Using Time Series Analysis to Predict Earthquake

Darren King Wijaya
Computer Science Department
Bina Nusantara University
Jakarta, Indonesia 11480
darren.wijaya002@binus.ac.id

Rafaell Widjaya
Computer Science Department
Bina Nusantara University
Jakarta, Indonesia 11480
rafaell.widjaya@binus.ac.id

Afdhal Kurniawan
Computer Science Department
Bina Nusantara University
Jakarta, Indonesia 11480
afdhal.kurniawan@binus.ac.id

Diana
Computer Science Department
Bina Nusantara University
Jakarta, Indonesia 11480
diana@binus.edu

***Abstract*—One of the most disastrous natural disasters is an earthquake; it has the potential to cause mass-scale destruction and extensive loss of human lives. An earthquake is a sudden, violent movement of the earth's surface, produced by the movement of tectonic plates at faults in the earth's crust. The energy thus released travels in seismic waves. The earthquakes are totally unpredictable; therefore, safe buildings become more important in regions where the seismic events are more frequent. Prediction of earthquakes through time series analysis is one of the most challenging tasks and has, over time, shown promising improvements. Some of the most common techniques applied in predicting earthquakes are AutoRegressive Integrated Moving Average, Exponential Smoothing, Moving Average, Generalized Autoregressive Conditional Heteroskedasticity, and Seasonal AutoRegressive Integrated Moving Average with eXogenous factors, all having some unique strengths in making better predictions. These models work by analyzing past seismic data for trends and patterns that could indicate the likelihood of a future earthquake. This paper aimed at establishing which of these models performs better in predicting earthquakes by analyzing them based on accuracy and reliability. ARIMA has, over time, emerged as the best since it found better predictive power in earthquake prediction. Its admirable performance may potentially enhance earthquake prediction and, in effect, disaster preparedness. This will require continued research into the area so that predictive capabilities can be further refined and improved to really help lessen the impacts of earthquakes on vulnerable communities.**

Keywords*—Earthquake prediction, Time Series, ARIMA, Exponential Smoothing, Moving Average, GARCH, SARIMAX

I. INTRODUCTION

Earthquakes are considered to be one of the most disastrous natural phenomenons, causing really bad damage to infrastructure and huge loss of human life [1][2]. Determination of an earthquake's magnitude with accuracy therefore is very important in anticipating possible damage and putting effective disaster mitigation plans in place [3]. Traditional earthquake predicting techniques have used mainly seismometers and other physical measurements as the primary component, often tedious and error-prone. Time series analysis has already proved to be quite useful in increasing the accuracy and efficiency of earthquake prediction in recent years [4].

Time series analysis encompasses all statistical methods that involve time series data [3]. It is the analysis of data points past and present, organized at regular time intervals, for extracting essential statistics and other important features from data. This approach is going to be quite suitable for earthquake prediction since seismic data is inherently sequential and time-dependent [3]. Several time series models, including ARIMA, Exponential Smoothing, Moving Average, GARCH, and SARIMAX, have been applied to predict earthquakes with different levels of success [5][6][7][8].

This research paper will highlight the application of time-series analysis towards earthquake prediction [9]. This piece of research is done in an attempt to shed some light on how machine-learning models can capture the complex patterns beneath seismic signals with a literature review, coupled with an empirical analysis using real-world seismic datasets. It also aims to contribute to the discussion of major problems and aspects related to the use of machine learning in earthquake monitoring and early warning. This paper provides the opportunity for reviewing the various building blocks that compose the machine learning pipeline—essentially, data preprocessing, feature engineering, model selection, and validation—highlighting best practices and potential pitfalls. We will present some of our own research related to the use of time-series analysis in earthquake predicting, review the state of art in this area, and talk about the advantages and pitfalls of using time-series analysis. We are trying to provide insight into this fascinating and very promising field of research.

II. LITERATURE REVIEW

Predicting earthquakes has long been a critical range of research, with different methods being explored over the decades [3]. This review is pointed at summarizing the key headways and discoveries within the space of earthquake prediction, especially centering on time arrangement investigation. Earthquake prediction once only depends on seismological data that includes seismic waves and their engendering [3]. This old-school method still struggles with accuracy and convenience. And for seismometers, while successful in recording earthquake events, it often still struggles to provide accurate predictions due to its complexity, and the tremendous number of impacting factors.

Time series analysis has been used as a powerful tool for earthquake prediction. Different types of time series analysis have been applied to seismic data with various degrees of success. **ARIMA** models have been in broad application since it is capable of dealing with different patterns in time series data [2]. The model decomposes time series data into auto-regressive and moving average components, therefore suitable for modeling linear relationships in seismic data as documented by Adhikari & Agrawal, 2019 [10]. **Exponential Smoothing** involves weighing the previous observations with a progressively smaller weight as one goes down the history of observations in order to predict future values. The sub-techniques for this technique at high levels include simple exponential smoothing, involving exponentially decreasing weights, Holt's linear trend model, which considers the trends in the data, and Holt-Winter seasonal, handling both trends and seasonality. Exponential smoothing works really well for short-term predictions where the most recent data are very relevant [11].

Moving Average models smooth out the short-term increases and bring into relief the longer-term trends in seismic data. The thinking behind this method is useful in reducing the noise of data and making underlying trends more apparent [12]. **GARCH** are usually found in financial markets, modeling volatility clustering. In the prediction of earthquakes, they involve the estimation of the volatility of seismic activities over time, which should be useful in gaining insights into the periodicity and intensity of earthquakes [13][14]. **SARIMAX** with eXogenous variables extends the original ARIMA models to include seasonal effects and external variables, making them very suitable for time series analysis data that contains periodic patterns and influences from external factors, very common in seismic data [15][16][17]. Different models have comparably been tested in various studies in search of the best approach towards earthquake prediction.

For example, in the paper "Comparative analysis of ARIMA and LSTM for predicting fluctuating time series data", ARIMA worked fine for short-term predictions, while LSTM did better when time is considered for a longer forecasting horizon, due to its capability in learning complex temporal patterns [18][19].

In another instance, "Covid-19 cases prediction using SARIMAX Model by tuning hyperparameter through grid search cross-validation approach," they applied the SARIMAX model in COVID-19 case forecasting, expressing how flexible and accurate the model was in handling time series analysis data but with influences from exogenous variables. The flexibility of the SARIMAX in this regard could be applied in earthquake prediction by encompassing a good number of exogenous factors, conditions, geological and anthropogenic activities [20].

Also, in the research paper "Single exponential smoothing method to predict sales multiple products", it was found that methods of Exponential Smoothing were brilliantly applied to tasks of short-term earthquake prediction; in it, the recent seismic activity was itself an excellent predictor, heralding immediate future events in the case of earthquakes. There has been profound improvement in the case of earthquake prediction by using time series analysis [21].

III. METHODOLOGY

Interest in earthquake prediction has been practiced for several decades with various methods and techniques being explored to predict earthquakes, which will be explained in the following section by method.

A. *Dataset*

This dataset comes from the Earthquake Repository, which is administered by the non-departmental Indonesian government organization, BMKG. Early in 2023, BMKG modified the layout of their website, producing two distinct datasets. The Preliminary Earthquake Catalog is the source of the new dataset, which contains focal mechanism information (if any). It includes data on earthquake events from November 1, 2008, to April 9, 2024, however some of the most recent earthquake events may not be accurate. This dataset has 7 variables, each of which has a name that is descriptive explained in TABLE I.

TABLE I. DATASET DESCRIPTION

Features	Description
tgl	the date of the event.

ot	the timestamp of the event.
lat	a geographic coordinate that gives the north-south location of a point on the Earth's surface, given in degrees, ranging from 6°N to 11°S.
lon	the angular distance of any point on the earth's surface east or west of the prime meridian, ranging from 94°E to 142°E.
depth	the distance from the hypocentre of the earthquake, the point within the Earth where the rupture begins, to the Earth's surface directly above it in km.
mag	a numerical measure of the energy released from the source of an earthquake. In this dataset the magnitude is ranging from 1 up to 9.5.
remark	Flinn-Engdahl seismic and geographic regions associated with taking place or being analyzed.

Feature extraction is an important part in the earthquake prediction process. Feature extraction is the process of identifying the most relevant and informative attributes in a given dataset, used subsequently during the training process of a machine learning model. For this study, we extracted depth, latitude, longitude and magnitude as the following features from the dataset [8].

B. *Pre-Processing*

Pre-processing is also important in machine learning models that involves cleaning and transforming the data to prepare it for modeling. First steps are managing missing value was addressed by imputing missing values using the mean or media of the corresponding characteristic. Second, to prevent features with wide ranges from taking over the model, scaling was applied to bring the features to a common range. Third, normalization was performed to make sure that every feature is on the same scale. Last is eliminating outliers from impairing the model's performance.

C. *AutoRegressive Integrated Moving Average (ARIMA)*

Time series analysis has persisted to be the main technique used in the managerial aspects of earthquake warning systems and predictions. ARIMA, meaning AutoRegressive Integrated Moving Average, is among the prime techniques in this field that already showed immense promise in being able to predict earthquake magnitudes. Because it has the autoregressive and moving average components, it positively captures complex dynamics over time series data; hence, ARIMA models are particularly appropriate in this application. An ARIMA model is specified using three parameters: p, d, and q. The p is the number of lagged observations taken into account, d is the number of times raw observation differs, and q is the size of the moving average window [15][22].

D. *Exponential Smoothing*

Exponential Smoothing is a time series forecasting method that gives more weight to recent observations and thus should easily capture the important hidden patterns and trends in seismic data. This technique thus fits very well for any non-stationary and noisy time series data, which is very common in earthquake records. It is self-adjusting to changes over time by dynamically changing the weights given to previous observations. Exponential Smoothing is a computationally efficient and simple approach that can fit perfectly within the real-time early earthquake warning applications. Exponential Smoothing comes in several forms: Simple Exponential Smoothing, suitable for data without trend or seasonality; Holt's Linear Trend Model, an extension of SES designed to manage linear trends; and the Holt-Winters Seasonal Model, which addresses seasonality in addition to trends [11].

E. *Moving Average*

Moving Average is another technique in time series forecasting where the average of a given number of previous observations is computed. This will create a smoothed representation of the data and, at the same time, showcase the basic trend. This method is more helpful in extracting time dependencies and periodicities that are relevant in seismic data for earthquake prediction. Moving Average is also insensi-tive to outliers and noisy observations; it acts as a corrective measure against the short-term noise. It has very low computational complexity and is simple to implement—therefore, suiting perfectly for any real-time earthquake early warning system. [12].

F. *Generalized Autoregressive Conditional Heteroskedasticity (GARCH)*

The application of a GARCH model will then be used to forecast future earthquake magnitudes, taking into account time-varying volatility inherent in seismic data. One major reason underlying the suitability of the GARCH model for this particular application is that it models clustering in periods of high and low volatility, very common features of earthquake time series data. Such a model, therefore, would have a major advantage in that it will be robust to heteroskedasticity and efficiently capture long-range dependencies and event clustering, hence improving this appraisal's performance. The parameters of the GARCH model itself give information about different mechanisms underlying seismic processes. It also serves an important role in estimating volatility normally witnessed in financial returns, especially in very liquid markets. This algorithm exploits the aspect of volatility clustering, which means periods of high volatility are followed by high volatility, and low volatility periods also normally follow low volatility. GARCH (p,q) specifically stipulates the order of the terms of GARCH as 'p, that of the lagged variance, while 'q' that of the ARCH, that is, squared lagged residuals [13][14].

G. Seasonal AutoRegressive Integrated Moving Average with eXogenous variables (SARIMAX)

This generalizes the ARIMA model in a way that makes SARIMAX able to deal with periodic time series data having an external influence. The seasonal and cyclic components captured make it more accurate in long-term forecasts. Surely, those exogenous variables, tectonic plate movements and stress accumulation would be excellent enhancements of the predictive power in this model. The suitability of SARIMAX includes very strong resistance to non-stationarity and heteroskedasticity. It also intrinsically handles missing data and irregular sampling; this makes it a very versatile tool in monitoring earthquakes [16]. This SARIMAX method is developed as a package of improved and enhanced time series analysis in general over the ARIMA model to handle the seasonality and the exogenous variables. Then it becomes very paramount when dealing with time-dependent data which shows periodic behavior at some time intervals [16][17].

H. MAPE interpretation

The table (TABLE II) that can be used to understand the meaning of the Mean Absolute Percentage Error related to forecasting accuracy. MAPE is a common evaluation metric for a model of forecasting performance. The lower the MAPE value the more accurate is the forecasting model. A MAPE below 10% is very accurate, and one above 50% is indicative of the fact that the forecasting model is not able to accurately fit and cannot be used for the purpose. In other words, the table very clearly interprets values of MAPE helpful in judging the quality and reliability of forecasting models as per the performance metric that is most widely used for it.

TABLE II. MAPE TABLE INTERPRETATION

MAPE	Interpretation
<10	Highly accurate forecasting
10-20	Good forecasting
20-50	Reasonable forecasting
>50	Inaccurate forecasting

Source: Lewis (1982, p. 40)

IV. RESULT

From 2016 to 2022, earthquakes frequency was highest with somewhat considerable magnitudes. After conducting tests to evaluate the performance of the regression model, it was found that the first model, which is the ARIMA, achieved the lowest error percentage, which is 0.3401 and has a MAPE value below than 10. Making it the best model to use shown in TABLE III.

TABLE III. EVALUATION MODEL

TYPE	RMSE	MAPE
ARIMA	0.3401	7.2790
Exponential Smoothing	0.3441	7.3611
Moving Average	0.4116	9.7461
SARIMAX	0.35558	7.6129
GARCH	3.3254	90.0144

After finding the value of each technique that is used in the model. We tried to make a plot based on earthquake data, forecast future values and tried to visualize prediction alongside the actual data (shown in Fig. 1-5).

The first plot (Fig. 1), "Earthquake prediction with ARIMA," contains a high level of variability in the blue history data line, with spikes. One could easily trace in the orange line how it closely follows the ups and downs of the historical data for the predictions by the ARIMA model. This model is suitable for a linear time series and can make a time series stationary through differencing and produces more accurate and stable predictions by capturing the linear patterns and dynamically adjusting to changes in trends and patterns, which in turn give lower errors.

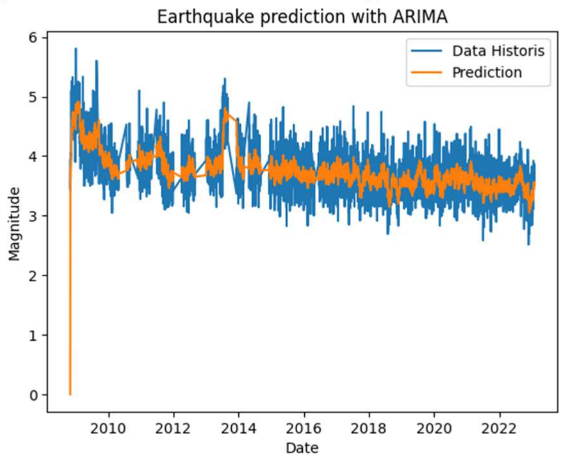

Fig. 1. Earthquake Prediction using ARIMA.

The second plot (Fig. 2), "Earthquake prediction with Exponential Smoothing," has the same time range and plots earthquake magnitudes. The blue line of historical data still preserves a high level of variability with frequent peaks. The orange line that comes out from the Exponential Smoothing algorithm is much smoother and stable. This predicted line does not catch the ups and downs of the historical data very well, meaning this approach is bad at capturing the short-term variability and leading to less accurate predictions.

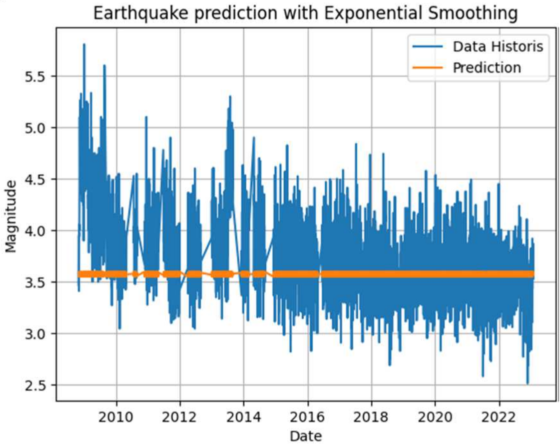

Fig. 2. Earthquake Prediction using Exponential Smoothing.

The third plot (Fig. 3) is entitled "Earthquake prediction with Moving Average." In this case, the historical data, represented by the blue line, preserves most of its high variability with frequent picks. The Moving Average predictions, represented by the orange line, in some sense balance the ARIMA and Exponential Smoothing methods. This smoothing of some short-term fluctuations while following the general trend of historical magnitudes resulted in medium accuracy in the prediction of earthquakes.

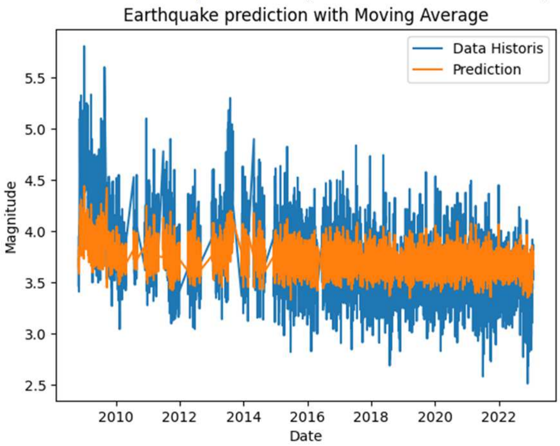

Fig. 3. Earthquake Prediction using Moving Average.

There was an issue with plotting Fig. 4 on the X-axis labels. It should be from 2009 – 2022. Though well, the appropriate labels were tried to be put, issues were faced with the representation of time range labels. The above problem shows the starching issues that may be involved when one wants to viz-represent precisely the time series data and how cautious one should be with the details while preparing a graphical outlay for analysis and interpretation.

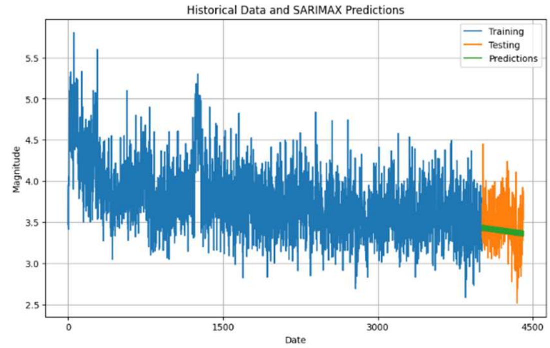

Fig. 4. Earthquake Prediction using SARIMAX.

Another issue faced was also with Fig. 5 when it was run. The best result could not be found, despite attempts having been made, due to problems pertaining to the representation of the prediction result. The above problem shows issues that may get involved while vizualization representing the time series data accurately and how much care one has to take at details while preparing the graphical outlay for analysis and interpretation. This models are directed toward issues of volatility and clustering, something irrelevant for seismic data that at times do not show any clear pattern of volatility.

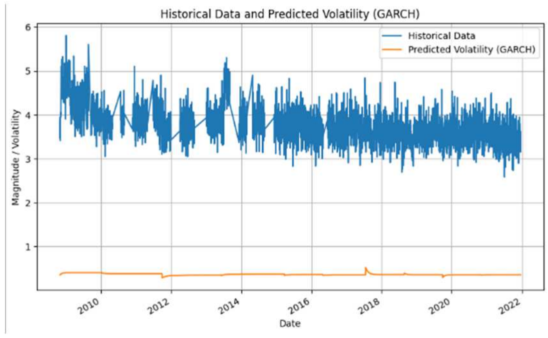

Fig. 5. Earthquake Prediction using GARCH.

V. CONCLUSION

This paper considers advanced time series forecasting methods for earthquake magnitude prediction ranging across the ARIMA, Exponential Smoothing, Moving Average, GARCH, and SARIMAX techniques. These models represent considerable advancements beyond simple implementation of the Gutenberg-Richter model because they account much better for such temporal and spatial complexity found with earthquakes. After seeing the evaluation model, ARIMA is a very flexible technique in dealing with different types of time series data, it was so widely applied in the forecast of earthquakes, and there is another very robust way of forecasting called Exponential Smoothing. Moving Average is a very simple technique, but it is always considered a foundation for all types of time series analysis methods. GARCH models are also quite useful for modelling, and forecasting variances that are seismically volatile. The SARIMAX model, being an extended version of the ARIMA framework, offers a comprehensive approach to the forecasting of earthquake magnitudes since it can incorporate seasonal patterns and other exogenous variables.

REFERENCES

[1] Wang, C., Li, C., Yong, S., Wang, X., & Yang, C. (2022). Time Series and Non-Time Series Models of Earthquake Prediction Based on AETA Data: 16-Week Real Case Study. *Applied Sciences*, *12*(17), 8536. https://doi.org/10.3390/app12178536

[2] Merdasse, M., Hamdache, M., Peláez, J. A., Henares, J., & Medkour, T. (2023). Earthquake Magnitude and Frequency Forecasting in Northeastern Algeria using Time Series Analysis. *Applied Sciences*, *13*(3), 1566. https://doi.org/10.3390/app13031566

[3] Celik, E., Kondiloğlu, A., & Atalay, M. (2016). THE EARTHQUAKE MAGNITUDE PREDICTION USED SEISMIC TIME SERIES AND MACHINE LEARNING METHODS. *IV. International Energy Technologies Conference (ENTECH'16)*.

[4] Et. al., S. M. (2021). TIME SERIES MODELLING ALGORITHM FOR PREDICTION OF EARTHQUAKES IN NEPAL. *INFORMATION TECHNOLOGY IN INDUSTRY*, *9*(2), 29–35. https://doi.org/10.17762/itii.v9i2.302

[5] KANBER, M., & SANTUR, Y. (2023). Time Series and Data Science Preprocessing Approaches for Earthquake Analysis. *European Journal of Science and Technology*. https://doi.org/10.31590/ejosat.1265261

[6] Nur Cahyo, E., & Susanti*, E. (2023). Analisis Time Series Untuk Deep Learning Dan Prediksi Data Spasial Seismik: Studi Literatur. *Jurnal Teknologi*, *15*(2), 124–136. https://doi.org/10.34151/jurtek.v15i2.3581

[7] Velicer, W. F., & Fava, J. L. (2003). Time Series Analysis. In *Handbook of Psychology* (pp. 581–606). Wiley. https://doi.org/10.1002/0471264385.wei0223

[8] Du, D., & Cui, Z. (2022). An Evolutionary Model for Earthquake Prediction Considering Time-Series Evolution and Feature Extraction and Its Application. *Journal of Physics: Conference Series*, *2333*(1), 012013. https://doi.org/10.1088/1742-6596/2333/1/012013

[9] Lok, S., & Karabatak, M. (2021). Earthquake Prediction by Using Time Series Analysis. *2021 9th International Symposium on Digital Forensics and Security (ISDFS)*, 1–6. https://doi.org/10.1109/ISDFS52919.2021.9486358

[10] Adhikari, R., & Agrawal, R. K. (2013). *An Introductory Study on Time Series Modeling and Forecasting*.

[11] Lukman, M., & Tanan, B. (2021). Time series modeling by using exponential smoothing technique for river flow discharge forecasting (case study: Cabenge, Walanae, and Cenranae rivers system). *IOP Conference Series: Materials Science and Engineering*, *1088*(1), 012100. https://doi.org/10.1088/1757-899X/1088/1/012100

[12] Dinesh, S., R, N. R., Anusha, S. P., & R, S. (2021). Prediction of Trends in Stock Market using Moving Averages and Machine Learning. *2021 6th International Conference for Convergence in Technology (I2CT)*, 1–5. https://doi.org/10.1109/I2CT51068.2021.9418097

[13] Smolović, J. C., Lipovina-Božović, M., & Vujošević, S. (2017). GARCH models in value at risk estimation: empirical evidence from the Montenegrin stock exchange. *Economic Research-Ekonomska Istraživanja*, *30*(1), 477–498. https://doi.org/10.1080/1331677X.2017.1305773

[14] Otto, P., & Schmid, W. (2019). *Spatial and Spatiotemporal GARCH Models -- A Unified Approach*.

[15] Laurenti, L., Tinti, E., Galasso, F., Franco, L., & Marone, C. (2022). Deep learning for laboratory earthquake prediction and autoregressive forecasting of fault zone stress. *Earth and Planetary Science Letters*, *598*, 117825. https://doi.org/10.1016/j.epsl.2022.117825

[16] Arunraj, N. S., Ahrens, D., & Fernandes, M. (2016). Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. *International Journal of Operations Research and Information Systems*, *7*(2), 1–21. https://doi.org/10.4018/IJORIS.2016040101

[17] Nurtas, M., Zhantaev, Z., & Altaibek, A. (2024). Earthquake time-series forecast in Kazakhstan territory: Forecasting accuracy with SARIMAX. *Procedia Computer Science*, *231*, 353–358. https://doi.org/10.1016/j.procs.2023.12.216

[18] Mondol, M. (2021). *Analysis and Prediction of Earthquakes using differentMachine Learning techniques*.

[19] Taslim, D. G., & Murwantara, I. M. (2024). Comparative analysis of ARIMA and LSTM for predicting fluctuating time series data. *Bulletin of Electrical Engineering and Informatics*, *13*(3), 1943–1951. https://doi.org/10.11591/eei.v13i3.6034

[20] Sah, S., Surendiran, B., Dhanalakshmi, R., & Yamin, M. (2023). Covid-19 cases prediction using SARIMAX Model by tuning hyperparameter through grid search cross-validation approach. *Expert Systems*, *40*(5). https://doi.org/10.1111/exsy.13086

[21] Gustriansyah, R., Suhandi, N., Antony, F., & Sanmorino, A. (2019). Single exponential smoothing method to predict sales multiple products. *Journal of Physics: Conference Series*, *1175*, 012036. https://doi.org/10.1088/1742-6596/1175/1/012036

[22] Fattah, J., Ezzine, L., Aman, Z., el Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, *10*, 184797901880867. https://doi.org/10.1177/1847979018808673