p-ISSN 2355-5785
e-ISSN 2550-0325
JPF (Jurnal Pendidikan Fisika)
Universitas Islam Negeri Alauddin Makassar

SEISMIC ACTIVITY ANALYSIS IN INDONESIA: INTEGRATING MACHINE LEARNING, GEOSPATIAL DATA, AND ENVIRONMENTAL FACTORS FOR RISK ASSESSMENT

Ayu Aprilia[1*], Alka Budi Wahidin[2], Syafriadi[3]

[1]Physics Department, University of Lampung
[2]Atmospheric and Planetary Science Department, Institute Technology of Sumatra
[3]Physics Department, University of Lampung

Corresponding Address: ayu.aprilia@fmipa.unila.ac.id

Info Artikel

Riwayat artikel

Dikirim: 2025-01-03
Direvisi : 2025-05-14
Diterima: 2025-05-19
(Palatino Linotype 9)

Keywords:

Earthquake Data Analysis
Seismic Hazard Mapping
Principal Component Analysis (PCA)
K-Means Clustering

DOI:
10.24252/jpf.v13i1.54081

ABSTRACT

Earthquake's phenomena are critical for understanding Earth's interior, tectonic processes, and disaster preparedness. Because of Indonesia location in the Pacific Ring of Fire, it's suffering from regular seismic activities which result in huge annual losses. This study investigates the earthquake data from 1992 to 2024 by applying clustering techniques such as K-means and geodata visualization. By integrating physics, geospatial analysis, and machine learning, the study processes earthquake data to calculate energy release and analyze spatial-temporal patterns. Principal Component Analysis (PCA) is applied to reduce data dimensionality, while K-Means clustering identifies seismic patterns based on magnitude, depth, and energy. Visual tools, including correlation heatmaps and spatial maps, are used to present findings that support earthquake risk management in Indonesia. The results reveal temporal patterns in earthquake activity, with peaks observed in 2004–2007, associated with significant seismic energy release. Spatial analysis highlights high energy concentrations in megathrust zones. PCA and K-Means clustering identify three distinct clusters with varying correlations between seismic and atmospheric variables, indicating the influence of thermal and tectonic factors. These insights contribute to seismic hazard mapping, risk reduction strategies, and the improvement of earthquake prediction models. Future research should extend datasets and refine machine learning techniques to achieve more accurate predictions.

INTRODUCTION

Earthquakes are a very important phenomenon to investigate and understand because of their highly destructive potential resulting in great loss of property and lives [1]. Earthquakes are indeed a very interesting area to study as one can learn a lot about the interior of the earth and the source mechanism of seismic activity. Seismic history illustrated in the form of seismograms is of great value since it allows long term research of the dynamics of development and cycles of important geophysical events, earthquakes [2]. The study of seismic events may also appreciate the correlation between movements of the earth's crust which expands the understanding of tectonics [3]. This knowledge is Critical for communities affected by earthquakes in terms of town planning and building construction [1]. In addition, studies associated with earthquakes can help further understand the mechanics of earthquakes and their possible precursors [3]. Many-sided the nature of the study of earthquakes – seismology, seismotectonic, and earthquake geology as well as the ways of their measuring demonstrate their relevance in many branches and applied sciences.

Earthquakes release elastic energy stored in the Earth's crust due to tectonic plate movements. The energy released can be calculated using various methods, including seismic moment and magnitude scales [4], [5]. The elastic rebound model provides a physical basis for understanding earthquakes, relating fault dimensions, dislocation, and propagation velocity to seismic energy release [6]. For large earthquakes, the elastic energy released is proportional to the excess pressure in the magma chamber and the volume decrease during eruption [7]. The spatial distribution of earthquake energy release can provide insights into the state of stress and strength of the lithosphere [8]. Most seismic energy is radiated at frequencies below 1-2 Hz, even near the source [5]. Understanding earthquake energy release is crucial for assessing seismic hazards and improving our knowledge of tectonic processes.

The Indonesian archipelago is located in the Pacific Ring of Fire, making the region very vulnerable to earthquakes and tsunamis [9]. Research on the existence of active faults is still a gap in Indonesia but is very important in terms of disaster preparedness [9]. He said that the country experiences more than 10 earthquakes every day due to its geological conditions [10]. Indonesia also experiences huge losses due to earthquakes every year amounting to US$ 4,745 million [11]. To address this challenge, initiatives to strengthen community preparedness and disaster risk reduction have been prepared [12]. This approach has evolved from a charity-driven approach to a more integrated and productive empowerment initiative [13]. Despite these programs, it is unfortunate that more strategies are needed to increase the understanding, involvement, and use of indigenous knowledge in community-based activities to improve earthquake resilience in Indonesia.

Various clustering methods have been used in previous earthquake studies in Indonesia. Machine learning techniques, particularly unsupervised methods, have proven invaluable in uncovering patterns and structures within complex seismic data [14]. In this study, K-Means Clustering is employed to classify earthquake events based on their spatial distribution, magnitude, and energy, allowing for the identification of high-risk zones. Additionally, Principal Component Analysis (PCA) is utilized to reduce data dimensionality, enabling a focus on the most critical features while retaining the integrity of the overall variability in the dataset [15].

Series of algorithms including CLARA and K-Medoids, K-Mean [16], Fuzzy Possibilistic C-Means [17], and Bisecting K-means [18] have been used to classify earthquakes based on their magnitude and depth. This study aims to determine trends and areas that may pose a threat. It has been noted that the CLARA algorithm is efficient for large datasets and is relatively less sensitive to outliers [16]. Regarding the Comparison of Kmeans and Bisecting K-Means, K-means is superior when assessed by, Silhouette Coefficient and Davies Bouldin Index [18]. The Fuzzy Possibilistic C-Means method can be adapted to successfully cluster earthquakes primarily based on depth even though the magnitude distribution is much more random across clusters [17]. This clustering technique is well suited for the practice of earthquake risk assessment and management in Indonesia.

Seismic events in Indonesia have been examined with the help of popular media focusing on topics such as disaster planning, prediction, and timelines. Media for data visualization has been used to understand the status of measuring the quality of earthquake information [19] and for the use of remote sensing tools to determine areas susceptible to seismic shaking [13]. These visualizations are designed with the aim of improving the understanding and utilization of earthquake information more widely [20]. Data mining and geospatial techniques are also used by utilizing earthquake periodicity and risk assessment of high-risk zones especially in North Maluku [21]. Therefore, this research activity is directed to improve the level of preparedness and earthquake risk management in Indonesia - a country known for its earthquakes, located on the Pacific Ring of Fire and the meeting of three tectonic plates.

The purpose of this study is to evaluate earthquake clustering and spatial patterns in Indonesia over the period from 1992 to 2024. It strives to identify possible segments and areas of clustering through which to gain an understanding of earthquakes in the region, estimate earthquake threats, and possibly create plans and policies to aid in disaster management and prevention improvement. The findings are expected to deliver several key benefits. Scientifically, the research will enhance the knowledge of the space and time occurrence of earthquakes in Indonesia and this knowledge will contribute to the body of knowledge in the discipline of seismology and disaster management. Practically, the study strives to assist the policymakers in designing more efficient disaster management policies through the localization of areas that are prone to sever earthquakes and their possible future occurrences. Moreover, the study illustrates the application of clustering and geospatial tools in the analysis of the risk of earthquakes, which can be utilized all around the world, and attempts to strengthen communities by focusing on areas that are most at risk.

The study spans a 3 decade (1992-2024) data set of earthquakes which offers valuable insights into earthquake chronology. It concentrates on Indonesia, which is known to be one of the most active zones along the Pacific Ring of Fire, focusing on both the regional and the local distribution of the occurrence of earthquakes. This study combines sophisticated clustering such as K-Means methods with mapping techniques to derive information about seismic events, their frequency, and spatial arrangement.

Other analyzes are done as well, such as the trends of previous earthquakes' frequency, strength and spatial arrangement, as well as how the earthquake activity relates to other environmental and geological features. These final findings are anticipated to enhance the body of knowledge on the dynamics of earthquakes in Indonesia and offer reliable solutions for effective disaster management, policy formulation and risk control. Combining scientific methods with real life effectiveness, this study highlights its possible contributions on academic works and preparedness of the society to the most dreadful earthquakes.

METHODS

This study presents a more structured and integrative methodological approach to earthquake data in Indonesia by incorporating more advanced data processing and machine learning techniques. The data include seismic events from 1992 to 2024 based on magnitudes, energy and operational distribution. At the same time, meteorological data is also examined to find the possible relations. In order to find structures and hierarchies in the data set, Principal Component Analysis (PCA) and K-Means clustering are applied. Additionally, various techniques of data visualization are used to illustrate the trends related to energy distribution and seismic activity. The research flow is illustrated in **Figure 1** that outlines the step-by-step processes of data collection, processing, analysis and evaluation in order to comprehend the comprehensive research design of the study at hand.

This research adopts a multidisciplinary strategy that integrates physics, geospatial information, and machine learning algorithm to delve into the seismic phenomena of Indonesia. The principal data is sourced from the earthquake database that records the magnitude, latitude and longitude coordinates, depth, and the date of occurrence since 1992. This information is sourced from USGS Earthquakes dataset and NASA POWER and is thereafter processed to remove duplicates, invalid comments, and errors. In addition, to evaluate the trend of earthquake intensity, earthquake energy is also determined based on physics defined equations and various levels of magnitude.

$$E = 10^{\,1.5M+4.8} \tag{1}$$

The dataset undergoes an explorative analysis whereby temporal factors and patterns are evaluated using visual methods such as the annual earthquake energy distribution and magnitude boxplots over a period. The first phase includes a temporal and spatial look-up of values, which assists in identifying periods which have the highest seismic activity levels and regions which have high accumulation of energy. Calculate the earthquake energy as in **Equation (1)** , where E is the earthquake energy in joules and M is the earthquake magnitude on the Richter scale [22]. In the same way, the obtained spatial data is then related to the Indonesian maps shapefile in order to relate the intensity of the earthquakes with the energy distribution in an exponential function for ease in the identification of prevalent regions of earthquakes.

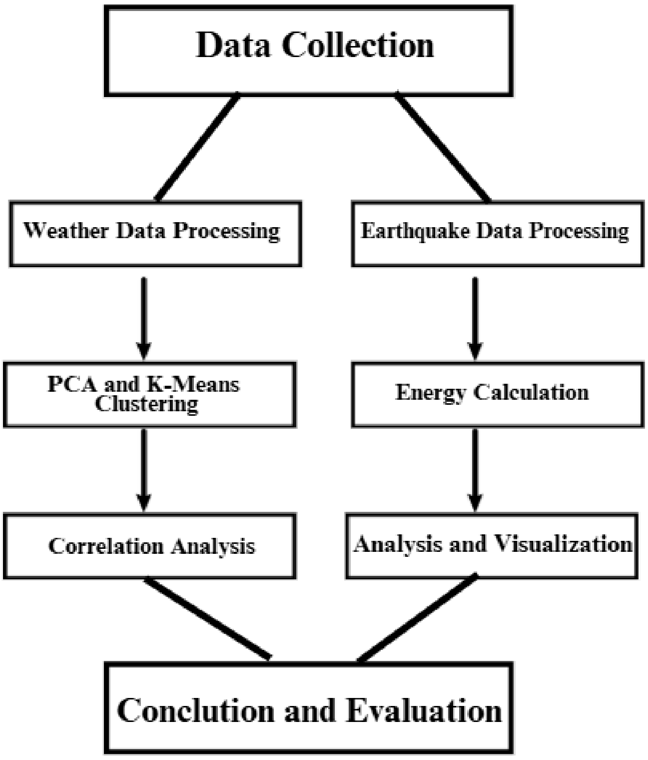


Figure 1. Flow Chart

In this study, machine learning is applied for a more in-depth analysis. The Principal Component Analysis (PCA) method is used to reduce the dimensionality of the data without losing important information, allowing focus on the main attributes that contribute to data variation. Dimensionality reduction is done by projecting the data into the principal vector space formulated in the **equation (2)**. Where Z represents the transformed data in a lower-dimensional space, X is the initial data, and W is the eigenvector matrix of the data covariance containing the principal components [23].

$$Z = X\,.W \tag{2}$$

Furthermore, the K-Means Clustering algorithm in the equation is used to group earthquakes based on geographic location, magnitude, depth, and energy. In the **equation (3)**, Where J represents the objective function representing the total inertia in all clusters, K is the number of clusters, Ci is *cluster-i*, x is the data set, and μ_i is the centroid of *cluster-i*. This algorithm was chosen because of its efficiency in handling large seismic data and its ability to identify hidden patterns that may be related to certain tectonic activities [24],[25]. The number of clusters is determined based on an evaluation using the Silhouette Coefficient to ensure optimal clustering results. The optimal number of clusters is determined using the Silhouette Coefficient presented in the equation. Based on the **equation (4)**, S is the Silhouette Coefficient, and the values a and b are the average distance between data in the same cluster and different clusters [26].

$$J = \sum_{i=1}^{k} \sum_{x \epsilon C_i} [\![x - \mu_i]\!]^2 \quad (3)$$

$$S = \frac{b - a}{\max(a, b)} \quad (4)$$

The clustering results are further analyzed to evaluate the relationship between variables in each cluster. A correlation heatmap is used to identify the most influential factors in earthquake distribution patterns. The results of the analysis are also visualized through earthquake distribution maps, graphs of the relationship between magnitude and energy, and visualization of cluster distribution in Indonesia. This research is expected to provide a significant contribution to understanding the dynamics of earthquakes in Indonesia and support disaster risk management based on physics and machine learning.

RESULTS AND DISCUSSION

The data processing results show a pattern of fluctuations in the number of events and the distribution of earthquake magnitudes over a three-year period from 1992 to 2024. In **Figure 1(a),** the number of earthquakes tends to increase from 1992-1995 to peak in 2004-2007 with 89 earthquakes. This is related to the major event of the Aceh Earthquake and Tsunami in 2004. After that, there was a decrease in the number of earthquake events, but the number of earthquakes increased again between 2020 and 2023. This pattern deserves further study to see the pattern of major earthquakes such as the Aceh Earthquake and Tsunami in 2024. **Figure 2(b)** shows the distribution of earthquake magnitudes, which are relatively the same for each period, with a median ranging from 6.0 to 8.0 on the Richter Scale. A higher pattern is seen for the median magnitude in 2004-2007. This supports the events and patterns of earthquakes that occurred during that period. In addition, each period has outliers that represent earthquakes with very large magnitudes, although the frequency of these outliers varies.

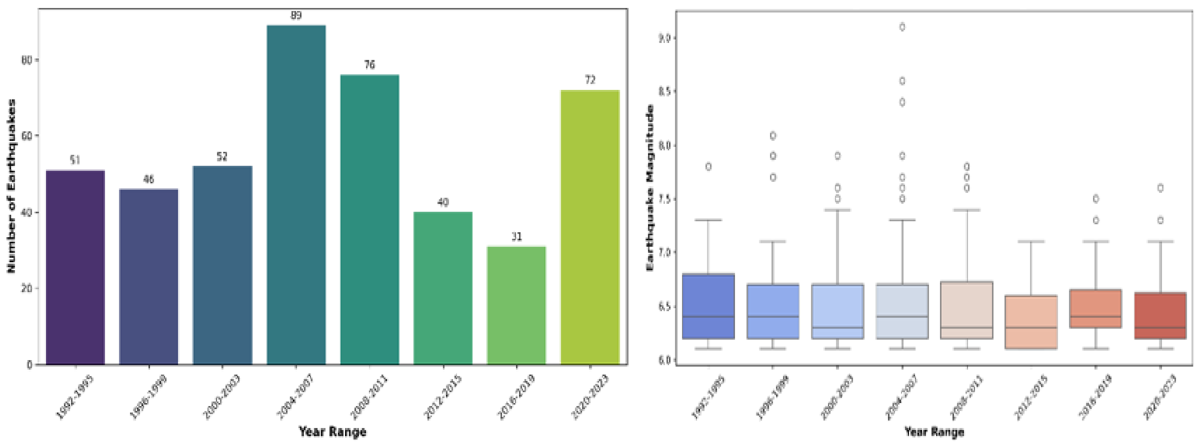


Figure 2. (a) The Number of Earthquakes Over 3-Year Period (b). Magnitude Based on Year Period

In general, in **Figure 2(a)** and **Figure 2(b)** images, there is a temporal pattern in earthquake activity in Indonesia. To see it further, further research is needed, with a range of earthquake data of up to 100 years. Furthermore, this pattern can be analyzed further to see indications of association with weather parameters and atmospheric conditions such as temperature, air pressure, and global climate phenomena. This analysis will be discussed further with a machine learning approach to understand the complex relationship between seismic activity and atmospheric factors.

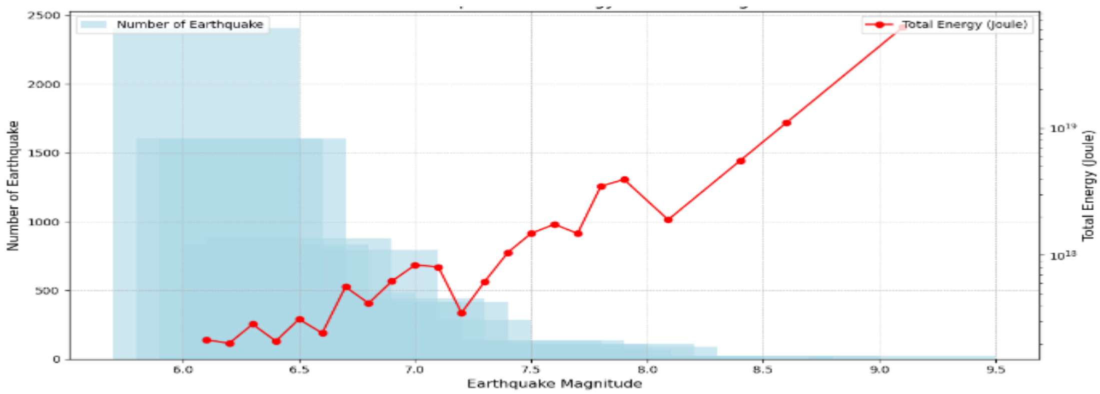


Figure 3. Number of Earthquake and Energy Based on Magnitude

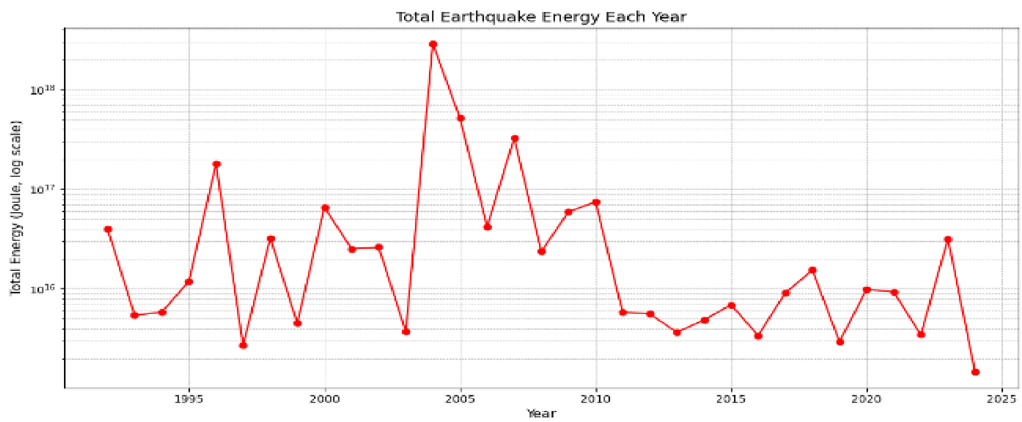


Figure 4. Total Earthquake Energy Each Year

Figure 3 shows the relationship between the number of earthquakes and energy based on the magnitude of the earthquake distribution. The histogram shows that the earthquakes that often occur in Indonesia are mostly with a magnitude of 6.0-6.5 on the Richter Scale. The number of earthquakes decreases exponentially as the magnitude increases. This pattern is in accordance with the Gutenberg-Richter law, which states that the number of earthquakes decreases logarithmically with increasing magnitude. Although large earthquakes with a magnitude > 7.5 on the Richter Scale are rare, the energy released increases exponentially. This is due to the non-linear relationship between magnitude and seismic energy. As mentioned in the **equation(1)**, an increase in magnitude by 1-unit results in an energy release of about 32 times greater. Therefore, earthquake prediction with machine learning must be developed to predict the occurrence of large earthquakes so that they can reduce their impact on humans, the environment, and infrastructure. **Figure 4** shows the annual seismic energy fluctuation with a significant peak in 2004, likely caused by the earthquake in Aceh (9.1 SR). The decrease in total energy in the following years shows that after the release of large energy, seismic activity tends to decrease. This is consistent with the theory that plates need time to re-accumulate energy before the next big one occurs. The earthquake energy releases every year shows the process of releasing elastic energy accumulated in active fault zones. This pattern is to the concept of the seismic cycle, where the accumulation of stress in the earth's crust will be followed by the release of accumulated energy through earthquakes.

The spatial energy distribution of earthquakes in Indonesia illustrates in **Figure 5** the pattern of seismic activities, which is closely controlled by the presence of a subduction zone and faulting systems. The Sumatra to Nusa Tenggara region is among those with the highest concentration of earthquake occurrences, reflecting the influence of the subduction of the Indo-Australian plate against the Eurasian plate. The map shows that the largest concentration of earthquake energy release occurred in the region surrounding the megathrust subduction zone, which included the western coast of Sumatra and south of Java. This is consistent with the history of great earthquakes, including the Sumatra megathrust, that has significantly impacted the region.

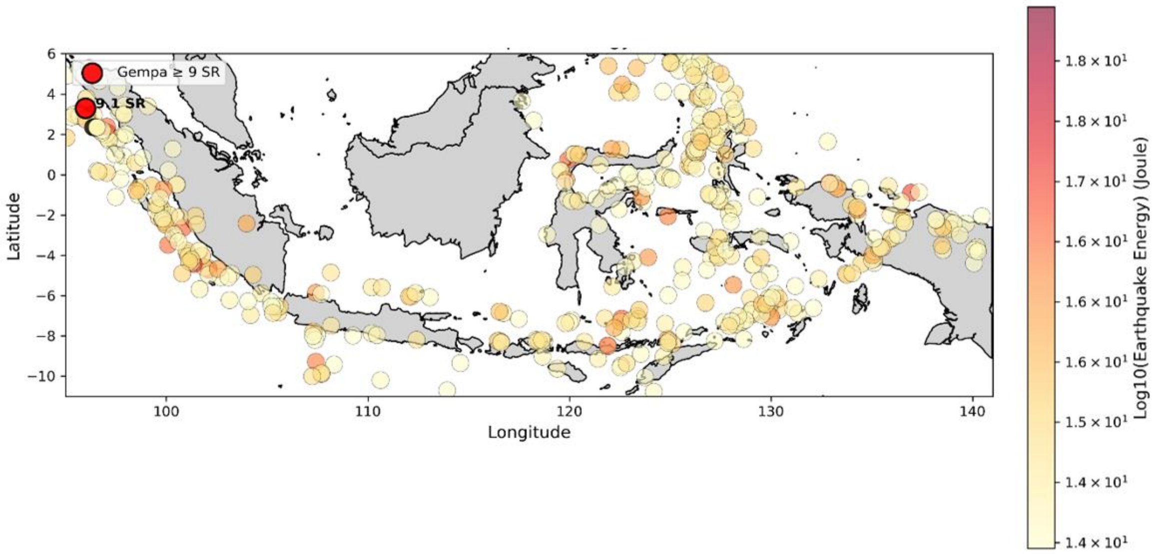


Figure 5. Distribution of Earthquake Energy in Indonesia

Moreover, there is a noticeable increase in the intensity of earthquake energy in the east part of Indonesia, in the areas of the Maluku Islands and Papua, which has been associated with high tectonic activity induced by the region's triple junction plate system. The distribution pattern of energy revealed in this study also suggests that, on the other hand, the earthquake activities in Indonesia may also include other nonrelated to the main subduction zone located throughout the eastern region that is influenced by local active faults.

The logarithmic scale utilized to illustrate the energy of earthquakes provides insight, showing that there is a significant difference in the energy released by small and large magnetism. The occurrence of earthquakes with extraordinary energy puts into focus the scale of measures that must be redoubled concerning mitigation planning in order to abate the risk of a major disaster. The darker colors on the map correspond to regions where there was very high energy release, which implies the unsuppressed activity of tremors that can cause destruction.

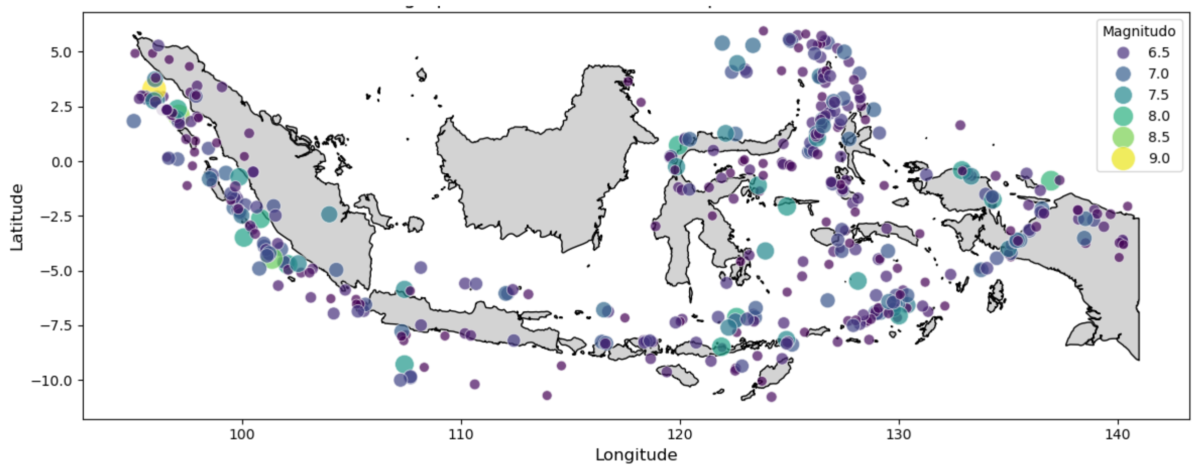


Figure 6. Geographical Distribution of Earthquakes in Indonesian

This spatial distribution in is not only applicable in the context of the seismic hazard mapping but is also useful in assisting the understanding of tectonic processes of Indonesia. This is significant in designing risk reduction strategies which are focused on the spatial distribution of seismic energy releases, with emphasis on areas of high energy release located in the proximity of densely populated areas.

This study uses Principal Component Analysis (PCA) and K-means in **Figure 7** clustering to analyze the relationship between earthquakes in Indonesia and environmental variables such as temperature, specific humidity, surface pressure, wind speed, and radiation. The PCA results show that the data can be reduced to two principal components (Principal Component 1 and 2), which represent most of the variability in the dataset. PCA visualization reveals that most of the data is well distributed around the center, while some outliers are seen in certain areas, indicating earthquakes with unique or extreme environmental conditions. Moreover, in the clustering analysis utilizing K-Means in **Figure 7(b)**, there are three groups of clusters formed, which are clusters 0, 1 and 2. The graphical presentation of the results of clustering shows that there is a reasonable degree of dispersion, even though clusters have some degree of overlap. The merit of clustering through K-Means is the manner in which it is possible to describe the features of an earthquake by its quantitative determinants without any preceding picture of the earthquakes which may help to establish the earthquake dangers characteristics in prone regions. Therefore, these types of results can be used for attempting to set up further more precise risk reduction approaches in accordance with the characteristics of the clusters.

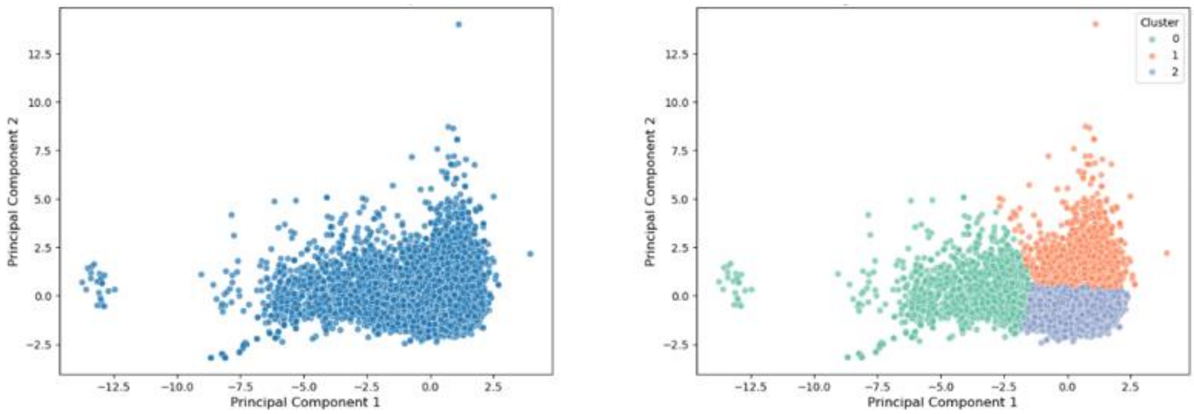
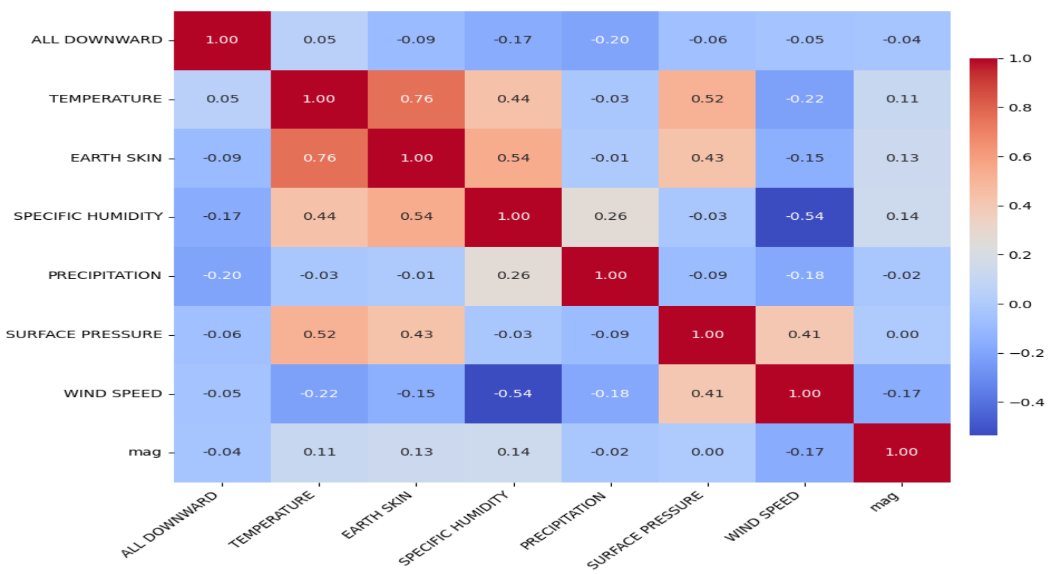


Figure 7. (a) Data Visualization with PCA 2 Component (b) Clustering Results with K-Means (3 Clusters)

Principal Component Analysis (PCA) in **Figure 7(a)** visualization shows effective dimensionality reduction, where two principal components are able to explain most of the variation in earthquake data. The data distribution in the two-dimensional PCA space indicates a scattered pattern with a certain density, representing the heterogeneity of earthquake parameters (e.g., magnitude, depth, and energy). PCA allows one to visualize the parameters associated with certain seismic phenomena and one may observe that a certain bulk of data density on the PCA plot suggests that most of the earthquakes occur at the same parameters on the mid-magnitude frequency, while those scattered outside suggest the presence of significant earthquakes which have extreme parameters. The combination of PCA and K-Means works towards better understanding the seismic data statistically and also works as a predictive model to identify probable events of concern. This pair indeed has the potential to be used in forecasting anomalies associated with precursors of earthquakes, both from tectonic and atmospheric data, thus providing a firm basis and direction for future multidisciplinary studies.



	ALL DOWNWARD	TEMPERATURE	EARTH SKIN	SPECIFIC HUMIDITY	PRECIPITATION	SURFACE PRESSURE	WIND SPEED	mag
ALL DOWNWARD	1.00	0.05	-0.09	-0.17	-0.20	-0.06	-0.05	-0.04
TEMPERATURE	0.05	1.00	0.76	0.44	-0.03	0.52	-0.22	0.11
EARTH SKIN	-0.09	0.76	1.00	0.54	-0.01	0.43	-0.15	0.13
SPECIFIC HUMIDITY	-0.17	0.44	0.54	1.00	0.26	-0.03	-0.54	0.14
PRECIPITATION	-0.20	-0.03	-0.01	0.26	1.00	-0.09	-0.18	-0.02
SURFACE PRESSURE	-0.06	0.52	0.43	-0.03	-0.09	1.00	0.41	0.00
WIND SPEED	-0.05	-0.22	-0.15	-0.54	-0.18	0.41	1.00	-0.17
mag	-0.04	0.11	0.13	0.14	-0.02	0.00	-0.17	1.00

Figure 8. Correlation Heatmap for Cluster 0

Correlation analysis of atmospheric variables and seismic parameters provides important insights into the relationship between atmospheric parameters and earthquake characteristics in each group. Starting from Cluster 0 in **Figure 8**, which is analyzed in more depth using a correlation heatmap between variables. This analysis shows a strong positive correlation between temperature and the earth's surface temperature (0.76), indicating a close relationship between atmospheric conditions and land temperature. In addition, surface pressure has a moderate correlation with specific humidity (0.41), indicating atmospheric dynamics affecting this cluster. A prominent negative correlation is found between wind speed and specific humidity (-0.54), which may be due to the movement of dry air masses or the effects of evaporation.

Cluster 1 in **Figure 9** , temperature shows a moderate positive correlation with surface pressure (0.33) and specific humidity (0.36). Cluster 1 indicates that earthquakes in this cluster are often associated with stable atmospheric conditions, where temperature affects energy release and surface pressure patterns. On the other hand, some factors reach low or almost zero correlations with earthquake magnitude such as wind speed and precipitation, which implies that extremely complicated atmospheric dynamics have no vital effects on earthquakes of this cluster. Earthquake magnitude (mag) shows a weak correlation to atmospheric variables (the highest correlation is 0.06), indicating that earthquakes in this cluster are more determined by geological factors than atmospheric factors.

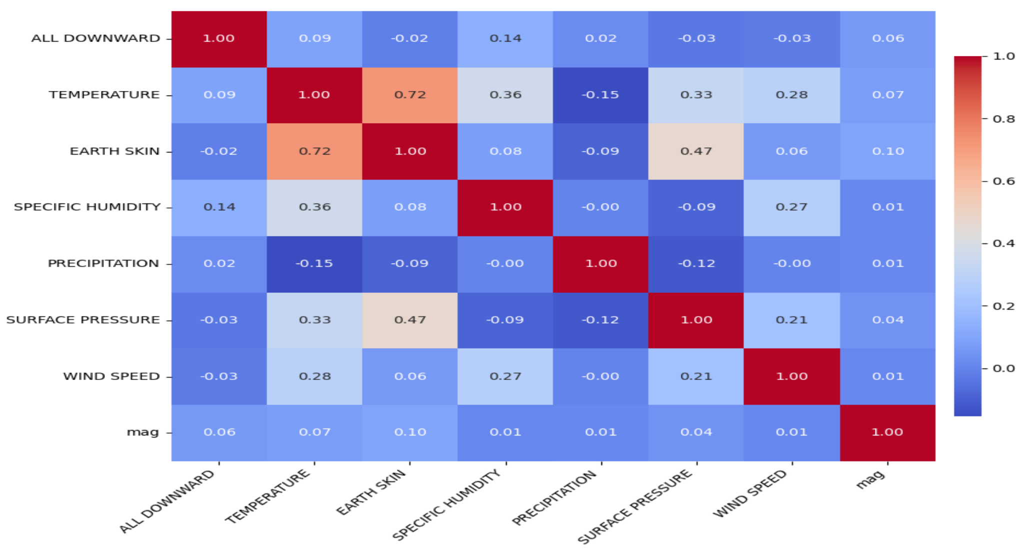


Figure 9. Correlation Heatmap for Cluster 1

In contrast to cluster 1, the correlation cluster 2 in **Figure 10** shows a more significant relationship between atmospheric variables. Temperature shows a very high correlation with the temperature of the earth's surface (earth skin) (0.83), indicating that local thermal conditions have a more dominant influence on earthquake activity in this cluster. Cluster 2 can refer to the relationship between the accumulation of thermal energy in the earth's crust and the event of energy release in the form of earthquakes. In addition, specific humidity also has a higher correlation with other variables, such as precipitation (0.35) and earth's surface temperature (0.40). Cluster 2 can indicate a relationship between local atmospheric activity (for example, rain patterns or high humidity) and tectonic conditions in this cluster area.

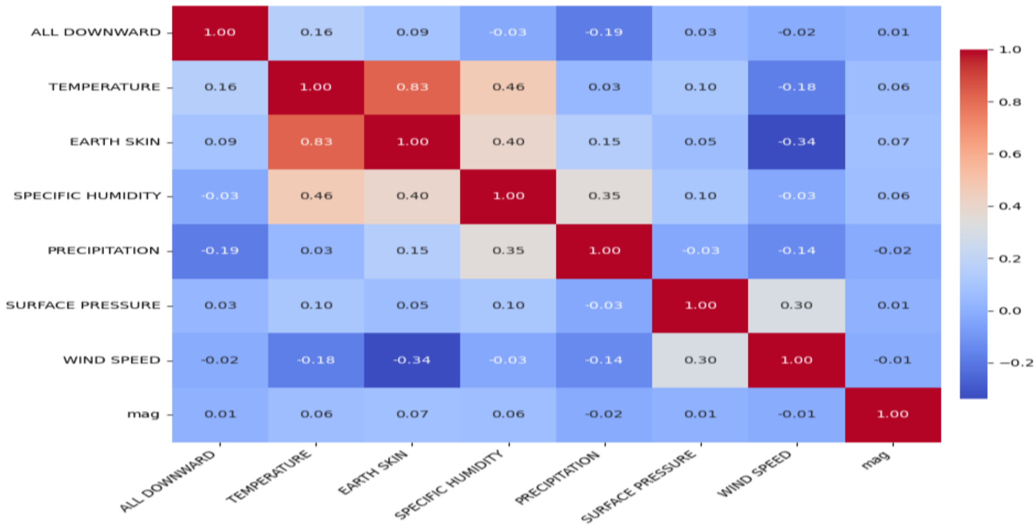


Figure 10. Correlation Heatmap for Cluster 2

SUMMARY

In this study, over a period of thirty years, the research attempt to find a link between the environmental factors and the earthquake occurring in Indonesia. The evidence suggests that the earthquakes occurred at different points in time and areas with varying intensities, magnitudes and energy strength. A notable surge in recorded earthquakes was seen between 2004 to 2007 during which the Aceh Earthquake (9.1 SR) and the subsequent Tsunami struck. The trend is also observed that while there are multiple earthquakes only a few have a timespan that vary between 6.0-6.5 which makes it more abundantly common which also coincides with the energy released in waves aligning with the curve graph presented by Gutenberg-Richter law. Geospatial data analyzed shows that areas around the Sumatra-Java region and the eastern portions of Indonesia are the most prone to seismic activity due to the collisions and movements of the tectonic plates in the triple junction region.

This research along with focusing on seismic activity correlation also involved analyzing the atmospheric data points through the use of PCA-K-Means clustering which includes humidity, temperature, pressure, wind speed, etc. PCA also managed in reducing the data set complexity to encompass two main components thereby allowing most of the variance within the data set to remain intact while the K-Means clustering enabled tilling earthquakes with common environmental features into three categories. Unique correlations in a specific cluster also enabled the researcher in understanding how each variable individually impacted the seismic parameters. An example of this would be in cluster two where the surface temperature was noted to be high during periods of seismic activity suggesting a possible correlation string between the two variables.

The research calls for additional multidisciplinary studies especially those using machine learning techniques to model earthquakes based on atmosphere and climate parameters. Findings stress the significance of performing seismic hazard mapping and risk minimization approaches in Indonesia with high population densities and high-energy seismic areas. In future research, it is suggested that the dataset should span 100-year time periods and the model be better configured to illustrate the tectonic and atmospheric relationship more elucidative.

REFERENCES

[1] Alex Copley, "Studying Earthquakes to Save Lives," 2017.

[2] W. H. K. . Lee, Hiroo. Kanamori, Paul. Jennings, and Carl. Kisslinger, *International Handbook of Earthquake & Engineering Seismology. Part A*. Academic Press, 2003.

[3] C. Wang and M. Manga, *Earthquakes and Water*, vol. 114, Lecture Notes in Earth Sciences. Berlin, Heidelberg: Springer, 2009. doi: 10.1007/978-3-642-00810-8.

[4] A. Gudmundsson, "Elastic energy release in great earthquakes and eruptions," *Front Earth Sci (Lausanne)*, vol. 2, no. May, May 2014, doi:10.3389/feart.2014.00010.

[5] M. S. Vassiliou and H. Kanamori, "The Energy Release In Earthquakes," 1982.

[6] M. Ohnaka and M. Ohnaka, "Earthquake-source parameters related to magnitude," 1978. [https://academic.oup.com/gji/article/55/1/45/680684

[7] A. Gudmundsson, "Elastic energy release in great earthquakes and eruptions," *Front Earth Sci (Lausanne)*, vol. 2, no. May, May 2014, doi: 10.3389/feart.2014.00010.

[8] C. A. L. Qingwen Miao, "Spatial Distribution of Earthquake Energy Release in the Central United States from a Global Point of View," 2008. [Online]. Available: http://folkworm.ceri.memphis.edu/

[9] Danny Hilman Natawidjaja, *Riset Sesar Aktif Indonesia dan Peranannya dalam Mitigasi Bencana Gempa dan Tsunami*. 2021.

[10] E. S. Z. Balok Budiyanto, "Penanganan Pasca Bencana Gempa di Lombok," 2018.

[11] M. Irsyam *et al.*, "Past earthquakes in Indonesia and new seismic hazard maps for earthquake design of buildings and infrastructures," in *Geotechnical, Geological and Earthquake Engineering*, vol. 25, Springer Netherlands, 2013, pp. 33–46. doi: 10.1007/978-94-007-5675-5_3.

[12] L. Maulida, R. Ocktadinata, and B. Adhayanti, "Kesiapsiagaan Masyarakat Kota Cilegon dalam Menghadapi Bencana Gempa Bumi dan Tsunami," *Preventia : The Indonesian Journal of Public Health*, vol. 7, no. 1, p. 20, Jun. 2022, doi: 10.17977/um044v7i12022p20-27.

[13] R. Akbar et al., "Impelementasi Business Intelligenci Menentukan Daerah Rawan Gempa Bumi di Indonesia dengan Fitur Geolokasi," *Jurnal Edukasi dan Penelitian Informatika*, 2018.

[14] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine Learning and Data Mining Methods in Diabetes Research," 2017, *Elsevier B.V.* doi: 10.1016/j.csbj.2016.12.005.

[15] C. Ding, "K-means Clustering via Principal Component Analysis."

[16] P. A. Kaban, R. Kurniawan, R. E. Caraka, B. Pardamean, B. Yuniarto, and Sukim, "Biclustering method to capture the spatial pattern and to identify the causes of social vulnerability in Indonesia: A new recommendation for disaster mitigation policy," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 31–37. doi: 10.1016/j.procs.2019.08.138.

[17] E. D. A. Sani Putriana, "Clustering Data Titik Gempa dengan Metode Fuzzy Possibilitic C-Means," *Jurnal Rekursif*, Mar. 2021.

[18] T. H. J. H. Adela Tania, "Perbandingan Antara Algortima K-Means dan Algoritma Bisecting K-Means Dalam Menganalisis Gempa Bumi di Indonesia," *Jurnal Sistem Informasi dan Teknik Komputer*, 2023.

[19] M. Chandra Kirana *et al.*, "Visualisasi Kualitas Penyebaran Informasi Gempa Bumi di Indonesia Menggunakan Twitter," 2019. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[20] A. Amirsardari, M. Sofi, E. Lumantarna, I. Imran, and C. Duffield, "Impact of Earthquakes on the Transportation Infrastructure of Indonesia: A Preliminary Study," *Civil Engineering Dimension*, vol. 21, no. 1, pp. 19–28, Mar. 2019, doi: 10.9744/ced.21.1.19-28.

[21] I. Arfiani, D. Normawati, and M. D. Suratin, "Analisis dan Visualisasi Periodisitas Gempa Bumi di Maluku Utara," *Jurnal Ilmiah FIFO*, vol. 16, no. 1, p. 37, May 2024, doi: 10.22441/fifo.2024.v16i1.004.

[22] B. Gutenberg, "The Quarterly Journal of Geological Society of London Volume CXII The Energy of Earthquakes," 2015. [Online]. Available: http://jgslegacy.lyellcollection.org/

[23] N. Salem and S. Hussein, "Data dimensional reduction and principal components analysis," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 292–299. doi: 10.1016/j.procs.2019.12.111.

[24] I. Hidayatur Rifa and H. Pratiwi, "Media Statistika 13(2) 2020: 194-205 194 CLUSTERING OF EARTHQUAKE RISK IN INDONESIA USING K-MEDOIDS AND K-MEANS ALGORITHMS," *MEDIA STATISTIKA*, vol. 13, no. 2, pp. 194–205, 2020, doi: 10.14710/medstat.13.1.194-205.

[25] H. J. Wattimanela, U. S. Pasaribu, S. W. Indratno, and A. N. T. Puspito, "Earthquakes clustering based on the magnitude and the depths in Molluca Province," in *AIP Conference Proceedings*, American Institute of Physics Inc., Dec. 2015. doi: 10.1063/1.4936449.

[26] V. N. H. G.-N. N. Jian Chen, *Knowledge and Systems Sciences*, vol. 1103. in Communications in Computer and Information Science, vol. 1103. Singapore: Springer Singapore, 2019. doi: 10.1007/978-981-15-1209-4.