from seismic and non-seismic sources: Deterministic modelling based on historical and modern data," *Geosci Lett*, vol. 8, no. 1, p. 20, Dec. 2021, doi: 10.1186/s40562-021-00190-y.

[21] F. Li, H. Sun, Y. Gu, and G. Yu, "A Noise-Aware Multiple Imputation Algorithm for Missing Data," *Mathematics*, vol. 11, no. 1, p. 73, Dec. 2022, doi: 10.3390/math11010073.

[22] M. Pazhoohesh, M. S. Javadi, M. Gheisari, S. Aziz, and R. Villa, "Dealing with Missing Data in the Smart Buildings using Innovative Imputation Techniques," in *IECON 2021 – 47th Annual Conference of the IEEE Industrial Electronics Society*, IEEE, Oct. 2021, pp. 1–7. doi: 10.1109/IECON48115.2021.9612650.

[23] J. C. F. de Winter, S. D. Gosling, and J. Potter, "Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data.," *Psychol Methods*, vol. 21, no. 3, pp. 273–290, Sep. 2016, doi: 10.1037/met0000079.

[24] M.-T. Puth, M. Neuhäuser, and G. D. Ruxton, "Effective use of Pearson's product–moment correlation coefficient," *Anim Behav*, vol. 93, pp. 183–189, Jul. 2014, doi: 10.1016/j.anbehav.2014.05.003.

[25] S. Panigrahi and H. S. Behera, "Effect of Normalization Techniques on Univariate Time Series Forecasting using Evolutionary Higher Order Neural Network," 2013.

[26] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl Soft Comput*, vol. 97, p. 105524, Dec. 2020, doi: 10.1016/j.asoc.2019.105524.

[27] S. Kim, Y. Noh, Y.-J. Kang, S. Park, J.-W. Lee, and S.-W. Chin, "Hybrid data-scaling method for fault classification of compressors," *Measurement*, vol. 201, p. 111619, Sep. 2022, doi: 10.1016/j.measurement.2022.111619.

[28] Z. Li, "A Generic Model of Global Earthquake Rupture Characteristics Revealed by Machine Learning," *Geophys Res Lett*, vol. 49, no. 8, Apr. 2022, doi: 10.1029/2021GL096464.

[29] L. Seydoux, R. Balestriero, P. Poli, M. de Hoop, M. Campillo, and R. Baraniuk, "Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning," *Nat Commun*, vol. 11, no. 1, p. 3972, Aug. 2020, doi: 10.1038/s41467-020-17841-x.

[30] K. J. Bergen, T. Chen, and Z. Li, "Preface to the Focus Section on Machine Learning in Seismology," *Seismological Research Letters*, vol. 90, no. 2A, pp. 477–480, Mar. 2019, doi: 10.1785/0220190018.

[31] S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza, "Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking," *Nat Commun*, vol. 11, no. 1, p. 3952, Aug. 2020, doi: 10.1038/s41467-020-17591-w.

[32] Z. Fan and X. Xu, "Application and visualization of typical clustering algorithms in seismic data analysis," *Procedia Comput Sci*, vol. 151, pp. 171–178, 2019, doi: 10.1016/j.procs.2019.04.026.

[33] B. Pokharel, O. F. Althuwaynee, A. Aydda, S.-W. Kim, S. Lim, and H.-J. Park, "Spatial clustering and modelling for landslide susceptibility mapping in the north of the Kathmandu Valley, Nepal," *Landslides*, vol. 18, no. 4, pp. 1403–1419, Apr. 2021, doi: 10.1007/s10346-020-01558-5.

[34] M. S. Boettcher, "Earthquake scaling relations for mid-ocean ridge transform faults," *J Geophys Res*, vol. 109, no. B12, p. B12302, 2004, doi: 10.1029/2004JB003110.

Anyela Aprianti is a student in the Physics Education Study Program at IKIP Muhammadiyah Maumere. She is joining the computational physics research group. (email: anyelaaprianti@gmail.com).

Adi Jufriansah is a lecturer at the Physics Education Study Program, IKIP Muhammadiyah Maumere, Indonesia. His area of expertise is in artificial intelligence. He has many publications in various reputable journals. Currently, he is focusing on research on artificial intelligence for earthquakes (email: saompu@gmail.com).

Pujianti Bejahida Donuata, is senior lecturer in Physics Education Study Program In IKIP Muhammadiyah Maumere, Indonesia. She received a Bachelor's in Education from Universitas Nusa Cendana in 2011 and a Master's Degree (Physics) from Universitas Brawijaya in 2014. She focuses on Quantum Teaching, Nuclear Physics, Medical Physics, and Biophysics research areas (email: pujinuna@gmail.com).

Azmi Khusnani is a lecturer at the Physics Education Study Program, IKIP Muhammadiyah Maumere, Indonesia. Her current research focus is earthquakes and disaster mitigation. He also has many publications in Scopus and accredited national journals. (email: husnaniazmi@gmail.com).

John Ayuba is a dedicated professional in the field of Science Laboratory Technology. He is affiliated with the Department of Science Laboratory Technology at the Ganye College of Agriculture in Adamawa State, Nigeria. (email: johnbuba5580@gmail.com).