Published *online* on the journal's web page : http://innovatics.unsil.ac.id

Innovation in Research of Informatics (INNOVATICS)

| ISSN (Online) 2656-8993 |

Natural Language Processing for Unstructured Data: Earthquakes Spatial Analysis in Indonesia Using Platform Social Media Twitter

Joko Ade Nursiyono [1], Rasya Khalil Gibran[2]

[1]Badan Pusat Statistik Provinsi Jawa Timur, Surabaya, Indonesia
[2]Polytechnic of Statistics STIS, Jakarta, Indonesia

[1]joko.ade@bps.go.id, [2]211911020@stis.ac.id

ARTICLE INFORMATION

Article History:
Received: 20 February 2023
Last Revision: 28 Macrh 2023
Published *Online*: 30 March 2023

KEYWORDS

Crawling,
Earthquake,
Natural Language Processing,
Spatial Analysis,
Unstructured Data,

CORRESPONDENCE

Phone: +6281244019483
E-mail: joko.ade@bps.go.id

ABSTRACT

As a country who had a high risk affected by the earthquake, social media have an important role. Besides to serving earthquake information, the spread of information on social media is so wide and fast. However, information on social media has a gap to reach validity and doesn't contain detailed information about spatial information. By leveraging crawling result data on Twitter, then data will be processed with Natural Language Processing (NLP), this research aims to proves about transformation of unstructured data into structured data with NLP for use on spatial analysis in Indonesia using data text on platform social media, Twitter. In addition, this research is also aims to reveal correlation between earthquake magnitude and earthquake frequency. The results proves that NLP can be used for spatial analysis with data text on Twitter related to earthquake. Besides that, the value of maximum magnitude is great significance to the earthquake frequency.

1. INTRODUCTION

Indonesia is one of the countries most at risk of being affected by earthquakes. The intensity of earthquakes throughout 2022 in Indonesia was recorded at 10,792 [1]. A total of 22 earthquake events were destructive earthquakes [2]. This is why earthquakes are considered the most dangerous type of disaster [3]. As an event of shaking the earth due to the spontaneous shifting or movement of the earth's skin [4], earthquake events are suspected by the presence of layer faults in the earth's crust or commonly termed plate tectonics [5].

In addition, earthquakes in Indonesia are also caused by volcanic activity in the Ring of Fire area. This is what makes Indonesia vulnerable to earthquakes [6]. The Ring of Fire region brings together three tectonic plates, namely the Indo-Australian Plate, the Eurasian Plate, and the Pacific Plate [7]. Please note, until now there has been no tool that can predict precisely and accurately when an earthquake will occur. For this reason, the existence of social media is crucial because it is able to provide information about disasters that are or are predicted to occur [8], especially earthquakes. In addition, social media is able to cause information quickly [9]. The speed of information dissemination is part of the Big Data element [10] as well as being a big challenge in how data can be collected in real time.

Among the efforts to improve accuracy and extract as many insights as possible from Big Data is the use of Natural Language Processing (NLP). With NLP, information sourced from unstructured data (e.g. text data) can be transformed into structured data. In fact, NLP can be a tool of converting unstructured data into spatial data containing geographic information. As research conducted by [11], with a multi-model coupling technique made from social media data shows the results of microblogs (Weibo) data containing a lot of information related to earthquakes. With NLP, text data containing earthquake information is used for the classification of Seismaesthesia as well as earthquake intensity. In addition, research related to earthquakes has also been published by [12] using BMKG data to predict the possibility of a possible earthquake.

However, research on earthquakes has a number of shortcomings, such as information on Microblog data that requires lag or time lag to reach valid status. In addition, the study has not provided a detailed explanation of the utilization of NLP in spatial analysis. To meet the research gap, this study was conducted by utilizing information from Twitter about the earthquake event in early 2023. In addition to focusing on the application of NLP in converting unstructured data into structured data, this study will also reveal data insights that have been structured for spatial earthquake analysis and the effect of earthquake magnitude on earthquake intensity. In addition, this study also provides discussions related to the results of spatial analysis as recommendations for earthquake disaster mitigation in Indonesia.

2. RELATED WORK

Studies of the use of NLP for spatial analysis have been produced by many studies. In Table 1 below, we present a number of these studies as a reference as well as a reference from this researcher.

TABLE 1. LITERATUR REVIEW

Research Title	Authors (Years)	Objective	Results
Classification of Seismaesthesia Information and Seismic Intensity Assessment by Multi-Model Coupling	Qingzhou et al. (2023)	Proposing a multi-model coupled seismic intensity assessment method based on BERT-TextCNN, constrained by seismaesthesia intensity attenuation model, and supplemented by method of ellipse-fitting inverse distance interpolation.	- Microbiolog contain a large amount of earthquake information. - The influence of subjectivity can be reduced using the seismaesthesia intensity attenuation model and method of ellipse-fitting inverse distance interpolation. - Accuracy of seismic intensity assessment based on coupled model is 70.81%.
Visualisasi Kualitas Penyebaran Informasi Gempa Bumi di Indonesia Menggunakan Twitter	Kirana et al. (2019)	Visualization to analyze quality of dissemination information in social media platform Twitter.	Quality of dissemination information in Twitter account BMKG is "Good" and effective. Data processing results show can be visualized from Twitter account BMKG.
Analisis Hubungan Magnitudo Gempa Bumi Terhadap Hasil Frekuensi Dominan Pada Rangkaian Gempa Aceh 2004, Yogyakarta 2006, Palu Dan Lombok 2018 Sebagai Upaya Mitigasi Bencana	Dwiyanti et al. (2020)	Analyze empirical relationship between earthquake magnitudes to earthquake dominan frequency in several earthquakes as Aceh earthquake in 2004, Yogyakarta earthquake in 2006, Palu and Lombok earthquake in 2018 as disaster mitigation.	Research proves that earthquakes magnitudes and earthquake dominan frequency have a significant relationship, and the bigger earthquake magnitudes then lower the earthquake dominan frequency.

Source: processed from various sources

3. METHODOLOGY

This research consists of several stages, ranging from collecting data, preprocessing data with NLP, data projection on spatial maps, and analysis.

3.1 Data Collection

This study used the Twitter Crawling technique to collect text data about earthquakes in Indonesia. The tool used to crawl Twitter is the R Studio package version 4.0.2 which is public.

To focus the topic on earthquakes, this study prepared an authenticated Twitter developer residency as well as the Crawling keywords "#Gempa" and "#BMKG". Then set a special language Indonesian because the scope of this research area is the occurrence of earthquakes in Indonesia (lang = "id").

The credentials were then used for Twitter Crawling 3 times. Twitter Crawling results obtained unstructured data in the form of mixed text containing both keywords (#Gempa and #BMKG) of 9,000 records and 90 variables during the time span of January 18 – February 9, 2023. Of these variables, this study only uses Twitter text variables to be processed into structured data. Here's a quick look at the text data gleaned from Twitter Crawling:

TABLE 2. TWITTER CRAWLED DATA

Screen_name	text
infoBMKG	#Gempa Mag:6.3, Kedlmn: 138 Km, 18-Jan-2023 07:34:46WIB, Lok: 0.07LS, 123.28BT (69 Km Tenggara BONEBOLANGO-GORONTALO), Tidak berpotensi tsunami #BMKG https://t.co/OiHiTwvX8x
infoBMKG	#Gempa Mag:6.3, 18-Jan-23 07:34:46WIB, Lok:0.07LS, 123.28BT (69 Km Tenggara BONEBOLANGO-GORONTALO), Kedlmn:138 Km, tdk berpotensi tsunami #BMKG https://t.co/tXrH48BRCA
infoBMKG	#Gempa Mag:6.3, Kedlmn: 138 Km, 18-Jan-2023 07:34:46WIB, Lok: 0.07LS, 123.28BT (Pusat gempa berada di laut 69 Km Tenggara Bone Bolango) #BMKG https://t.co/OiHiTwvX8x
infoBMKG	#Gempa Mag:7.1, 18-Jan-23 13:06:14WIB, Lok:2.80LU, 127.11BT (141 Km Tenggara MELONGUANE-SULUT), Kedlmn:64 Km, tdk berpotensi tsunami #BMKG https://t.co/UTm0VjPmTX
infoBMKG	"#Gempa Mag:4.4, 18-Jan-2023 13:33:02WIB, Lok:2.90LU, 127.12BT (131 Km Tenggara MELONGUANE-SULUT), Kedlmn:24 Km #BMKG

Source: processed with R 4.0.2

3.2 Data Preprocessing with NLP

Data preprocessing is the first stage in data preparation. As a type of Big Data, Twitter data needs to be treated in the form of preprocessing data. This is because data processing is an important stage in Big Data research [13]. NLP is one part of data preprocessing, especially for processing unstructured data in the form of text data. The application of NLP in this study uses a number of functions in the R package, namely gsub( ), filter( ), str_detect_all( ), str_replace_all( ), mutate( ), str_extract_all( ), anytime( ), lapply( ), and duplicated( ). In summary, the benefits of each of these functions are described in Table 3 below:

TABLE 3. TWITTER CRAWLED DATA

Package	Function Name	Application
dplyr	mutate( ); filter ( )	To create a new variable; Perform a data filter
stringr	str_detect_all( ); str_replace_all( ); str_extract_all( )	To filter data containing specific text; to perform the replacement of certain characters (text) into other characters; for specific text extraction in a set of text
base R	gsub( ); lapply( ); duplicated( )	Remove specific characters; apply functions to text frame data; delete duplicate text data
anytime	anytime( )	To change character text to datetime

Source: R 4.0.2

Extraction of latitude and longitude coordinates from the text of this study uses a combination of several functions, namely mutate (), str_extract_all (), then continued with str_detect( ) and str_extract( ). At this stage, 2 conditions are applied, remembering that for text with the labels "LS" and "BB" it is necessary to add a minus sign so that it can be projected correctly on the spatial map. At a glance, the results of extracting coordinates from the text are listed in Table 4 below:

TABLE 4. TWITTER CRAWLED DATA

Before	After (Latitude, Longitude)
#Gempa Mag:6.3, Kedlmn: 138 Km, 18-Jan-2023 07:34:46WIB, Lok: 0.07LS, 123.28BT (69 Km Tenggara BONEBOLANGO-GORONTALO), Tidak berpotensi tsunami #BMKG https://t.co/OiHiTwvX8x	(-0.07, 123.28)
#Gempa Mag:3.8, 18-Jan-2023 14:50:46WIB, Lok:2.87LU, 127.19BT (137 Km Tenggara MELONGUANE-SULUT), Kedlmn:16 Km #BMKG Disclaimer:Informasi ini mengutamakan kecepatan, sehingga hasil pengolahan data belum stabil dan bisa berubah seiring kelengkapan data https://t.co/3VGwDAghRZ	(2.87, 127.19)
#Gempa Mag:3.1, 18-Jan-2023 19:26:02WIB, Lok:2.44LS, 140.71BT (100 Km BaratLaut KEEROM-PAPUA), Kedlmn:10 Km #BMKG Disclaimer:Informasi ini mengutamakan kecepatan, sehingga hasil pengolahan data belum stabil dan bisa berubah seiring kelengkapan data https://t.co/zDVhFzSulL	(-2.44, 140.71)
#Gempa Mag:7.1, 18-Jan-23 13:06:14WIB, Lok:2.80LU, 127.11BT (141 Km Tenggara MELONGUANE-SULUT), Kedlmn:64 Km, tdk berpotensi tsunami #BMKG https://t.co/UTm0VjPmTX	(2.80, 127.11)
"#Gempa Mag:4.4, 18-Jan-2023 13:33:02WIB, Lok:2.90LU, 127.12BT (131 Km Tenggara MELONGUANE-SULUT), Kedlmn:24 Km #BMKG	(2.90, 127.12)

Source: processed with R 4.0.2

Time extraction of text data containing earthquake topics is performed by uniformizing the time format first. Because, some texts have different time formats, some are written "DD MM YY" and some are in the form of "DD-MM-YY". The uniformity of the time format is set in the form "DD-MM-YY".

The extraction of the earthquake time is then carried out by detecting text ending in the word "WIB" or Western Indonesia. Thus, the previously mixed text into a time variable with the format "YY-MM-DD HH:MM: SS". In this condition, adjustments are also made to the year format because there is also a time text that only writes the year by 2 digits, for example 2024 is only written 24. Therefore, the 2-digit year text is converted to 4 digits using str_replace_all( ). The extraction results are shown in Table 5 below:

TABLE 5. TIME EXTRACTION RESULTS FROM TEXT

Before	After
#Gempa Mag:6.3, Kedlmn: 138 Km, 18-Jan-2023 07:34:46WIB, Lok: 0.07LS, 123.28BT (69 Km Tenggara BONEBOLANGO-GORONTALO), Tidak berpotensi tsunami #BMKG https://t.co/OiHiTwvX8x	2023-01-18 07:34:46
#Gempa Mag:3.8, 18-Jan-2023 14:50:46WIB, Lok:2.87LU, 127.19BT (137 Km Tenggara MELONGUANE-SULUT), Kedlmn:16 Km #BMKG Disclaimer:Informasi ini mengutamakan kecepatan, sehingga hasil pengolahan data belum stabil dan bisa berubah seiring kelengkapan data https://t.co/3VGwDAghRZ	2023-01-18 14:50:46

Source: processed with R 4.0.2

This stage is done by detecting text that begins with the text "Mag:". However, some texts have a format that says "Magnitude: (with spaces)". Therefore, uniformization of the format is carried out first. The results of the extraction of earthquake megnitude from the text are listed in Table 6 below:

TABLE 6. EARTHQUAKE MAGNITUDE EXTRACTION RESULTS FROM TEXT

Before	After
#Gempa Mag:6.3, Kedlmn: 138 Km, 18-Jan-2023 07:34:46WIB, Lok: 0.07LS, 123.28BT (69 Km Tenggara BONEBOLANGO-GORONTALO), Tidak berpotensi tsunami #BMKG https://t.co/OiHiTwvX8x	6.3
#Gempa Mag:3.8, 18-Jan-2023 14:50:46WIB, Lok:2.87LU, 127.19BT (137 Km Tenggara MELONGUANE-SULUT), Kedlmn:16 Km #BMKG Disclaimer:Informasi ini mengutamakan kecepatan, sehingga hasil pengolahan data belum stabil dan bisa berubah seiring kelengkapan data https://t.co/3VGwDAghRZ	3.8

Source: processed with R 4.0.2

Earthquake depth extraction is performed by uniformizing the format first. Because sometimes the format of the text followed by the depth information is written "(space) Depth:". To that end, the uniformization of text containing depth information is changed entirely to "Kedlmn:" to then be extracted with the str_extract_all()

function. The results of the earthquake depth extraction from the text are listed in Table 7 below:

TABLE 7. EARTQUAKE DEPTH EXTRACTION RESULTS FROM TEXT

Before	After
#Gempa Mag:6.3, Kedlmn: 138 Km, 18-Jan-2023 07:34:46WIB, Lok: 0.07LS, 123.28BT (69 Km Tenggara BONEBOLANGO-GORONTALO), Tidak berpotensi tsunami #BMKG https://t.co/OiHiTwvX8x	138
#Gempa Mag:3.8, 18-Jan-2023 14:50:46WIB, Lok:2.87LU, 127.19BT (137 Km Tenggara MELONGUANE-SULUT), Kedlmn:16 Km #BMKG Disclaimer:Informasi ini mengutamakan kecepatan, sehingga hasil pengolahan data belum stabil dan bisa berubah seiring kelengkapan data https://t.co/3VGwDAghRZ	16

Source: processed with R 4.0.2

3.3 Data Projection on Spatial Maps

The conversion results from unstructured data (text) to structured data (excel) are then projected on a map of Indonesia with an extension *shp. Some of the packages used in this projection process consist of ggplot2, maps, and gganimate. In addition, to increase insights, the results of text data extraction also display the presence of earthquakes, namely on land or at sea.

3.4 Analysis

The analysis used in this study consists of descriptive analysis and inference. Descriptive analysis is used to describe insights into transformations from text data to structured data resulting from the application of NLP, including Pearson correlations between formed variables. Meanwhile, the inference analysis of this study revealed the relationship between the frequency and magnitude of earthquakes which were sorted into 3 types, namely the minimum magnitude (minmag), average magnitude (avemag), and maximum magnitude (maxmag), as the famous formulation proposed by seismologists Gutenberg and Richter in 1941 [14] follows:

$$lg\, N = a_0 - b_0 M \qquad (1)$$

Where is the magnitude of the earthquake; is the frequency of occurrence of earthquakes; and an intercept and coefficient: a_0, b_0

The interrelation of such formulations translates as regression. According to [15], regression shows a causality (causality) relationship between a free (independent) variable and a non-free variable (dependent). The results of this regression were then tested for feasibility using several classical assumption tests, namely the normality test (Kolmogorov Smirnov test), the non-autocorrelation test (Durbin Watson test), and the homoskedasticity test (Breush Pagan test). If p-value > 0.05 then it is said that the residual regression has met the classical assumption.

The regression models formed in this study are 3 models. Then the selection of the best model is carried out based on the significance value of the partial test (t test), simultaneous test (F test), as well as the highest R square value.

4. RESULT AND DISCUSSION

Modeling the frequency of earthquakes on earthquake magnitudes which are divided into maximum magnitude, minimum magnitude, and average earthquake magnitude is carried out to determine the influence and significance of the variables tested. The assumptions made on this regression model are normality, homogeneity, and nonautocorrelation. Consider the table 8 follows:

TABLE 8. P-VALUE OF CLASSIC ASSUMPTION REGRESSION

Model	Classic Assumption Test		
	Normality	Homoskedasticity	Nonautocorrelation
Log Freq ~ Min-Mag	1	0.8365	0.4815
Log Freq ~ Ave-Mag	0.9979	0.2792	0.2008
Log Freq ~ Max-Mag	0.9979	0.985	0.4055

Source: processed with R 4.0.2

The table shows that all three models meet classical assumptions so that regression modeling can be carried out. Non-multicholinearity testing is not done because it uses simple linear regression. The results of the formation of the three models are listed in the following table 9:

TABLE 9. REGRESSION MODEL RESULTS

Model	Estimate Parameters		F-statistics	p-value	R^2
	Intercept	Beta			
Log Freq ~ Min-Mag	6.2548	-1.3176	10.78	0.0031**	0.3099
Log Freq ~ Ave-Mag	0.08551	0.7808	1.728	0.2011	0.0672
Log Freq ~ Max-Mag	-2.0213	1.0209	32.73	6.80E-06**	0.5769

Note: **) Significant at 5% level
Source: processed with R 4.0.2

Based on these results, models that have a significant influence and have met the classical assumption test on earthquake frequency are models with a minimum earthquake magnitude and a maximum earthquake magnitude. However, when compared according to the significance of F-statistics and R-square, the maximum magnitude model is the best model. Thus, the selected model can be written as follows:

$$lg\, N = -2.0213 + 1.0209 MaxMag \qquad (2)$$

Based on the selected model, it is interpreted that every time an earthquake occurs, the maximum magnitude increases by 1 Richter Scale (SR), the frequency of earthquakes that occur will increase by 1.0209 assuming other variables are of constant value. In addition, by reviewing the magnitude of the R-square of 0.5769, it can be interpreted that 57.69% of the influence on the frequency of earthquakes can be explained by the maximum magnitude model. These results are in accordance with research [16] that the frequency of earthquakes is influenced by the magnitude of earthquakes, in particular the maximum magnitude. This occurs due to seismic waves that are affected by soil structure, soil layer, and soil rigidity. In addition, a correlation test was also formed with Pearson's Product-Moment Correlation as follows:

Table 10. Correlation Of Magnitude With Earthquake Frequency

Pearson's Product-Moment Correlation	Max-Mag	Min-Mag	Ave-Mag
t-statistics	5.064	-1.520	-1.515
p-value	3.54E-05**	0.1416	0.1429
Correlation	0.719	-0.296	0.295

Note: **) Significant at 5% level
Source: processed with R 4.0.2

The results showed that the model that had a significant relationship to the frequency of earthquakes was the model using a maximum magnitude with a t-statistics value of 5,064 (p-value < 0.05). The relationship in the model is 0.719, which is interpreted as the direction of the relationship between the maximum earthquake magnitude and the frequency of earthquakes with a positive and statistically significant earthquake frequency. To facilitate the understanding of the relationship between frequency and magnitude, a correlation matrix is formed as follows:

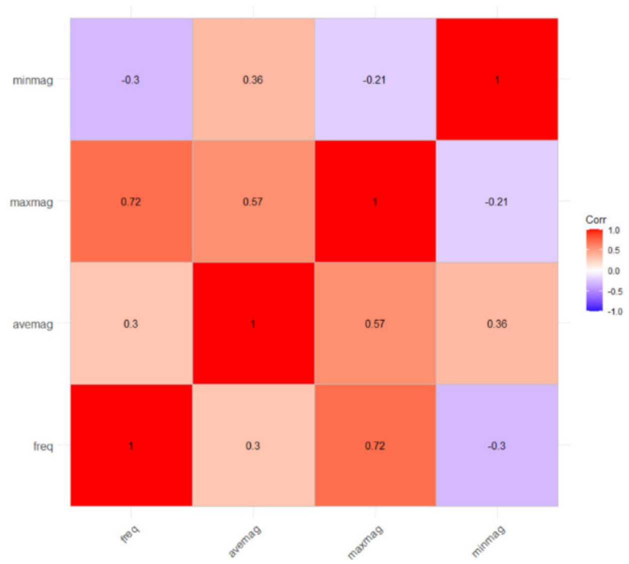

Figure 1. Correlation Matrix of Earthquake Frequency with Earthquake Magnitude
Source: processed with R 4.0.2

The correlation matrix explains the closeness and direction of the relationship between earthquake frequency, average earthquake magnitude, minimum earthquake magnitude, and maximum earthquake magnitude. The variable that has the strongest relationship is the relationship between the earthquake frequency variable and the maximum earthquake magnitude compared to the relationship between other variables. This result is different from the research by [17] which shows that the greater the magnitude of the earthquake, the smaller the dominant frequency will be or can be said to have an inversely proportional correlation. This can happen because earthquakes of great magnitude are likely to cause aftershocks so that the frequency of earthquakes can increase.

Discussing the frequency of earthquakes, it was found that there are other characteristics related to magnitude, namely earthquake depth. Here is a table of the relationship between magnitude and earthquake depth.

Table 11. Correlation Of Magnitude With Earthquake Depth

Pearson's Product-Moment Correlation	Depth-Magnitudo
t-statistics	2.436
p-value	1.50E-02**
Correlation	0.0796

Note: **) Significant at 5% level
Source: processed with R 4.0.2

Reviewing table 11, magnitude has a significant relationship to earthquake depth. However, although the relationship is significant, it turns out that the relationship between the magnitude and depth of the earthquake is not very strong, and it can even be said to be weak. This result is proven because there are some earthquakes that have a relatively low magnitude but a deep earthquake depth, and there are also some earthquakes that have a relatively high magnitude but have an earthquake depth that is not too deep. There are many theories that shallow earthquakes are more destructive, but in general this happens because the epicenter is close to the surface so that the vibrations are more strongly felt. Speaking of the results we have gotten so far, earthquakes are closely related to the location as well as the time when the earthquake occurred. Thus, exploration is carried out using pie chart visualization to find out the location and time offrequent earthquakes.

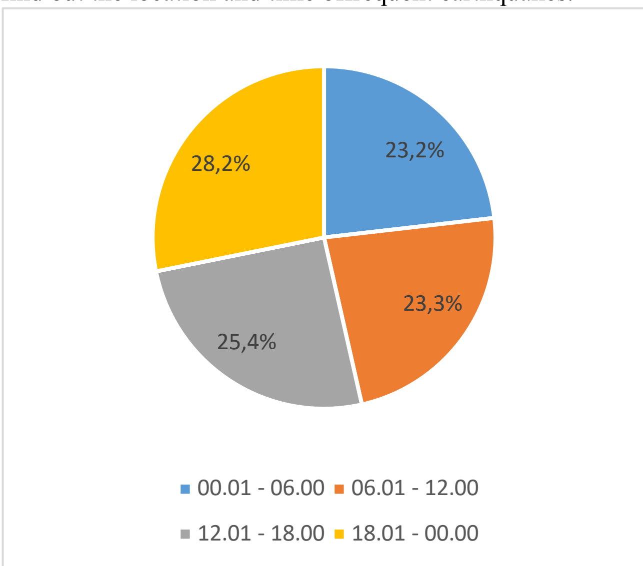

Figure 2. Pie Chart of Earthquake Times
Note: Reference times in Western Indonesia (WIB)
Source: processed with R 4.0.2

TABLE 12. EARTHQUAKE EVENTS ACCORDING TO TIME, PERCENTAGE, AND NUMBER OF EVENTS

Time Event*	00.01 - 06.00	06.01 - 12.00	12.01 - 18.00	18.01 - 00.00
Percentage Event	23.2%	23.3%	25.4%	28.2%
Total Event	216	217	237	263

Note: Reference times in Western Indonesia (WIB)
Source: processed with R 4.0.2

Based on the visualization of figure 12, almost every time an earthquake occurs evenly. However, the most frequent time for earthquakes is 18:01 to 00:00. There are no studies that can confirm when the most frequent times of earthquakes occur because earthquakes are difficult to predict. This is because the earthquake process occurs suddenly depending on the activity of the movement of the earth's plates, the cracking of the earth's plates, or the presence of volcanic activity. In addition, this finding at least provides an early warning of earthquake disaster mitigation so that active vigilance is carried out 24 hours as well as the pursuit of earthquake disaster mitigation strategies at times when the community is resting (sleeping), namely 00.00-06.00 or 12.01-18.00.
To add insight, the results of NLP utilization in this earthquake data are also visualized according to the location of the earthquake event in categories, namely land and sea with the results shown in figure 3.

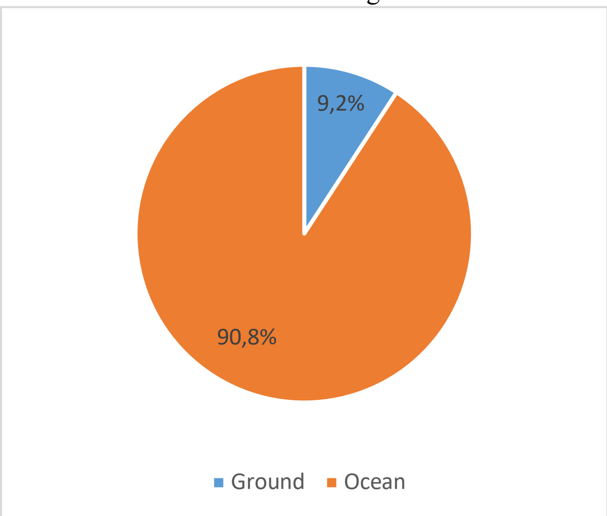

Figure 3. Pie Chart of Earthquake Locations
Source: processed with R 4.0.2

TABLE 13. EARTHQUAKE EVENTS ACCORDING TO LOCATION EVENTS

Event Location	Ground	Ocean
Percentage Event	9.2%	90.8%
Total Event	86	847

Source: processed with R 4.0.2

Based on visualization figure 3 shows that earthquakes in Indonesia often occur in sea areas rather than land. The results in the 13 m table show that the number of earthquakes occurring at sea is almost 10 times the number of earthquakes on land. However, quoting from [18], that earthquakes on the ground can cause damage to more casualties because they are close to residential areas. However, basically, all earthquake sites can be dangerous if not properly anticipated. Thus, visualization is carried out using land locations by considering the existence of residential areas based on a predetermined time.

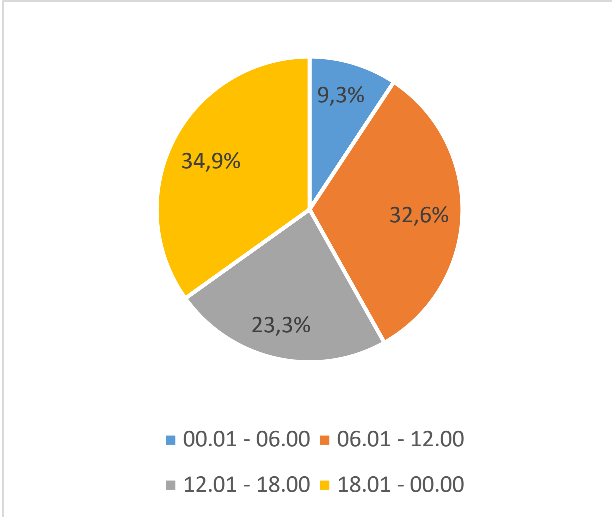

Figure 4. Pie Chart of Earthquake Times on Land
Note: Reference times in Western Indonesia (WIB)
Source: processed with R 4.0.2

TABLE 14. EARTHQUAKE EVENTS ON LAND ACCORDING TO TIME

Time Event*	00.01 – 06.00	06.01 – 12.00	12.01 – 18.00	18.01 – 00.00
Percentage Event	9.3%	32.6%	23.3%	34.9%
Total Event	8	28	20	30

Note: Reference times in Western Indonesia (WIB)
Source: processed with R 4.0.2

Based on figure 4 and table 14, the time when the most earthquakes occur on land is 18:01 to 00:00 with a frequency of 30 earthquakes. These results can be used as material for earthquake disaster mitigation at these times, especially at 18.01-00.00.

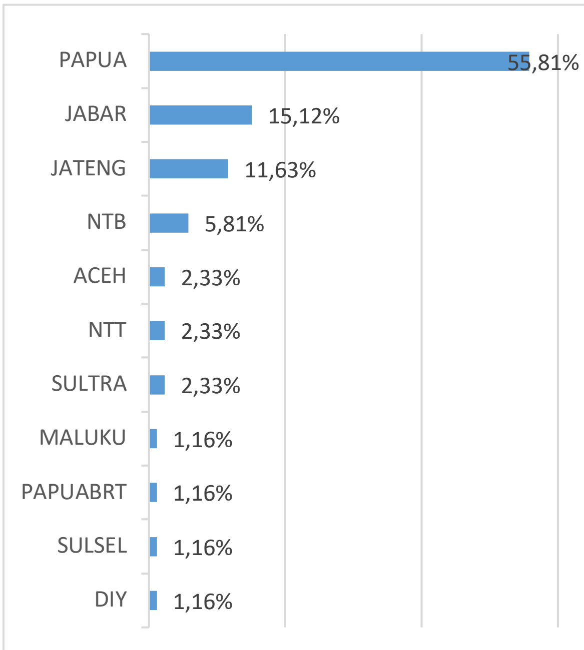

Figure 5. Percentage of Earthquake Events on Land Based on Provinces
Source: processed with R 4.0.2

Based on the findings of the number of earthquakes on land, Papua province is the area with the most frequent earthquakes on land, recorded as many as 48 earthquake events. Then followed by the provinces of West Java and Central Java. This finding needs to be the focus of the government so that the preparation of a mitigation strategy for earthquakes that occur in Indonesia is prioritized in these three regions by considering population density, building capacity, and soil and rock structures.

In addition to exploring the timing and location of earthquakes, exploration of certain patterns of earthquakes is also carried out according to the central province or around the earthquake event. As a result, earthquakes that occurred starting in the Maluku region immediately occurred in the Sumatra region, continued to occur on the island of Sumatra, then occurred in the Sulawesi region, and again occurred in the Maluku region. These results show that no specific patterns of earthquake events were found in Indonesia. But what is clear is that a number of earthquakes that occurred in Indonesia showed the active movement of the Pacific Plate and the India-Australia Plate.

5. CONCLUSIONS

Based on the results of the study, several points of conclusion can be drawn as follows:

1. A simple linear regression modeling of the earthquake frequency log that has a significant influence as the best model is the model with a maximum magnitude that can explain the information in the model as much as 57.69% and the rest is explained by other variables.
2. Models with a maximum magnitude of0.72 have a significant correlation so they can be categorized as quite strong.
3. The relationship between earthquake magnitude and earthquake depth also has a significant, but very weak, relationship.
4. Based on the visualization presented, earthquakes occur in many sea areas and the time of occurrence of earthquakes is quite even. In addition, the time of the most earthquakes on the ground occurs at 18.01 to 00.00 WIB.
5. The province with the most earthquakes according to BMKG Twitter data is Papua province with 48 events.
6. There is no specific pattern in the occurrence of earthquakes in Indonesia.

REFERENCES

[1] BeritaSatu.com, "BMKG: Indonesia Diguncang 10.792 Kali Gempa Sepanjang 2022," *beritasatu.com*, 2022.

[2] N. W. Koesmawardhani, "BMKG: Ada 10.792 Gempa di RI Selama 2022, 22 di Antaranya Merusak," *https://www.detik.com/edu/edutainment/d-6489347/bmkg-ada-10792-gempa-di-ri-selama-2022-22-di-antaranya-merusak* , 2022.

[3] R. Tehseen, M. S. Farooq, and A. Abid, "Earthquake Prediction Using Expert Systems: A Systematic Mapping Study," *Sustainability*, vol. 12, no. 6, p. 2420, Mar. 2020, doi: 10.3390/su12062420.

[4] B. D. W. Sari, D. Djayus, S. Supriyanto, and B. Hendrawanto, "Penentuan Nilai Parameter Gempabumi Menggunakan Metode Geiger dan Hukum Laska pada Pulau Lombok," *GEOSAINS KUTAI BASIN*, vol. 5, no. 1, Feb. 2022, doi: 10.30872/geofisunmul.v5i1.706.

[5] A. H. Santoso, O. D. Wahyuni, T. Tarcisia, and D. Denny, "Penapisan Hipertensi melalui Pelayanan Pengukuran Tekanan Darah bagi Warga Desa Kampung Baros Ciherang Pacet Paska Bencana Gempa Cianjur," *Nusantara*, vol. 3, no. 1, Jan. 2023.

[6] D. L. Ramatillah, D. A. C. Agustin, S. E. Susilowati, R. Astiani, A. Rofii, and S. Lukas, "PENANGGULANGAN SANITASI DAN PENYULUHAN UNTUK MENINGKATKAN KESEHATAN MASYARAKAT DESA BENJOT PASCA GEMPA CIANJUR," *Berdikari*, vol. 6, no. 1, 2023.

[7] R. Akbar, R. Darman, F. Marizka, J. Namora, and N. Ardewati, "Implementasi Business Intelligence Menentukan Daerah Rawan Gempa Bumi di Indonesia dengan Fitur Geolokasi," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 4, no. 1, p. 30, Jun. 2018, doi: 10.26418/jp.v4i1.25518.

[8] S. Fahriyani, D. Harmaningsih, and S. Yunarti, "PENGGUNAAN MEDIA SOSIAL TWITTER UNTUK MITIGASI BENCANA DI INDONESIA," *Jurnal Sosial dan Humaniora*, vol. 4, no. 2, 2019.

[9] K. K. Rafiah and D. H. Kirana, "Analisis Adopsi Media Sosial Sebagai Sarana Pemasaran Digital Bagi UMKM Makanan dan Minuman di Jatinangor," *Jesya (Jurnal Ekonomi & Ekonomi Syariah)*, vol. 2, no. 1, pp. 188–198, Jan. 2019, doi: 10.36778/jesya.v2i1.45.

[10] M. Aiello, C. Cavaliere, A. D'Albore, and M. Salvatore, "The Challenges of Diagnostic Imaging in the Era of Big Data," *J Clin Med*, vol. 8, no. 3, p. 316, Mar. 2019, doi: 10.3390/jcm8030316.

[11] Q. Lv, W. Liu, R. Li, H. Yang, Y. Tao, and M. Wang, "Classification of Seismaesthesia Information and Seismic Intensity Assessment by Multi-Model Coupling," *ISPRS Int J Geoinf*, vol. 12, no. 2, p. 46, Jan. 2023, doi: 10.3390/ijgi12020046.

[12] D. P. Utomo and B. Purba, "Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia," *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, vol. 1, p. 846, Sep. 2019, doi: 10.30645/senaris.v1i0.91.

[13] A. Phinyomark, R. N. Khushaba, E. Ibáñez-Marcelo, A. Patania, E. Scheme, and G. Petri, "Navigating features: a topologically informed chart of electromyographic features space," *J R Soc Interface*, vol. 14, no. 137, p. 20170734, Dec. 2017, doi: 10.1098/rsif.2017.0734.

[14] D. Marchetti *et al.*, "Quick Report on the ML = 3.3 on 1 January 2023 Guidonia (Rome, Italy) Earthquake: Evidence of a Seismic Acceleration," *Remote Sens (Basel)*, vol. 15, no. 4, p. 942, Feb. 2023, doi: 10.3390/rs15040942.

[15] J. A. Nursiyono and P. P. H. Nadeak, *Setetes Ilmu Regresi Linier: Untuk Penelitian*, 1st ed. 2015.

[16] J. Nia Shohaya, U. Chasanah, A. Mutiarani, L. Wahyuni P, and M. Madlazim, "SURVEY DAN ANALISIS SEISMISITAS WILAYAH JAWA TIMUR

BERDASARKAN DATA GEMPA BUMI PERIODE 1999-2013 SEBAGAI UPAYA MITIGASI BENCANA GEMPA BUMI," *Jurnal Penelitian Fisika dan Aplikasinya (JPFA)*, vol. 3, no. 2, p. 18, Dec. 2013, doi: 10.26740/jpfa.v3n2.p18-27.

[17] N. E. Dwiyanti *et al.*, "ANALISI HUBUNGAN MAGNITUDO GEMPA BUMI TERHADAP HASIL FREKUENSI DOMINAN PADA RANGKAIAN GEMPA ACEH 2004, YOGYAKARTA 2006, PALU DAN LOMBOK 2018 SEBAGAI UPAYA MITIGASI BENCANA," *Jurnal Meteorologi Klimatologi dan Geofisika*, 2021.

[18] B. Setiawan, "Mengenali Gempa dari Letak Kejadian, Darat dan Laut," *https://tekno.tempo.co/read/1552195/mengenali-gempa-dari-letak-kejadian-darat-dan-laut*, 2023.

AUTHORS

First Author

Joko Ade Nursiyono, working in the Cross-Sectoral Statistical Analysis and Big Data team of BPS Provinsi Jawa Timur since 2022. The author is an alumnus of the Sekolah Tinggi Ilmu Statistik (STIS) Jakarta in 2013 with 34 books. Some of the book titles are Pengantar Statistika Dasar, Kalkulus Dasar, Saripati Aljabar Linier, Pengantar Data Mining dengan R Studio, Visualisasi Data dengan Tableau, Setetes Ilmu Regresi Linier, dan Kompas Teknik Pengambilan Sampel.

Second Author

Rasya Khalil Gibran, who can be called Gibran, is an only child living in Surabaya. Gibran is also studying at the Polytechnic of Statistics STIS since 2019. Gibran has participated in various student activities, especially in public relations. In addition, Gibran has also participated in various organizations both internally and externally on Polytechnic of Statistics STIS. One of the organizations that Gibran have been participated in are the student representative council as well as the legislative organizations on Polytechnic of Statistics STIS.