Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)
17/1 (2024), 77-87. DOI: http://dx.doi.org/10.21609/jiki.v17i1.1203

Predicting Earthquake Magnitudes in Indonesia: Exploring the Potential of the Prophet Algorithm

Susi Nurindahsari, Slamet Wiyono, Dairoh

Informatics Engineering, Politeknik Harapan Bersama, Kota Tegal, Indonesia.

sarindahsn23@gmail.com, oc_slametwiyono@poltektegal.ac.id, dairoh@poltektegal.ac.id.

Abstract

Research on earthquakes has been extensively conducted by previous studies using various methods and specific discussions. Similarly, research to predict the magnitude of earthquakes that will occur in the future has also been conducted. This study employs the Prophet algorithm to test its capability in predicting a case study's magnitude using data with numerous missing values and outliers. The study is conducted without transformation and with Box-Cox and log-transformations. Transformations are applied to handle outliers. The results indicate that across the three experiments, the difference between the predicted and actual data ranges from 0.1 to 0.5 or even more. Performance metrics reveal that the log-transform is superior to the other two experiments, with a smaller MAE of 0.27 and a MAPE of 5.96%. Nevertheless, the use of the Prophet algorithm in this case study needs further investigation with different treatments to achieve more accurate results.

Keywords: *Prophet, Magnitude, Prediction, Box-Cox Transformation, Log-Transformation.*

1. Introduction

Based on its geographical location, Indonesia is an archipelago that has great earthquake potential because it is located along a seismically turbulent pathway, the Pacific Ring of Fire. Based on earthquake data taken from the United States Geological Survey (USGS) from 2017 to date, the 3 provinces in Indonesia that frequently occur or are prone to earthquakes are North Maluku, Maluku, and Papua. Provinces that rarely experience earthquakes are North Kalimantan, South Kalimantan, West Kalimantan, and Jakarta [1]. Earthquakes that occur have a variety of hypocentral depths, epicenters, and magnitudes. Recording the magnitude of these events is important to determine whether the earthquake has the potential for large damage, tsunamis for earthquakes that occur at or near the sea, and for recording historical data to see the calculation cycle of future earthquakes. Many earthquakes that have occurred are followed by aftershocks, both small and large earthquakes that are similar to the previous earthquake or even larger. Repeated earthquakes cause buildings, infrastructure, and soil to become increasingly weakened [2]. These effects certainly cause many victims and financial losses to society [3]. Likewise, earthquakes have an important impact on the ecological environment in the affected areas, whose impact on the natural ecological environment is enormous. Earthquake disasters not only cause great damage to the original ecological environment system, but even form a new ecological system [4]. In fact, not only the effects on land, if a large earthquake occurs under the sea, it may cause a tsunami.

Earthquakes are destructive natural disasters that occur almost without prior warning and are certainly unavoidable [3]. his disaster also cannot be known exactly when it will occur, where the location with its depth, and the strength of the magnitude it carries. Many researchers are currently trying to predict the occurrence of earthquakes before this disaster actually occurs, especially for the Government of Indonesia and related parties to anticipate and minimize the impact of the consequences based on the level of earthquake strength measured based on the Richter Scale (SR) value for local earthquakes and magnitude for wider coverage [5]. In addition to anticipating the impact of earthquakes, the importance of recording earthquake data needs to be done so that it is well recorded because this earthquake data will often be used, both for development and for the environment [6].

Many methods have been used to create an algorithm to predict the occurrence of earthquakes based on existing historical data. However, the results of this prediction certainly need to be calculated in more detail based on the occurrence cycle of each earthquake that occurs in the area. In this case, this research was conducted using the Prophet algorithm. Prophet is known to be good at handling missing values and outliers in time series data prediction [7]. Therefore, this study aims to test whether Prophet is able to produce accurate earthquake prediction values with 'hollow data' in the sense that only at certain times it occurs (not every time) without special treatment of the data (the magnitude data in this case study certainly has missing values and outliers) or whether optimization is still needed with data transformation.

In this study, three experiments were conducted, namely without transformation, with box-cox transformation, and with log-transformation. Box-cox and log-transform were chosen in this treatment to see how well Prophet handles data that has been adjusted to reduce the effects of outlier data and correct asymmetrical data distributions to be symmetrical or normal. Box-cox characteristics that can handle data with a wider range of values by calculating the change in values within a given data range, measured by lambda ($\boldsymbol{\lambda}$) values as used in [8] to evaluate the phenotypic and genomic background of litter size variability which showed that it is very important to perform Box-Cox Transformation for skewness data in order to correctly describe the phenotypic and genomic properties of litter size variability in Landrace pigs and used to demonstrate the use of Box-Cox Transformation for skewness data. cox for skewed data and predicted the results in the original scale using the Cost of Pyelonephritis in Type-2 Diabetes (COPID) data by [9] which resulted in the conclusion that the box-cox transformation can almost transform skewed data to normal. Meanwhile, the log transform converts skewness data to normal by taking the natural logarithm of each data point, such as in a study using healthy male and female health record data which concluded that the log transform almost succeeded in making the data close to normal [10].

2. Prophet

FBProphet or commonly called Prophet is a library available in R and Python that is used for time series data analysis and prediction. It was developed by Facebook and released as an open source library. Prophet focuses on additive models that are able to handle data that has non-linear trends and seasonal effects. This approach allows Prophet to address many of the problems common in predictive analysis. Prophet has become a popular algorithm in various fields, including economics, finance, marketing, and others. With its innovative approach and flexibility, Prophet allows users to perform predictive analysis easily and obtain accurate results [11]. Therefore, Prophet is used in this study to find out how good this algorithm is in producing earthquake magnitude prediction data where the data used is time series data which is not every time (in this case days) recorded for each earthquake in a particular region.

Prophet is a powerful and fast open-source time-series model developed by Facebook, the cool thing is that Prophet can handle missing values and outliers in forecasting [7]. This earthquake data does not have a seasonal pattern like fruits that grow in a certain time. The Prophet equation is as given in equation (1).

$$\boldsymbol{y(t) = g(t) + s(t) + h(t) + \varepsilon(t)} \quad (1)$$

$\boldsymbol{g(t)}$ represents the trend function responsible for modeling non-periodic changes in earthquake data, $\boldsymbol{s(t)}$ represents seasonality occurring daily, weekly, yearly, $\boldsymbol{h(t)}$ represents holidays occurring at any given time, and $\boldsymbol{\varepsilon(t)}$ an error term not accommodated by the model. The regressor used by Prophet for fitting to a saturating growth model (non-linear) or piecewise linear model (linear) as a component is time, by default Prophet uses fitting data to a linear model and can be changed to a non-linear model as needed by changing the arguments in the model [12]. Equation (2) is a non-linear equation.

$$\boldsymbol{g(t) = \frac{C}{1 + \exp(-k(t - m))}} \quad (2)$$

$\boldsymbol{C}$ is the carrying capacity, $\boldsymbol{k}$ is the growth rate, and $\boldsymbol{m}$ is the offset parameter. The offset parameter needs to be adjusted to connect the segment endpoints when the rate $\boldsymbol{k}$ is adjusted [13]. Then, the equation of the piecewise linear model is in equation (3).

$$\boldsymbol{g(t) = (k + a(t)\,\delta)t(m + a(t)^T\gamma)} \quad (3)$$

$\boldsymbol{k}$ is the growth rate, δ is the adjustment rate, and $\boldsymbol{m}$ is the offset parameter. Prophet uses Fourier series to forecast the seasonality effect, and the seasonality model is specified as a periodic function of $\boldsymbol{t}$. The seasonality effect can be represented as in equation (4).

$$\boldsymbol{s(t) = \sum_{n=1}^{N}\left(a_n \cos\cos\left(\frac{2\pi nt}{P}\right) + b_n sin(\frac{2\pi nt}{P})\right)} \quad (4)$$

$\boldsymbol{P}$ is the period, for annual seasonality $\boldsymbol{P}$ = 354.25 and for weekly seasonality $\boldsymbol{P}$ = 7. Holidays and special events do not follow any period cycle, then to model the holiday function $\boldsymbol{h(t)}$ can provide a specific matrix containing the dates and details of these holidays [12]. $\boldsymbol{N}$ for a given value, to fit the seasonality model, the parameters $\boldsymbol{a_1}$, $\boldsymbol{a_2}$, ..., $\boldsymbol{a_n}$ dan $\boldsymbol{b_1}$, $\boldsymbol{b_2}$, ..., $\boldsymbol{b_n}$ need to be estimated [7]. The Prophet model considers the effect of different holidays in a year on the change of time series trend as an independent model and assigns a separate dummy variable to each model. Equation 5 is the holiday equation.

$$\boldsymbol{h(t)} = \mathbf{Z(t)}\boldsymbol{k} = \sum_{\mathbf{i=1}}^{\boldsymbol{L}} \boldsymbol{k_i} \times \mathbf{1}_{\{\boldsymbol{t \in D_i}\}} \tag{5}$$

Where $\boldsymbol{k_i}$ represents the effect of holidays on forecasting values and $\boldsymbol{D_i}$ represents dummy variables [14].

Prophet has its own column name requirements for ease of forecasting, namely the date column named "ds" or date stamp (in datetime format) and the forecasted value column with "y" with forecasting measurements that must be in numeric values [15]. If there are more columns of values to be forecasted, it can be customized, but both columns are mandatory. When modeling is performed, it is necessary to enter the desired prediction length period, as in this study the prediction period is 30 days. The prediction results are contained in the "ds" and "yhat" columns with "yhat_upper" and "yhat_lower" as the upper and lower limits of the prediction results. The following Fig. 1 to Fig.3 shows the prophet workflow in this study with each treatment.

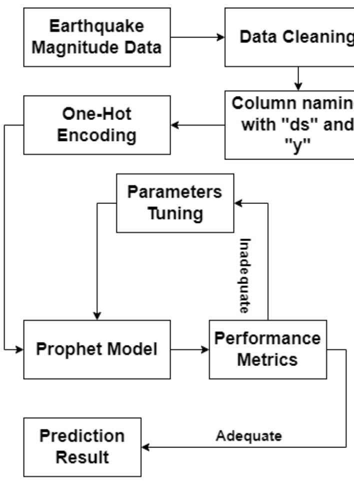

Fig. 1. Workflow of Prophet Without Data Transformation

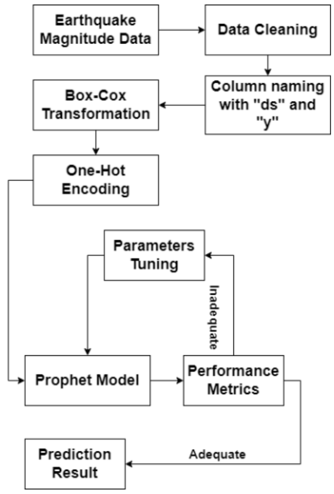

Fig. 2. Workflow of Prophet with Box-Cox Transformation

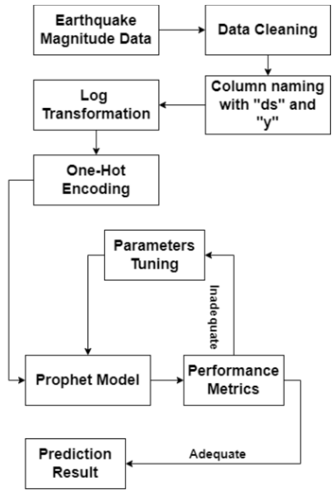

Fig. 3. Workflow of Prophet with Log Transformation

3. Literature Review

This section discusses research that has been done to predict magnitude and the use of Prophet in several case studies that show Prophet's performance which produces better prediction results than other algorithms.

Research [5] created a Neural Network (NN) model to predict the occurrence of earthquakes based on magnitude values. The dataset used is data taken from the Indonesian Meteorology, Climatology and Geophysics Agency (BMKG) from January 1, 2021 to January 11, 2021 based on its strength. The result obtained is that NN gets an RMSE value of 0.718.

Research [6] was conducted as an experiment to see whether the Artificial Neural Network (ANN) using Back Propagation can analyze the 2018 Bengkulu Province earthquake data sourced from BMKG along with the calculation of the error value it produces. The results obtained from this study are that ANN successfully predicts the strength of earthquakes on dates that are not inputted with the smallest error value of 2.93% and the largest of 59.1%.

Research [16] aims to forecast future earthquake trends using LSTM and FFNN. This study also compared the accuracy of LSTM earthquake forecasting results with FFNN in several regions in the Asian continent using magnitude and depth data. LSTM was shown to outperform FFNN. In this study, a trend-based approach was adopted and LSTM was used to capture trends involving statistical techniques. The trend-based method involves identifying patterns of seismicity that precede earthquakes. A comparison of the two methods resulted in the R2 score of the LSTM model being 59% greater than that of the FFNN.

In research [3] also compared several algorithms to forecast the magnitude of earthquakes that will occur in the next week using Support Vector Machine (SVM), Decision Tree (DT), and Shallow Neural Network (SNN) with contemporary Deep Neural Network (DNN). The case study used Iran earthquake data (longitude between 24.5 and 40 and latitude between 43.5 and 64) obtained from USGS and IIEES websites from January 1973 to July 2019. The results show satisfactory performance of DNN and SVM in predicting high magnitude classes. However, the performance of DT is more promising in dealing with events of both high and low magnitude.

Research [17] aimed to create a workflow to estimate the maximum magnitude of earthquakes that may occur in Italy using geological and geophysical data based on potential brittle volume and strain rate. The largest predicted values were 7.3 ± 0.25 for thrust faults, 7.6 ± 0.77 for normal faults and 7.6 ± 0.37 for horizontal faults.

Research [12] this study proposes a hybrid method using Prophet and LSTM models to overcome some limitations in an effort to predict accurate loads. This hybrid model was applied to the Bangladesh power system and the results showed higher accuracy in predicting electricity load compared to the non-hybrid model. This hybrid model has the ability to accurately predict load by reducing the complexity of AI and improving the accuracy of conventional methods by adding the advantages of both models while overcoming their respective limitations. The proposed hybrid technique can accurately forecast electricity consumption by incorporating self-history data without additional data and handcrafted feature selection operations. The proposed hybrid model within the forecasting time horizon of one day ahead forecasts 81.36 (RMSE), 0.91% (MAPE), and 80.11 (MAE) on average. Similarly, the proposed hybrid model forecasted an average of 89.04 (RMSE), 0.91% (MAPE), 71.23 (MAE), in one week ahead time horizon and one month ahead time horizon at an average of 249.60 (RMSE), 2.11% (MAPE), 189.81 (MAE) forecasts.

Five-year daily air temperature forecasts in Bandung in [13] were modeled with LSTM and Prophet. The results showed that Prophet worked better at maximum air temperature, while LSTM worked better at minimum air temperature. However, the difference in RMSE value is not too significant.

Research [7] proposed ARIMA, SARIMA and Prophet models to predict daily new cases and cumulative confirmed cases in the US, Brazil and India for the next 30 days based on a data set of new confirmed cases and cumulative confirmed cases of Covid-19. Through fitting and prediction of daily new case data, Prophet has more advantages in US Covid-19 prediction, which can structure data components and capture periodic characteristics when data changes significantly, while SARIMA is more likely to appear over-fitting in the US.

4. Methodology

The stages in the development of the Prophet model for earthquake forecasting include data collection, data preprocessing, modeling, and evaluation to obtain conclusions from the resulting model whether it produces a small error value.

A. Dataset

The research material used is earthquake catalog data in Indonesia from January 1, 2017 - July 7, 2023 obtained from the United States Geological Survey (USGS). The obtained dataset contains a time series with many variables (multivariate), so in the analysis, the data is converted into univariate because the main attention is focused on one particular variable (magnitude). The data taken is Indonesian earthquake data at longitude [91.758, 141.24], latitude [-10.644, 8.163 on the USGS website. In the retrieved data, there are time, latitude, longitude, depth, mag, magType, nst, gap, dmin, rms, net, id, updated, place, type, horizontalError, depthError, magError, magNst, status, locationSource, and magSource. For usage, it only uses date, mag, and place data (which has been grouped per province). Mag in this case is the magnitude of the earthquake. These three data are inputted into Prophet to be studied in order to produce magnitude prediction data per province 30 days ahead of the last data in the dataset. The total data from the five provinces is 6123 data. The filtered data was exported into a csv file for ease of data processing. An additional note for the provincial data in this dataset is that out of 38 provinces in Indonesia (according to the latest provincial data) only 5 provinces are the object of research, namely North Maluku, Maluku, Papua, Aceh and East Nusa Tenggara. These five provinces have more earthquake records than other provinces with a total number of data above 500. In other words, these provinces have the most frequent earthquakes of all provinces in Indonesia in the last 7 years. Grouping by province is done for ease of analysis.

B. Data Preprocessing

This process is done by checking missing values and outliers, cleaning unnecessary data, adjusting data types, retrieving data in the provinces of North Maluku, Maluku, Papua, Aceh, and East Nusa Tenggara, and visualizing magnitude data. The data was then subjected to three experiments, namely without transformation (after changing the names to "ds" and "y", one-hot encoding was performed), with box-cox transformation, and with log transformation. Transformation was performed, because the earthquake data has a skewness of: North Maluku = 1.504484, Maluku = 1.808268, Papua = 1.397431, Aceh = 1.405302, East Nusa Tenggara = 1.669407 which values are more than 0, so the data skewness towards the left with a longer tail on the right. Fig. 4. Depicting the skewness of the earthquake data.

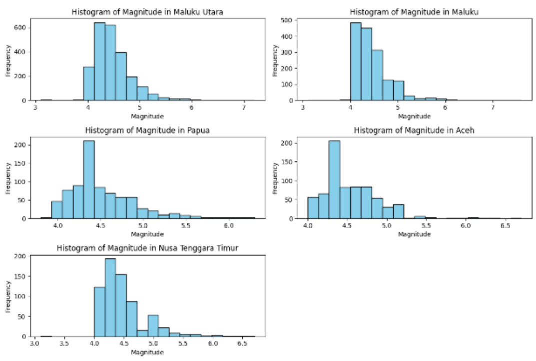

Fig. 4. Unsymmetrical Data

After transforming, the data was subjected to a one-hot encoding process to make it easier for Prophet to study the magnitude data per province. One-hot encoding is done by converting each category value (in this case province names) into a separate column in the new data and indicated by a binary value (0 or 1). Each column represents one category, and if the observation has that value, then the column value will be 1, while otherwise it will be 0.

C. Prophet Modeling

Prophet modeling is done by giving Prophet commands to study magnitude data in each province to make predictions for the next 30 days per province. The data predicted by Prophet is in the form of transformation data in experiments with box-cox transformation and log transform. After Prophet obtained the prediction results, the prediction results were converted back to the initial values to be analyzed and to calculate the error using RMSE, MAE, and MAPE.

D. Model Evaluation

Model evaluation is done by cross validation using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Absolute Error (MAPE). RMSE calculates the average of the square root of the difference between predicted and actual values. MAE measures the average of the absolute values of the difference between predicted and actual values, while MAPE is the percentage value of MAE. In this research, model evaluation focuses on the RMSE, MAE, MAPE values. The RMSE formula is given in equation 6.

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \hat{x}_i)^2}{n}} \quad (6)$$

Where n is the amount of data, x_i is the true data value, and $\hat{x}_i$ is the value predicted by the model. While the MAE formula is given in equation 7.

$$\text{MAE} = \sum_{i=1}^{n} \frac{|x_i - \hat{x}_i|}{n} \quad (7)$$

Where n is the number of data, x_i is the actual data value, and $\hat{x}_i$ is the predicted value of the model. The MAPE formula is given in equation 8. is the predicted value of the model. The MAPE formula is given in equation 8.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100 \quad (8)$$

Where n is the number of data, x_i is the actual data value, and $\hat{x}_i$ is the predicted value of the model, $\left| \frac{x_i - \hat{x}_i}{x_i} \right|$ is the absolute ratio of the difference between the actual and predicted values to the actual value, $\times 100$ converts the comparison value into percentage form. Cross validation is done with the initial value = 730 days, period = 30 days, and horizon = 30 days.

5. Result and Analysis

Based on the earthquake data retrieved, there is a total of 6123 clean data that has gone through various processes with each per province: North Maluku with 2365 data, Maluku with 1570 data, Papua with 792 data, Aceh with 716 data, and East Nusa Tenggara with 680 data. Fig. 5. shows a graph of the amount of data.

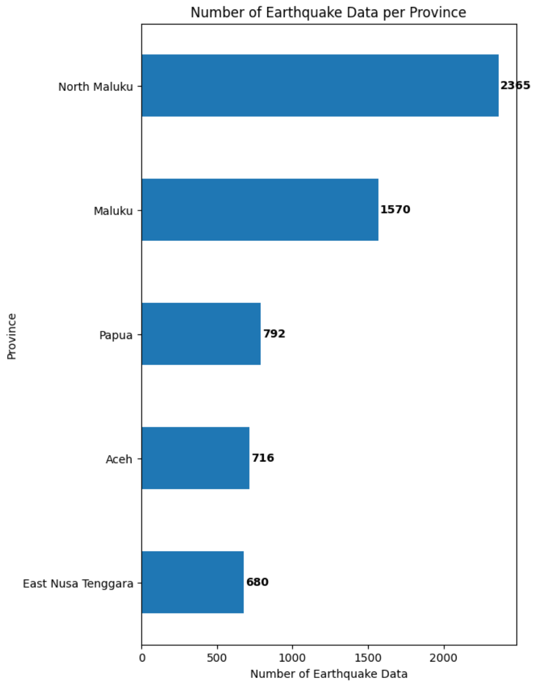

Fig. 5. Number of Earthquake Data per Province

The magnitude of each province also varies in pattern. The similarity of the magnitude of the five provinces used as the object of research is that the dominant value is between 4.0 and 5.0 with the highest and lowest values varying from province to province. The following Table 1 summarizes the highest and lowest magnitude values for each province. Fig. 6 to Fig. 10 display the magnitude per province.

Table 1. Highest and Lowest Magnitude per Province

Provinces	Highest Magnitude	Lowest Magnitude
North Maluku	7.2	3.1
Maluku	7.6	3.1
Papua	6.3	3.8
Aceh	6.7	4.0
East Nusa Tenggara	6.7	3.1

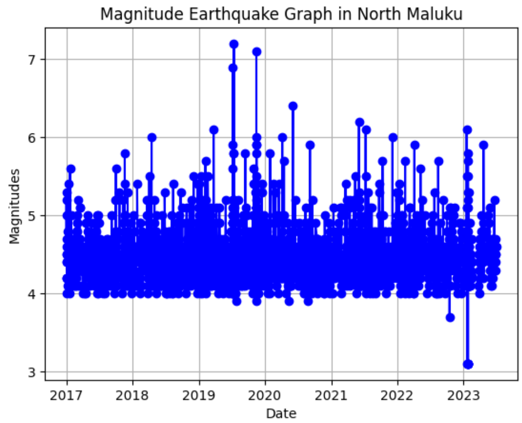

Fig. 6. Magnitude Earthquake Graph in North Maluku

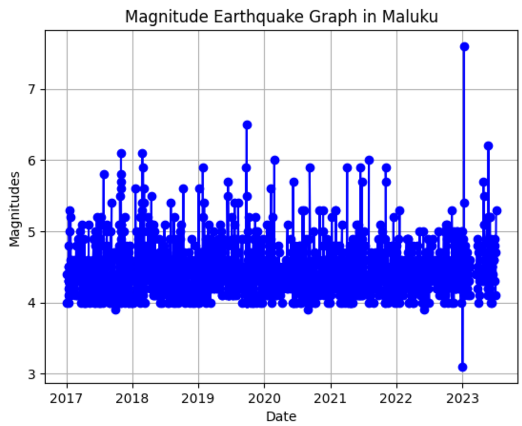

Fig. 7. Magnitude Earthquake Graph in Maluku

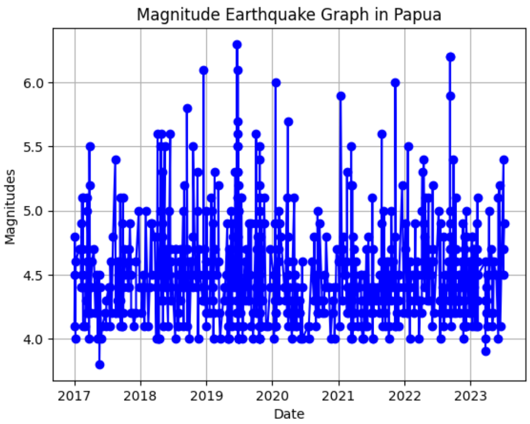

Fig. 8. Magnitude Earthquake Graph in Papua

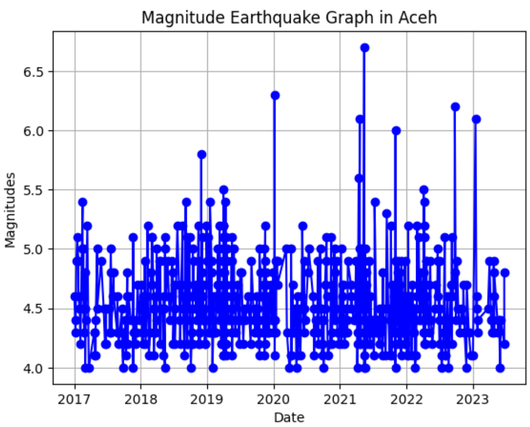

Fig. 9. Magnitude Earthquake Graph in Aceh

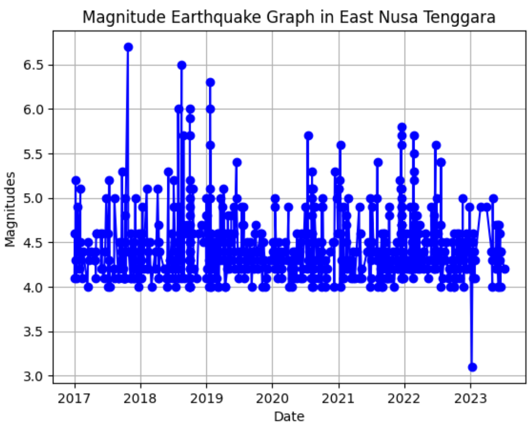

Fig. 10. Magnitude Earthquake Graph in East Nusa Tenggara

The magnitude graph above shows that a lot of data is above and below the dominant value, which can be considered as outliers. The treatment of outliers varies, one of which is to remove them. However, for this earthquake case study, the outliers cannot be removed, as earthquakes are threatening events when the magnitude is very large, such as in Maluku Province where the highest magnitude is 7.6 and there is only 1 data point. If this value is removed, Prophet will learn and predict data in the dominant range only. Therefore, outliers are treated by transforming the data into symmetric or normal distribution using box-cox transformation and log-transform. These two transformations have been frequently used to handle outliers. However, experiments without transformation are still conducted to compare the error values obtained and to find out which of these three experiments produces the lowest performance metrics values.

The following Table 2 gives the Prophet prediction results in the treatment without transformation for the provinces of North Maluku (NM), Maluku (M), Papua (P), Aceh (A), and East Nusa Tenggara (ENT). Table 3. prediction results on treatment with box-cox transformation and Table 4. prediction results on treatment with log transform transformation.

Table 2. Prediction Results Without Transformation

Date	NM	M	P	A	ENT
2023-07-08	4.5	4.6	4.5	4.5	4.4
2023-07-09	4.5	4.5	4.5	4.4	4.4
2023-07-10	4.5	4.6	4.5	4.5	4.5
2023-07-11	4.5	4.6	4.4	4.5	4.5
2023-07-12	4.5	4.6	4.4	4.5	4.4
2023-07-13	4.5	4.6	4.4	4.5	4.5
2023-07-14	4.5	4.6	4.5	4.5	4.5
2023-07-15	4.5	4.6	4.4	4.5	4.5
2023-07-16	4.5	4.5	4.4	4.4	4.4
2023-07-17	4.5	4.6	4.4	4.5	4.5
2023-07-18	4.4	4.6	4.3	4.5	4.6
2023-07-19	4.4	4.6	4.4	4.5	4.5
2023-07-20	4.4	4.6	4.3	4.5	4.5
2023-07-21	4.4	4.6	4.4	4.5	4.5
2023-07-22	4.4	4.6	4.4	4.5	4.5
2023-07-23	4.4	4.5	4.3	4.5	4.5
2023-07-24	4.4	4.6	4.4	4.6	4.5
2023-07-25	4.4	4.6	4.3	4.5	4.6
2023-07-26	4.4	4.6	4.4	4.5	4.5
2023-07-27	4.4	4.6	4.3	4.5	4.5
2023-07-28	4.4	4.6	4.4	4.5	4.5
2023-07-29	4.3	4.6	4.4	4.5	4.5
2023-07-30	4.4	4.5	4.4	4.5	4.5
2023-07-31	4.4	4.6	4.4	4.5	4.5
2023-08-01	4.3	4.6	4.3	4.5	4.6
2023-08-02	4.3	4.6	4.4	4.5	4.5
2023-08-03	4.3	4.6	4.4	4.5	4.5
2023-08-04	4.4	4.6	4.5	4.5	4.5
2023-08-05	4.3	4.6	4.4	4.4	4.5
2023-08-06	4.4	4.5	4.4	4.4	4.4

Table 3. Prediction Results with Box-Cox Transformation

Date	NM	M	P	A	ENT
2023-07-08	4.4	4.6	4.4	4.4	4.3
2023-07-09	4.5	4.5	4.4	4.4	4.3
2023-07-10	4.5	4.6	4.4	4.5	4.3
2023-07-11	4.4	4.6	4.3	4.4	4.3
2023-07-12	4.4	4.5	4.4	4.4	4.3
2023-07-13	4.4	4.5	4.3	4.4	4.4
2023-07-14	4.4	4.6	4.4	4.4	4.3
2023-07-15	4.4	4.6	4.3	4.4	4.3
2023-07-16	4.4	4.5	4.3	4.4	4.3
2023-07-17	4.4	4.6	4.3	4.5	4.3
2023-07-18	4.4	4.6	4.3	4.5	4.4
2023-07-19	4.4	4.5	4.3	4.4	4.3
2023-07-20	4.4	4.5	4.2	4.4	4.4
2023-07-21	4.4	4.6	4.3	4.5	4.3
2023-07-22	4.4	4.6	4.3	4.5	4.4
2023-07-23	4.4	4.5	4.3	4.4	4.3
2023-07-24	4.4	4.6	4.3	4.5	4.4
2023-07-25	4.3	4.6	4.2	4.5	4.4
2023-07-26	4.3	4.5	4.3	4.5	4.3
2023-07-27	4.3	4.5	4.2	4.4	4.4
2023-07-28	4.3	4.5	4.3	4.5	4.3
2023-07-29	4.3	4.6	4.3	4.5	4.4
2023-07-30	4.3	4.5	4.3	4.4	4.3
2023-07-31	4.3	4.5	4.3	4.5	4.4
2023-08-01	4.3	4.5	4.3	4.4	4.4
2023-08-02	4.3	4.5	4.3	4.4	4.3
2023-08-03	4.3	4.5	4.3	4.4	4.4
2023-08-04	4.3	4.5	4.4	4.4	4.3
2023-08-05	4.3	4.5	4.3	4.4	4.3
2023-08-06	4.4	4.5	4.3	4.4	4.3

Table 4. Prediction Results with Log-Transformation

Date	NM	M	P	A	ENT
2023-07-08	4.5	4.5	4.5	4.4	4.4
2023-07-09	4.5	4.5	4.5	4.4	4.4
2023-07-10	4.5	4.5	4.5	4.5	4.4
2023-07-11	4.5	4.5	4.4	4.5	4.5
2023-07-12	4.5	4.5	4.4	4.4	4.4
2023-07-13	4.5	4.5	4.4	4.4	4.4
2023-07-14	4.5	4.5	4.5	4.5	4.4
2023-07-15	4.5	4.5	4.4	4.5	4.5
2023-07-16	4.5	4.5	4.4	4.4	4.4
2023-07-17	4.5	4.5	4.4	4.5	4.5
2023-07-18	4.4	4.4	4.3	4.5	4.5
2023-07-19	4.4	4.4	4.4	4.4	4.5
2023-07-20	4.4	4.4	4.3	4.5	4.5
2023-07-21	4.4	4.4	4.4	4.5	4.5
2023-07-22	4.4	4.4	4.4	4.5	4.5
2023-07-23	4.4	4.4	4.3	4.4	4.4
2023-07-24	4.4	4.4	4.4	4.5	4.5
2023-07-25	4.4	4.4	4.3	4.5	4.6
2023-07-26	4.4	4.4	4.4	4.5	4.5
2023-07-27	4.4	4.4	4.3	4.5	4.5
2023-07-28	4.4	4.4	4.4	4.5	4.5
2023-07-29	4.3	4.3	4.4	4.5	4.5
2023-07-30	4.4	4.4	4.4	4.4	4.4
2023-07-31	4.4	4.4	4.4	4.5	4.5
2023-08-01	4.3	4.3	4.3	4.5	4.5
2023-08-02	4.3	4.3	4.4	4.4	4.5
2023-08-03	4.3	4.3	4.4	4.4	4.5
2023-08-04	4.4	4.4	4.5	4.5	4.5
2023-08-05	4.3	4.3	4.4	4.4	4.5
2023-08-06	4.4	4.4	4.4	4.4	4.4

If we pay attention, all the prediction results are in the number 4 and only differ in the number behind the comma. This is related to the dominant data in each dataset which is in the range of 4.0 - 5.0. The prediction results between the three experiments are also different, although some are the same in some provinces. For more details, let's compare the prediction results from the three experiments with the actual data per province. The following Table 5. contains a comparison between the prediction results Without Transformation (WT), with Box-Cox Transformation (BCT), and with Log-Transformation (LT) with Actual Data (AD) in North Maluku Province. Table 6. shows the comparison in Maluku Province. Table 7. contains a comparison of values in Papua Province. Similarly, Table 8. contains the results of the comparison in Aceh Province. Also, Table 9. which compares the results in East Nusa Tenggara Province.

Table 5. Comparison of Predicted Results and Actual Values in North Maluku Province

Date	WT	BCT	LT	AD
2023-07-09	4.5	4.5	4.5	4.3
2023-07-10	4.5	4.5	4.5	4.4
2023-07-11	4.5	4.4	4.5	4.5
2023-07-11	4.5	4.4	4.5	3.1
2023-07-11	4.5	4.4	4.5	4.2
2023-07-12	4.5	4.4	4.5	4.8
2023-07-13	4.5	4.4	4.5	4.2
2023-07-16	4.5	4.4	4.5	4.3
2023-07-16	4.5	4.4	4.5	4.9
2023-07-17	4.5	4.4	4.5	4.2
2023-07-17	4.5	4.4	4.5	4.9
2023-07-17	4.5	4.4	4.5	5.6
2023-07-19	4.4	4.4	4.4	4.1
2023-07-19	4.4	4.4	4.4	4.3
2023-07-20	4.4	4.4	4.4	4.7
2023-07-20	4.4	4.4	4.4	4.5
2023-07-20	4.4	4.4	4.4	4.6
2023-07-20	4.4	4.4	4.4	4.7
2023-07-20	4.4	4.4	4.4	5.2
2023-07-21	4.4	4.4	4.4	4.5
2023-07-21	4.4	4.4	4.4	4.7
2023-07-23	4.4	4.4	4.4	4.6
2023-07-24	4.4	4.4	4.4	4.4
2023-07-25	4.4	4.3	4.4	4.3
2023-07-25	4.4	4.3	4.4	5.1
2023-07-28	4.4	4.3	4.4	4.1
2023-07-29	4.3	4.3	4.3	4.1
2023-08-03	4.3	4.3	4.3	4.6
2023-08-03	4.3	4.3	4.3	4.4
2023-08-04	4.4	4.3	4.4	4.5

Table 6. Comparison of Predicted Results and Actual Values in Maluku Province

Date	WT	BCT	LT	AD
2023-07-08	4.6	4.6	4.5	4.5
2023-07-09	4.5	4.5	4.5	4.1
2023-07-11	4.6	4.6	4.5	4.3
2023-07-12	4.6	4.5	4.5	4.2
2023-07-13	4.6	4.5	4.5	4.2
2023-07-18	4.6	4.6	4.4	4.2
2023-07-19	4.6	4.5	4.4	4.4
2023-07-19	4.6	4.5	4.4	4.4
2023-07-19	4.6	4.5	4.4	4.3
2023-07-19	4.6	4.5	4.4	4.6
2023-07-20	4.6	4.5	4.4	4.6
2023-07-21	4.6	4.6	4.4	4.3
2023-07-21	4.6	4.6	4.4	4.4
2023-07-21	4.6	4.6	4.4	4.6
2023-07-25	4.6	4.6	4.4	4.4
2023-07-27	4.6	4.5	4.4	4.4
2023-07-28	4.6	4.5	4.4	4.4
2023-08-04	4.6	4.5	4.4	4.3

Table 7. Comparison of Predicted Results and Actual Values in Papua Province

Date	WT	BCT	LT	AD
2023-07-08	4.5	4.4	4.5	4.4
2023-07-13	4.4	4.3	4.4	4.8
2023-07-13	4.4	4.3	4.4	4.7
2023-07-17	4.4	4.3	4.4	4.7
2023-07-18	4.3	4.3	4.3	4.4
2023-07-19	4.4	4.3	4.4	5.2
2023-07-23	4.3	4.3	4.3	4.5

Table 8. Comparison of Predicted Results and Actual Values in Aceh Province

Date	WT	BCT	LT	AD
2023-07-10	4.5	4.5	4.5	4.7
2023-07-10	4.5	4.5	4.5	4.5

Table 9. Comparison of Predicted Results and Actual Values in East Nusa Tenggara Province

Date	WT	BCT	LT	AD
2023-07-12	4.4	4.3	4.4	4.6
2023-07-12	4.4	4.3	4.4	4.5
2023-07-12	4.4	4.3	4.4	4.1
2023-07-14	4.5	4.3	4.4	4.2
2023-07-21	4.5	4.3	4.5	4.9

If we look closely at the tables above, we can see that each experiment that successfully predicts according to the actual data marked with blue blocks is: no transformation = 6 correct, box-cox transformation = 5 correct, and log-transform = 10 correct from a total of 62 actual data values from all provinces. When there is the same actual data date with a different magnitude value, it still uses the predicted magnitude on that day with the same value. This is because Prophet only produces 1 day 1 prediction value. Similarly, if the actual data has no earthquake recordings on a particular day, Prophet still produces a predicted value, because Prophet produces predicted values from the specified time range. Unfortunately, in East Nusa Tenggara Province, there are not even any prediction results that match the actual data. Regarding Prophet that produces predictive values every day within the specified time range, this is the same as the research conducted by [6] where the Neural Network successfully fills in the empty times from the data entered. This means that the Neural Network also produces predictive values according to the specified time range. The performance metric results of the three experiments are given in Table 10, Table 11, and Table 12.

Table 10. Without Transformation Performance Metrics Result

Provinces	RMSE	MAE	MAPE
North Maluku	0.38	0.27	5.93%
Maluku	0.36	0.25	5.53%
Papua	0.42	0.30	6.55%
Aceh	0.36	0.27	5.91%
East Nusa Tenggara	0.37	0.28	6.24%
Average performance	0.38	0.28	6.03%

Table 11. Box-Cox Transformation Performance Metrics Result

Provinces	RMSE	MAE	MAPE
North Maluku	0.39	0.28	5.99%
Maluku	0.36	0.25	5.46%
Papua	0.42	0.30	6.45%
Aceh	0.38	0.28	5.91%
East Nusa Tenggara	0.37	0.28	6.15%
Average performance	0.38	0.28	5.99%

Table 12. Log-Transformation Performance Metrics Result

Provinces	RMSE	MAE	MAPE
North Maluku	0.38	0.27	5.89%
Maluku	0.36	0.25	5.48%
Papua	0.42	0.30	6.47%
Aceh	0.36	0.27	5.86%
East Nusa Tenggara	0.37	0.28	6.10%
Average performance	0.38	0.27	5.96%

In the performance metrics above, it can be seen that the results do not differ much between the three experiments. When referring to the prediction results that have been compared with the actual data, the order is the least box-cox results, without transformation in the middle, and log-transformation the most. However, if we look back at the average performance metrics, log-transformation is indeed the smallest among the others, comparable to the prediction results that match the actual data the most. However, the box-cox performance metrics are smaller than those without transformation, but it turns out that in the comparison of results, without transformation there are more 1 correct data. This is because in the prediction of other data that does not match the actual data, at least box-cox still seems to be more successful in producing predicted values that are close to the actual data. When looking deeper into the performance metrics, the numbers obtained are actually fairly good, because the resulting error is fairly small. Especially with a MAPE in the range of 5%. However, the magnitude data is also small, only in the range of 3.0 to 7.6, which should make the error value much smaller than this magnitude data range. It should be noted that RMSE and MAE are still in the range of 0.2 - 0.5, if you want to be more accurate, the value of these two metrics should be much smaller which is evident in the comparison of results, the difference between actual data and experimental results that do not match the actual data has the most difference in the range of 0.1 - 0.5, some even more than that. Therefore, Prophet still needs to be re-examined using other transformation methods or combining with other algorithms in the case of predicting earthquake magnitude.

6. Conclusion

After going through 3 experiments, namely without transformation, with box-cox transformation, and with log-transformation performed to reduce skewness, Prophet still cannot produce truly accurate magnitude predictions. Further research is needed to explore the potential of Prophet in this case study. However, this research has been fairly good in producing magnitude predictions with a small difference between the predicted results and the actual data. Unfortunately, even though the difference is small, only 10 data are correct out of 62 total actual data recorded. The RMSE metrics results for the three experiments have the same average of 0.38. The MAE value of the experiment without transformation with box-cox transformation is the same at 0.28 on average, while the log-transformation is 0.001 smaller at 0.27. The MAPE values see slightly more average differences than the RMSE and MAE where the no-transformation experiment has the largest average MAPE with 6.03%, box-cox with a value of 5.99%, and log-transformation the smallest with 5.96%.

References

[1] "Search Earthquake Catalog." https://earthquake.usgs.gov/earthquakes/search/ (accessed May 27, 2023).

[2] S. Gentili and R. Di Giovambattista, "Forecasting strong subsequent earthquakes in California clusters by machine learning," *Physics of the Earth and Planetary Interiors*, vol. 327, Jun.

2022, doi: 10.1016/j.pepi.2022.106879.

[3] M. Yousefzadeh, S. A. Hosseini, and M. Farnaghi, "Spatiotemporally explicit earthquake prediction using deep neural network," *Soil Dynamics and Earthquake Engineering*, vol. 144, May 2021, doi: 10.1016/j.soildyn.2021.106663.

[4] Institute of Electrical and Electronics Engineers and IEEE Geoscience and Remote Sensing Society, *2019 IEEE International Geoscience & Remote Sensing Symposium : proceedings : July 28-August 2, 2019, Yokohama, Japan*.

[5] O. Somantri, J. Teknik Informatika, and P. Negeri Cilacap, "Seminar Nasional Informatika Bela Negara (SANTIKA) Prediksi Kekuatan Gempa Bumi Indonesia Berdasarkan Nilai Magnitudo Menggunakan Neural Network".

[6] F. Firdausa, "Prediksi dan Analisis Data Gempa Bumi di Provinsi Bengkulu dengan Metode Artificial Neural Network," *CANTILEVER*, vol. 8, no. 2, pp. 45–49, Jan. 2020, doi: 10.35139/cantilever.v8i2.5.

[7] Y. Wang *et al.*, "Prediction and analysis of COVID-19 daily new cases and cumulative cases: times series forecasting and machine learning models," *BMC Infect Dis*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12879-022-07472-6.

[8] G. Cieleń, M. F. L. Derks, E. F. Knol, and E. Sell-Kubiak, "The impact of Box-Cox transformation on phenotypic and genomic characteristics of litter size variability in Landrace pigs," *Animal*, vol. 17, no. 5, May 2023, doi: 10.1016/j.animal.2023.100784.

[9] S. Marimuthu, T. Mani, T. D. Sudarsanam, S. George, and L. Jeyaseelan, "Preferring Box-Cox transformation, instead of log transformation to convert skewed distribution of outcomes to normal in medical research," *Clin Epidemiol Glob Health*, vol. 15, May 2022, doi: 10.1016/j.cegh.2022.101043.

[10] R. M. West, "Best practice in statistics: The use of log transformation," *Ann Clin Biochem*, vol. 59, no. 3, pp. 162–165, May 2022, doi: 10.1177/00045632211050531.

[11] "Prophet | Forecasting at scale." https://facebook.github.io/prophet/ (accessed May 27, 2023).

[12] T. Bashir, C. Haoyong, M. F. Tahir, and Z. Liqiang, "Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN," *Energy Reports*, vol. 8, pp. 1678–1686, Nov. 2022, doi: 10.1016/j.egyr.2021.12.067.

[13] T. Toharudin, R. S. Pontoh, R. E. Caraka, S. Zahroh, Y. Lee, and R. C. Chen, "Employing long short-term memory and Facebook prophet model in air temperature forecasting," *Commun Stat Simul Comput*, 2020, doi: 10.1080/03610918.2020.1854302.

[14] B. Zhang, C. Song, Y. Li, and X. Jiang, "Spatiotemporal prediction of O3 concentration based on the KNN-Prophet-LSTM model," *Heliyon*, vol. 8, no. 11, Nov. 2022, doi: 10.1016/j.heliyon.2022.e11670.

[15] C. B. Aditya Satrio, W. Darmawan, B. U. Nadia, and N. Hanafiah, "Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 524–532. doi: 10.1016/j.procs.2021.01.036.

[16] T. Bhandarkar, V. K, N. Satish, S. Sridhar, R. Sivakumar, and S. Ghosh, "Earthquake trend prediction using long short-term memory RNN," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 2, p. 1304, Apr. 2019, doi: 10.11591/ijece.v9i2.pp1304-1312.

[17] P. Petricca, E. Carminati, and C. Doglioni, "Estimation of the maximum earthquakes magnitude based on potential brittle volume and strain rate: The Italy test case," *Tectonophysics*, vol. 836, p. 229405, Aug. 2022, doi: 10.1016/j.tecto.2022.229405.