Original Article

e-ISSN: 2581-0545 - https://journal.itera.ac.id/index.php/jsat/

Statistical Pattern Recognition of Lithosphere Anomalous Activity Along the Indonesian Ring of Fire

Received 03rd June 2024
Accepted 19th April 2025
Published 19th April 2025

Open Access

DOI: 10.35472/jsat.v9i1.1850

Mika Alvionita[a], Ardika Satria[a*], Triyana Muliawati[b], Fuji Lestari[c], Danni Gathot Harbowo[d]

[a] Department of Science Data, Institut Teknologi Sumatera, 35365, Wayhui, Lampung, Indonesia

[b] Department of Mathematics, Institut Teknologi Sumatera, 35365, Wayhui, Lampung, Indonesia

[c] Department of Actuarial Science, Institut Teknologi Sumatera, 35365, Wayhui, Lampung, Indonesia

[d] Center for Sustainable Development Goals Studies (ITERA SDGs Center), Institut Teknologi Sumatera, 35365, Wayhui, Lampung, Indonesia

* Corresponding E-mail: ardika.satria@sd.itera.ac.id

Abstract: The introduction of statistical pattern recognition becomes highly important for assessing disaster threats such as earthquakes. This approach is significantly more comprehensive and suitable for long-term event forecasting. Therefore, in the future, efforts can be promptly made to reduce the risk of disasters resulting from anomalies in lithospheric activity, especially frequent earthquakes in the Sumatra Island region, Indonesia. Statistical pattern analysis of lithospheric activity anomalies can be categorized through classification. Earthquake classification is performed based on magnitude scale and mathematical calculations of earthquake parameter unit conversion. The classification method employed in this research includes machine learning methods like k-nearest neighbor and support vector machine. The evaluation metrics used for machine learning models are model accuracy and confusion matrix tables.

Keywords: Earthquakes, Magnitude, Classification

Introduction

Indonesia is situated within the Pacific Ring of Fire, which is in the Pacific Ocean, making it prone to earthquakes and volcanic eruptions. Earthquakes occur because of their geographic location between the Indo-Australian Plate, the Eurasian Plate and the Pacific Plate. The movement between these plates causes earthquakes that occur in Indonesia. One example of movement between plates is the meeting between the Indo-Australian plate and the Eurasian plate in the Java Island region, which is known as the Sunda subduction zone. As a result, the Indo-Australian plate sinks beneath the Eurasian plate, which causes earthquakes and volcanic eruptions. According to experts, approximately 81% of earthquakes occur in Indonesia [1].

Earthquakes in Indonesia have had a significant impact on society and the environment. According to the Ministry of Energy and Mineral Resources (ESDM), 25 regions in Indonesia are prone to earthquakes [2]. One of these areas is Sumatra Island, which is crossed by active faults, volcanic paths, and subduction zones [4]. The earthquakes that occurred on Sumatra Island involved large shifts such as the Semangko Fault, Sumatra Fault and Mentawai Fault, with the most significant earthquake occurring in West Sumatra in 2004 [5].

In this research, earthquake locations were determined to carry out statistical analysis of the frequency and distribution of seismic data in the Sunda Strait region using a machine learning approach. The machine learning approach used is classification using the Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) methods. SVM is a learning system that uses a hypothesis space in the form of a linear function in high-dimensional features based on optimization theory. The goal of using SVM is to build a model to predict given test data. Meanwhile, KNN aims to classify based on the closest distance to other objects based on attributes and training data. This research aims to compare the SVM and KNN methods to identify earthquake-prone locations, with the hope of enhancing the understanding of seismic patterns and activities, as well as evaluating geohazard risks in the Sunda Strait region.

Original Article *Journal of Science and Applicative Technology*

Method

Dataset

The review of lithospheric dynamic phenomena in this study includes the timing of earthquake events in Indonesia [6]. Tectonic earthquake activity data will be compiled from the global earthquake database at earthquake.usgs.gov; in this case, specially on the sumatera region. The earthquake data collected includes the time of the earthquake, epicenter coordinates, earthquake magnitude and depth of the earthquake epicenter.

The compiled data that has been compiled will also be cross-reviewed with experts from the Indonesian Center for Volcanology and Geological Disaster Mitigation (PVMBG) and the Indonesian Meteorology, Climatology and Geophysics Agency (BMKG) [7]. The amount of data acquired was 16,263, with 22 columns of features. Then, proceed with data cleaning for data with NaN values and incomplete data. So, 14,607 earthquake data remain with the MB (body-wave magnitude) and Mw (moment magnitude) magnitude scales. This scale measures the energy released by an earthquake based on primary wave signals (P-wave) that propagate through the earth's core. Over time, the magnitude scale has evolved, and today, Mb magnitude is often used together with Mw (Moment Magnitude) [8], which is considered to be a more accurate measure of the total energy released by an earthquake in **Table 1.**

Table 1 Dataset of Sumatran earthquakes from 1990-2023

	latitude	longitude	depth	mag	rms	class	class_depth	years	months	days	Hours	Minutes	Seconds
0	-9.4120	107.2600	31.900	4.7	1.10	0	0	1990	1	1	14.0	10.0	3.0
1	-5.0060	102.4290	33.000	4.8	1.60	0	0	1990	1	11	21.0	23.0	35.0
2	7.5040	94.1740	33.000	4.1	1.10	0	0	1990	1	13	19.0	11.0	47.0
3	-5.0340	102.6780	122.300	4.1	1.10	0	1	1990	1	24	11.0	23.0	39.0
4	-5.8830	105.6700	104.700	5.3	1.30	1	1	1990	1	28	3.0	58.0	49.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
14112	-5.3791	102.0612	24.947	4.8	0.79	0	0	2023	7	21	3.0	40.0	21.0
14113	0.5611	92.3497	10.000	4.4	0.56	0	0	2023	7	21	21.0	30.0	52.0
14114	-7.9804	107.4013	54.701	4.4	0.46	0	0	2023	7	21	22.0	41.0	40.0
14115	-2.0959	99.8654	34.244	4.4	0.72	0	0	2023	7	22	20.0	27.0	35.0
14116	-4.8747	102.9517	64.566	4.5	0.30	0	0	2023	7	23	0.0	43.0	59.0

14117 rows × 13 columns

Exploratory Data Analysis

After cleaning the data, continue with Exploratory Data Analysis (EDA) to review information that may not have been revealed. Then, see the correlation between features through data visualization graphs. Correlation analysis can be seen from the scatter graph shown in **Figure 1** on the magnitude of all features. Based on the scatter graph, it was found that the relationship between the features was normally and evenly distributed, except for 2 features, namely the magnitude class and the depth class. These two features are the result of a classification of earthquakes on the scale of magnitude and depth of the Sumatra earthquake. According to research on earthquake classification according to the magnitude scale, it is shown in **Table 1**.

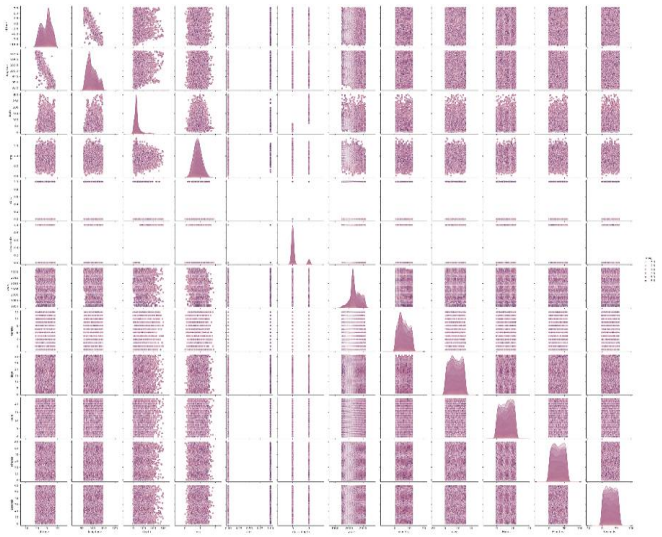

Figure 1 Sumatera Earthquake Magnitude Scatter Graph

Table 2 Classification of Earthquakes Based on the Magnitude Scale

Small Earthquake	Light Earthquake	Moderate Earthquake	Strong Earthquake	Great Earthquake
Magnitude less than 4.0	Magnitude 4.0 – 4.9	Magnitude 5.0 -5.9	Magnitude 6.0 – 6.9	Magnitude 7.0 – 7.9
Impact: not felt and recorded	Impact: Often felt, but damage often occurs	Impact: Can cause damage to buildings in the earthquake source area.	Impact: May cause serious damage to buildings in the affected area.	Impact: Can cause significant damage within a large radius.

The Sumatra earthquake hotspot area is shown in **Figure 2**. In this study, the earthquake hotspots were limited to the island of Sumatra and its surroundings.

Figure 2 Sumatra earthquake hotspot area and surrounding areas

Next, to determine the earthquake depth classification features, they are sorted based on the research [9,10,11] presented in **Table 3**. Shallow earthquakes occur at depths relatively close to the earth's surface. The impact can be

greater in densely populated areas because earthquake vibrations are felt more directly.

Table 3 Earthquake depth classification features

Shallow Earthquake	Medium Earthquake	Deep Earthquake
0 – 70 km	70 – 300 km	More than 300 km
Shallow earthquakes occur at depths relatively close to the earth's surface. The impact can be greater in densely populated areas because earthquake vibrations are felt more directly.	Intermediate earthquakes occur at greater depths within the earth's crust. Earthquakes at these depths are often associated with subduction zones and can cause significant damage, especially around the epicentre	Deep earthquakes occur at great depths in the Earth's mantle. This type of earthquake usually does not cause significant damage to the surface because the vibrations have propagated over a greater distance from the epicentre.

Intermediate earthquakes occur at greater depths within the Earth's crust and are often associated with subduction zones. These earthquakes can cause significant damage, especially near the epicenter, especially around the epicenter deep earthquakes occur at great depths in the Earth's mantle. This type of earthquake usually does not cause significant damage to the surface because the vibrations have propagated over a greater distance from the epicenter.

Classification of Earthquakes

Based on **Table 2** and **Table 3**, this study uses two classifications of earthquakes according to the magnitude scale and according to the depth of the earthquake. The number of classes that appear based on the magnitude scale is shown in **Figure 3a**. Meanwhile, based on the depth of the earthquake, the number of classes is shown in **Figure 3b**. The classification is divided into two classes, namely 0 and 1 with specifications for small and light earthquakes, then the rest are medium and strong earthquakes according to the magnitude scale. Next is the classification of earthquakes according to earthquake depth which is divided into two classes, namely, shallow earthquakes and medium earth quakes.

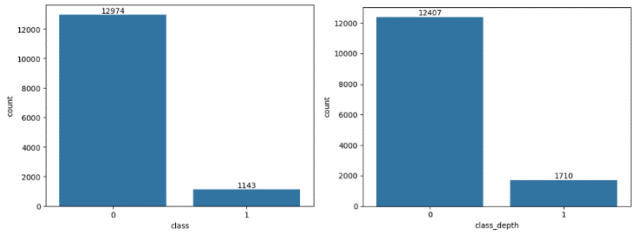

Figure 3 The Number of Classes in Earthquake Classification Based on (a) Magnitude Scale and (b) Earthquake Depth

KNN (K-Nearest Neighbor)

The K-Nearest Neighbors (KNN) classification method is an approach in machine learning that is used to predict the class of new data based on the majority class of the K nearest neighbors [12]. The main process of KNN begins by determining the K value, which represents the number of nearest neighbors that will be used to determine the class of new data. In feature space, the distance between new data $((u))$ with each data in the training dataset $((v))$ is measured using the Euclidean distance formula: $d(u,v) = \sqrt{\sum_{i=1}^{n}(u_i - v_i)^2}$. After that, KNN selects K training data points with the smallest distance to the new data point and determines the majority class of the K neighbors as the predicted class $((\hat{y}))$ [13]. In a mathematical formula, this can be represented as $(\hat{y} = arg\ arg\ \sum_{i=1}^{K} I(y_i = j))$, where $(\hat{y})$ is predicted class, (y_i) is the class of the i th neighbor, and (I) is the indicator function (1 if true, 0 if false) [14].

SVM (Support Vector Machine)

Support Vector Machine (SVM) is a machine learning method that works with the Structural Risk Minimization (SRM) principle, which aims to find the best hyperplane that separates two classes in the input space [15]. This method uses hypotheses in the form of linear functions in a high-dimensional feature space by implementing bias learning derived from statistical learning theory [16]. The level of accuracy in the model that will be produced by the switching process with SVM is very dependent on the kernel function and parameters used [17]. The classification process will use **Equation 1**. With b being the bias.

$$y = \{+1,\ if\ b + \alpha^T x \geq +1\ -1,\ if\ b + \alpha^T \leq -1 \qquad (1)$$

The kernel function used is as follows in equation 2.

Published by: Lembaga Penelitian dan Pengabdian Masyarakat (LPPM)
Institut Teknologi Sumatera, Lampung Selatan, Indonesia

$$K(x, x_i) = exp(-\gamma||x - y||^2) \quad (2)$$

The strength of SVM is that it avoids overfitting with small samples and is less sensitive to imbalanced distributions.

Model Evalution

The classification method in this research will be evaluated based on the accuracy and performance of the classification technique to determine its effectiveness. Evaluation is carried out to find the optimal solution resulting from various complex classification methods and is carried out repeatedly [18]. Testing of this research was carried out on test data for the classification function that had been trained. The results of this test are in the form of a prediction class, which will be compared with the fact class in the test data using a confusion matrix, as in **Table 4** [19].

Table 4 Confussion Matrix

		Actual	
		TRUE	FALSE
Prediction	TRUE	TP (True Positive)	FP (False Positive)
	FALSE	FN (False Negative)	TN (True Negative)

True Positive (TP) shows the number of testing data that the model classifies as the true class. False Positive (FP) shows the number of testing data in the column that matches the class but does not include TP. False Negative (FN) shows the number of testing data in rows that match the class but do not include TP. True Negative (TN) shows the amount of testing data in each column and row but does not include the class column and row [20].

Several standard measurements can be used to measure classification with two class labels, namely accuracy, true positive rate, false positive rate, true negative rate, false negative rate, and precision. The measurement used in this research is accuracy [21]. Accuracy is used to show accuracy of sentiment analysis; to calculate accuracy, Equation 3 is used.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Results And Discussion

In this study, the ratio of training data to test data was 80:20, with training data being 80% and test data being 20%. The accuracy results and f1 score with the KNN classification model are shown in **Table 5**. Meanwhile, the accuracy results and f1 score with the SVM classification model with radial basis function (RBF) kernel parameters are shown in **Table 6.**

Table 5 Accuracy Results and F1 Score of KNN Classification Model

Class	Precision	Recall	F1-score
0	0.99	1.00	1.00
1	1.00	0.95	0.97
Accuracy			0.99

Table 6 Accuracy Results and F1 Score of SVM Classification Model

Class	Precision	Recall	F1-score
0	0.96	1.00	0.98
1	1.00	0.65	0.79
Accuracy			0.96

The confusion matrix obtained from the two models above is shown in **Table 7** and **Table 8**. The number of test data for each class 0 and 1 is 2523 and 301 data.

Table 7 Confussion Matrix of KNN Classification Model

Class		True	False
0	True	2523	0
1	False	15	286

Table 8 Confussion Matrix of SVM Classification Model

Class		True	False
0	True	2523	0
1	False	105	196

Based on the results above, the accuracy of the KNN classification model is 0.9922 or 99%. This result is proven by the error found to be very small with a K value of 24. The error value is 0.0046, which is shown in the KNN graph with each K value from 1 to 40 in **Figure 4**.

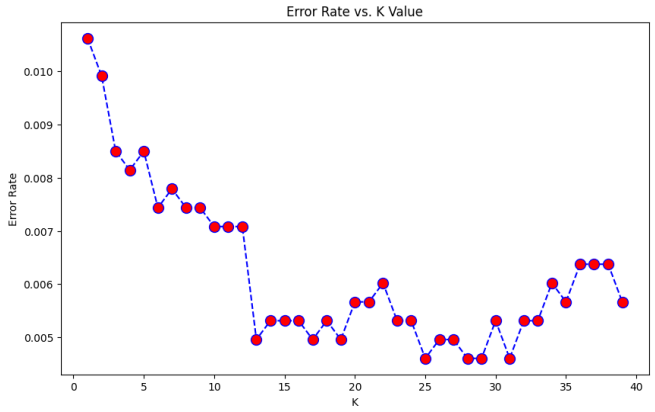

Figure 4 Error Rate Graph with K Value

Meanwhile, the confusion matrix value of the KNN classification model obtained was that in class 1, there were 15 incorrect guesses found. Next, check the value of each C parameter in the SVM classification model. The parameter C in the SVM model controls the trade-off between obtaining the maximum margin and allowing some data points to violate the decision boundary (margin). Parameter C is called the penalty parameter for misclassification of the training data. The best C value in the model is 1. Because the results shown are values 1 to 50, a stable value is found throughout the increase in the C value, as shown in Figure 5.

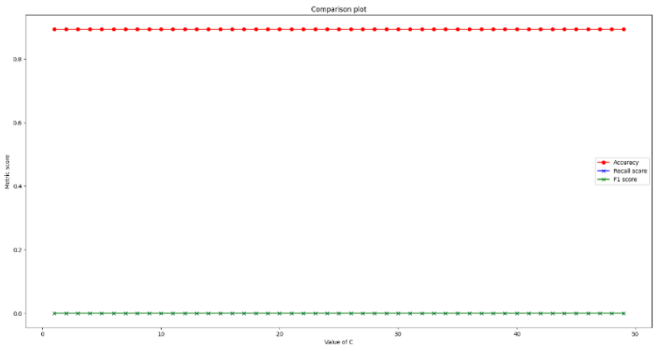

Figure 5 Graph of C values against evaluation metrics

Conclusions

Based on the results found in the KNN classification model, the optimal K value is 24, which yields a very low error rate and increases the accuracy of each class to 99 percentMeanwhile, in the SVM classification model, the C value is stable up to 50, so the margin of error for each class is very small. However, the value shown in the confusion matrix guesses table is that the KNN model is superior in guessing class 1 in earthquake classification. The SVM accuracy value obtained was 96 percent. A comparison of the accuracy results and the f1 score proves that not much data processing is needed to get an accurate value

Conflicts of interest

In accordance with our policy on Conflicts of interest please ensure that a conflicts of interest statement is included in your manuscript here. Please note that this statement is required for all submitted manuscripts. If no conflicts exist, please state that "There are no conflicts to declare".

Acknowledgements

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

References

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

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

[3] Sharma, R., Sinha, A. and Kautish, P., 2020. Examining the impacts of economic and demographic aspects on the ecological footprint in South and Southeast Asian countries. Environmental Science and Pollution Research, 27, pp.36970-36982.

[4] Fabinyi, M., Belton, B., Dressler, W.H., Knudsen, M., Adhuri, D.S., Aziz, A.A., Akber, M.A., Kittitornkool, J., Kongkaew, C., Marschke, M. and Pido, M., 2022. Coastal transitions: Small-scale fisheries, livelihoods, and maritime zone developments in Southeast Asia. Journal of Rural Studies, 91, pp.184-194.

[5] Jamil, F. Mukhaiyar, R. Dan Husnaini. (2020) "Kajian Literatur Rekonstruksi Mata Kuliah (Studi Kasus Mata Kuliah Pengolahan Sinyal Teknik Elektro URIP),"Jurnal Teknik Elektro Danvokasional, 6(2).

[6] Lorito, S., Romano, F. and Lay, T., 2022. Tsunamigenic

Published by: Lembaga Penelitian dan Pengabdian Masyarakat (LPPM)
Institut Teknologi Sumatera, Lampung Selatan, Indonesia

major and great earthquakes (2004–2013): source processes inverted from seismic, geodetic, and sea-level data. Complexity in Tsunamis, Volcanoes, and their Hazards, pp.247-298.

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

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

[9] Webb, A. R. (2003). Statistical pattern recognition. John Wiley & Sons.

[10] Brancato, A., Buscema, P. M., Massini, G., & Gresta, S. (2016). Pattern recognition for flank eruption forecasting: an application at Mount Etna volcano (Sicily, Italy). Open Journal of Geology.

[11] Sandri, L., Marzocchi, W., & Gasperini, P. (2005). Some insights on the occurrence of recent volcanic eruptions of Mount Etna volcano (Sicily, Italy). Geophysical Journal International, 163(3), 1203-1218.

[12] Curilem, M., Vergara, J., San Martin, C., Fuentealba, G., Cardona, C., Huenupan, F., & Yoma, N. B. (2014). Pattern recognition applied to seismic signals of the Llaima volcano (Chile): An analysis of the events' features. Journal of Volcanology and Geothermal Research, 282, 134-147

[13] San-Martin, C., Melgarejo, C., Gallegos, C., Soto, G., Curilem, M., & Fuentealba, G. (2010). Feature extraction using circular statistics applied to volcano monitoring. In Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 15th Iberoamerican Congress on Pattern Recognition, CIARP 2010, Sao Paulo, Brazil, November 8-11, 2010. Proceedings 15 (pp. 458-466). Springer Berlin Heidelberg.

[14] Novelo-Casanova, D. A., & Valdes-Gonzalez, C. (2008). Seismic pattern recognition techniques to predict large eruptions at the Popocatépetl, Mexico, volcano. Journal of Volcanology and Geothermal Research, 176(4), 583-590.)

[15] Burl, M. C., Fayyad, U. M., Perona, P., Smyth, P., & Burl, M. P. (1994). Automating the hunt for volcanoes on Venus.

[16] Ida, Y., Fujita, E., & Hirose, T. (2022). Classification of volcano-seismic events using waveforms in the method of k-means clustering and dynamic time warping. Journal of Volcanology and Geothermal Research, 429, 107616.

[17] Di Giuseppe, M. G., Troiano, A., Patella, D., Piochi, M., & Carlino, S. (2018). A geophysical k-means cluster analysis of the Solfatara-Pisciarelli volcano-geothermal system, Campi Flegrei (Naples, Italy). Journal of Applied Geophysics, 156, 44-54.

[18] Corradino, C., Amato, E., Torrisi, F., Calvari, S., & Del Negro, C. (2021). Classifying major explosions and paroxysms at Stromboli volcano (Italy) from space. Remote Sensing, 13(20), 4080

[19] Novianti, P., Setyorini, D., & Rafflesia, U. (2017). K-Means cluster analysis in earthquake epicenter clustering. International Journal of Advances in Intelligent Informatics, 3(2), 81-89.

[20] Rydelek, P. A., Davis, P. M., & Koyanagi, R. Y. (1988). Tidal triggering of earthquake swarms at Kilauea volcano, Hawaii. Journal of Geophysical Research: Solid Earth, 93(B5), 4401-4411.

[21] Petrosino, S., Cusano, P., & Madonia, P. (2018). Tidal and hydrological periodicities of seismicity reveal new risk scenarios at Campi Flegrei caldera. Scientific reports, 8(1),1-12.