Received: 06 Jan 2025 | Revised: 10 Dec 2025 | Accepted: 19 Dec 2025 | Available online: 26 Feb 2026
https://doi.org/10.13057/ijas.v8i2.97739

E-ISSN: 2621-086X
https://jurnal.uns.ac.id/ijas

Indonesian Journal of Applied Statistics | Vol. 8, No. 2, pp. 102-112, 2025

Gold Price Forecasting with Long Short Term Memory (LSTM) and ARIMAX Method

Raisa Naura Adila*, Abdurakhman
Department of Mathematics, Universitas Gadjah Mada, Yogyakarta, Indonesia
*Corresponding author: raisanadl31@gmail.com

Abstract

Gold is very popular investment instrument due to its annual prices increases. In the long term, gold prices follow a nonlinear pattern, but in the short term, there are fluctuations influenced by various factors, including global market dynamics, monetary policy, and overall economic conditions. Therefore, predicting gold prices is an important step in minimizing risk and maximizing profits for investors. In this study, we analyze the performance of two methods for forecasting global gold prices, namely long short term memory (LSTM) and autoregressive integrated moving average with exogenous variables (ARIMAX). Data used is weekly global gold price data from August 1, 2000, to June 1, 2024. The variables used are close as the dependent variable and open as the exogenous variable. The data used is stationary data through the differencing process and algorithmic transformation to overcome non-stationarity issues. The best LSTM model uses the Tanh activation function with 30 LSTM units, 10 timesteps, and a dropout of 0.01, resulting in a MAPE value of 5.323%. The best ARIMAX model obtained was the ARIMAX (0,1,1) model, with a MAPE value of 0.55% for the test data and 0.61% for the training data. The research results, indicate that the higher accuracy of ARIMAX reflects its suitability for linear data such as gold prices, but the accuracy of LSTM which is below 10% still performs well for more complex data patterns.

Keywords: gold price; forecasting; LSTM; ARIMAX.

How to Cite:
R. N. Adila and Abdurakhman, "Gold price forecasting with long short term memory (LSTM) and ARIMAX method," *Indonesian Journal of Applied Statistics*, vol. 8, no. 2, pp. 102-112, 2025, doi: 10.13057/ijas.v8i2.97739.

1. INTRODUCTION

Gold is very popular investment instrument due to its annual prices increases. Gold is dense, shiny, and considered the most malleable metal among other metals, which makes its value and price very valuable. In the world of investment, the price of gold has a linear pattern in the long term, but in the short term, it tends to be non-linear due to fluctuations influenced by supply and demand, currency exchange rate changes, inflation, and monetary policies of central banks in various countries.

Old prices can be categorized as time series data. Time series data is a series of data that has a certain time span in the same interval [1]. The main basis in solving time series problems is to collect and select the right variables, then choose the best model that has the best accuracy [2]. Time series modeling can be grouped based on stationarity as well as the number of variables observed [3]. The two main methods in predicting gold prices are long short term memory (LSTM) and autoregressive integrated moving average with exogenous variables (ARIMAX). LSTM is a development of reccurent neural network (RNN) and effectively handles complex patterns and long-term dependencies. Meanwhile, ARIMAX is an extension of ARIMA that adds external variables such as inflation and exchange rates, providing more accurate predictions.

Previous research shows the good performance of LSTM and ARIMAX. LSTM excels in processing complex data, while ARIMAX effectively uses external variables. This study compares the two methods in predicting world gold prices (2000-2024) to determine the best model based on accuracy using MAPE. This study uses the stationary LSTM method to overcome data non-stationarity, ensuring statistical stability of the model inputs. On the other hand, ARIMAX requires stationary data as one of the basic assumptions, so the data used has also gone through a stationary process. Both methods are tested with stationary data to ensure a fair and objective comparison in handling equally stationary data. The results are expected to provide insight into the advantages of each method and help investors in dealing with global economic uncertainty.

2. METHODS

2.1. Basic Concepts of Time Series

A time series is a series of data that is bound to a specific time span and equal intervals. Time series modeling can be categorized based on stationarity and the number of variables observed. The grouping of time series models based on stationarity is as follows [3].

1. Stationary models are models with statistical properties that do not change over time. Examples of stationary models include i.i.d, white noise, MA, ARMA, and ARMAX.
2. Non-stationary models are models with statistical properties that can change over time. Examples of nonstationary models include ARIMA, SARIMA, ARIMAX, ARCH and GARCH.

The grouping of time series models based on the number of observed variables is as follows [3].

1. The univariate model is a model using one variable that is observed sequentially at a certain time interval. Examples of univariate models are daily stock prices, daily rainfall.
2. Multivariate models involve more than one variable observed simultaneously, where the variables affect each other. An example of a multivariate model is a stock market prediction with the variables used being the opening, closing and trading volume prices.

2.2. Machine Learning

Machine learning is a technique used to handle and predict big data by presenting the data through learning algorithms [4]. One of the important steps in developing machine learning models is proper dataset assignment. This process aims to avoid problems such as overfitting and ensure that the model can generalize well on data that has never been seen. Data splitting is the process of dividing a dataset into two or more parts. The purpose of this division is to avoid overfitting and evaluate the generalization ability of the model [5]. Dataset splitting is generally divided as follows.

1. Training set, used to train the prediction of model.
2. Validation set, used to optimize the prediction of model during training.
3. Testing set, used to test the prediction of model after completion of the training process.

The general dataset split ratio is 80%:10%:10% or 90%:5%:5%. For small datasets, the validation set may not be used, and only the training and testing sets are used. Data normalization is an important step in preparing a dataset for use in machine learning algorithms. One of the most commonly used normalization techniques is min-max transformation, which aims to adjust the scale of the data to be in a uniform range, such as between 0 and 1. This technique is particularly useful for scale-sensitive algorithms, like SVM and neural networks. The min-max transformation equation is as follows [6].

$$X_{min-max} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

2.3. Deep Learning

Deep learning is a part of artificial intelligence and machine learning that develops neural networks with multiple layers to improve accuracy in tasks such as object detection, speech recognition, and language translation. In deep learning, algorithms attempt to learn at various levels of difficulty, using neural networks consisting of high-level layers and low-level layers. Higher layers use information from

https://doi.org/10.13057/ijas.v8i2.97739
E-ISSN: 2621-086X

lower layers to recognize more complex patterns. For example, in image recognition, the initial layer recognizes lines and colors, the middle layer recognizes shapes, and the final layer recognizes objects or faces [7].

2.4. Activation Function

Activation function is a mathematical function used to calculate the output of a neuron based on the input received. This function describes the relationship between activation levels which can be a linear or nonlinear equation. In the topic of deep learning, the selection of activation functions is often done with a trial and error approach because there is no one activation function that can be considered the best for every problem. Some activation functions that are often used are as follows [8].

1. Sigmoid, a nonlinear activation function that produces an output in the range of 0 to 1, is used for binary classification problems. The sigmoid equation is presented as follows.
$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{2}$$
2. Hyperbolic tangent (Tanh), which is a nonlinear activation function with outputs in the range $(-1, 1)$. This function is used in the input and hidden layer. The Tanh equation is presented as follows.
$$Tanh(z) = \frac{1-e^{-2z}}{1+e^{-2z}} \tag{3}$$
3. Rectified linear unit (ReLU), is an activation function that converts input values greater than 0 to the original data value, and converts negative values to 0. The ReLU equation is presented as follows.
$$ReLU(z) = max(0, z) \tag{4}$$

2.5. Data Stationarity

Time series data is considered stationary if the data fluctuations occur around a constant mean value and has a constant variance. But in practice, many time series data are not stationary which occurs when the average or variance changes over time. To overcome average non-stationarity, a differencing method can be used which is done by subtracting the data value in a period with the previous period's data [9]. The number of differencing performed to achieve stationary is denoted as d.

2.6. Long Short Term Memory (LSTM)

The LSTM method is a modification of RNN designed to overcome the weakness of RNN in predicting data based on information stored for a long time [10]. LSTM is equipped with memory cells that function to store information from previous time periods, improving its ability to remember long-term information. LSTM uses three main gates: forget gate, input gate, and output gate, each of which has a specific function in collecting, classifying, and processing data. Figure 1 is the structure of LSTM:

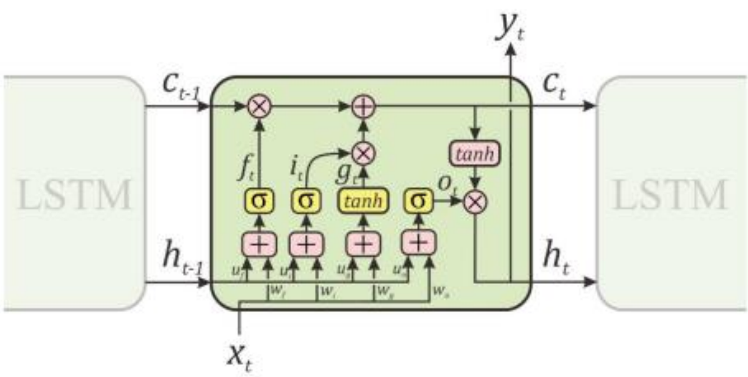

Figure 1. Structure of LSTM

1. The forget gate functions to decide what information will be discarded based on the previous hidden state with an output value between 0 and 1 and determine which information is forgotten or stored. The equation used in the forget gate is as follows.
$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_f\boldsymbol{x}_t + \boldsymbol{U}_f\boldsymbol{h}_{t-1} + \boldsymbol{b}_f) \tag{5}$$
2. The input gate functions to determine what information will be added to the cell state by using two parts, namely input gate and input cell, to update the cell state value with new information. The equation used in the input gate is as follows
$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_i\boldsymbol{x}_t + \boldsymbol{U}_i\boldsymbol{h}_{t-1} + \boldsymbol{b}_i) \tag{6}$$
$$\boldsymbol{g}_t = tanh(\boldsymbol{W}_g\boldsymbol{x}_t + \boldsymbol{U}_g\boldsymbol{h}_{t-1} + \boldsymbol{b}_g) \tag{7}$$
3. The output gate functions to produce LSTM output by processing the updated hidden state and cell state to produce prediction output. The equation used in the output gate is as follows.
$$\boldsymbol{O}_t = \sigma(\boldsymbol{W}_o\boldsymbol{x}_t + \boldsymbol{U}_o\boldsymbol{h}_{t-1} + \boldsymbol{b}_o) \tag{8}$$

2.7. ARIMAX

Time series modeling involves adding variables that are considered to have a significant influence on the data to increase forecasting accuracy. ARIMAX modeling is an extension of the ARIMA model that incorporates exogenous factors. In this framework, the dependent variable Y_t at time t is modeled not only by its own past values (autoregressive) and past forecast errors (moving average) but also by the current value of an independent exogenous variable X_t.
$$\phi_p(B)(1-B)^d Y_t = \beta X_t + \theta_q(B)\epsilon_t \tag{9}$$
In In this equation, the term βX_t explicitly demonstrates how the external variable is inserted into the model to explain variations in the series that cannot be captured solely by the historical patterns of Y_t.

In performing ARIMAX modeling, stationarity testing on exogenous variables is the first step that needs to be done [11]. After that, diagnostic checks are important to ensure the fit of the obtained model to the data. Diagnostic checks aim to verify whether the model fulfills basic assumptions, such as white noise properties and residual normality [12]. The white noise test is performed with the *Ljung-box* test, which tests whether there is residual autocorrelation. The hypothesis of this test states that the model meets the white noise assumption if there is no significant autocorrelation in the residuals. The *Jarque-Bera* test is used to verify whether the residuals are normally distributed, by comparing the value of the test statistic against the chi-square distribution. If the *p-value* of both tests is smaller than the specified significance level, then the null hypothesis is rejected, indicating that the model does not meet the assumptions tested [13].

2.8. Mean Absolute Percentage Error (MAPE)

A measure of forecasting error can be used to gauge how well a time series model performs. The measurement to test how much deviation between actual and forecasting data is called MAPE. The MAPE calculation is done by averaging the percentage of the result of the absolute error in each period divided by the real observation value in the same period. The highest limit is said to be good forecasting results if MAPE does not exceed 20%, because the smaller the MAPE, the better the results of forecasting [14]. The equation for calculating MAPE is as follows.
$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t} \times 100\% \tag{10}$$

3. RESULTS AND DISCUSSION

3.1. Data Description

This study uses weekly gold price data in USD per troy ounce obtained from Yahoo! Finance for the period August 1, 2000 to June 1, 2024. The unit of weight used, troy ounce, is equivalent to 31.1035 grams.

https://doi.org/10.13057/ijas.v8i2.97739
E-ISSN: 2621-086X

This dataset consists of six main variables, namely open, high, low, close, adjusted close, and volume. In this research, only 2 variables are used, namely open is used as the independent variable and close as the dependent variable to analyze the relationship between the opening price and closing price of gold weekly. The opening price reflects global market sentiment before trading begins, while the closing price reflects the results of market activity during the week. Changes in gold prices from open to close are influenced by various factors such as demand and supply, economic events, as well as changes in market sentiment [15].

Figure 2. Plot variabel close

Figure 3. Plot variabel open

Based on Figure 2 and Figure 3, there is an upward trend in gold prices over a 24-year period with fluctuations reflecting market volatility. The difference between the open and close prices indicates weekly price movements, where a close price higher than the open indicates an increase, while the opposite indicates a decrease. A significant period occurred between 2008 and 2011, with a sharp rise likely influenced by the global financial crisis. After peaking in 2011, the gold price declined until 2015, then increased again since 2016, reaching its highest level in 2024. Large fluctuations in some periods also emphasize the degree of volatility in gold prices over the time span.

3.2. Data Prepocessing

3.2.1. Data Transformation

Data transformation is using min-max transformation. Table 1 are the results of data transformation using min-max transformation.

Table 1. Results of data transformation

Before Transformation		After Transformation	
Open	Close	Open	Close
273.89	277	0.00824	0.00886
256.10	257.89	0	0
⋮	⋮	⋮	⋮
2358.30	2412.19	0.97337	1
2415.80	2335.5	1	0.963
2336.89	2322.89	0.96346	0.95854

3.2.2. Splitting of Training and Testing Data

This research splits the data into two main parts, that is training and testing to build and evaluate the prediction model. The data split ratio used is 70:30, with training data covering the period August 1, 2000 to April 17, 2017 and testing data covering the period April 24, 2017 to June 1, 2024 (presented in Figure 4).

https://doi.org/10.13057/ijas.v8i2.97739
E-ISSN: 2621-086X

Figure 4. Splitting of training and testing data

3.3. Identification of Data Stationarity

In time series analysis, checking stationarity of the data is an important step in building an accurate model. In ARIMAX models, stationarity is a key requirement where the data should fluctuate around an average with constant variance. In the LSTM model, data stationarity can also improve prediction accuracy. Data stationarity affects the model's ability to capture patterns and predict future values. One commonly used method to test for stationarity is the augmented dickey-fuller (ADF) test. This test detects the presence of a unit root, which indicates the data is not stationary. If the ADF test p-value is smaller than the significance level (α), then the data can be considered stationary. The ADF test results are used to ensure that the data meet the stationarity requirements before building the model. Below are the ADF test results for Open and Close data.

Table 2. ADF test results for close and open data

Transformation	*P-value* Close	*P-value* Open	Result
Originally	0.665	0.708	Not stationary
Transformation and Differencing	0.000	3.031e-27	Stasionary

Based on Table 2, the close and open data in the original data are not stationary, with p-values of 0.665 and 0.708 respectively. This shows that the null hypothesis (H_0), which states that the data is not stationary, cannot be rejected. However, after logarithmic transformation and first differencing, the data became stationary. This can be seen from the p-value for close data of 0.000 and for open data of 3.031×10^{-27}, which is much smaller than the significance level (α) of 0.05. Thus, the null hypothesis (H_0) is rejected, which means the close and open data become stationary after the transformation.

3.4. LSTM Model

The process of forming a prediction model using the LSTM method is carried out by exploring various combinations of hyperparameters to find the optimal configuration. The optimal combination for learning rate is 0.1, 0.01, and 0.0001 with batch size 32 [16]. Hyperparameters tested include timesteps (10, 20, 30, and 50), LSTM units (30, 50, and 100), dropouts (0.1, 0.2, and 0.5) and Adam's optimization algorithm [17].

In this research, parameter testing includes activation functions (ReLU, sigmoid, and tanh), LSTM units (10, 20, 30, and 50), timesteps (10, 15, 20, 25, and 30), dropout values (0.01, 0.05, and 0.1), epochs (60, 80, and 100), and batch size (32, 64, and 128). Based on Table 3, the best combination was found with the Tanh activation function, 30 LSTM units, dropout 0.01, timesteps 10, resulting the MAPE value is 5.323%, which represents excellent prediction accuracy.

In addition, this LSTM model uses three LSTM layers to capture long-term patterns in the data. The first and second layers use 50 LSTM units, while the third layer uses 30 units. The activation function

https://doi.org/10.13057/ijas.v8i2.97739
E-ISSN: 2621-086X

used in each layer is Tanh, and dropout 0.01 is applied to each layer to prevent overfitting. For optimization, the Adam optimizer algorithm is used with a learning rate setting of 0.001, which has proven effective in model convergence.

The loss curve and validation curve were generated during the training process to monitor the model's performance. The loss curve shows that the loss curve decreased steadily, indicating that the model successfully minimized training errors. The validation curve also shows consistent performance, with little fluctuation, indicating that the model did not overfit the training data.

Based on the performance analysis of each activation function, it can be concluded that the model with Tanh activation function produces the best performance compared to other models. This is because the Tanh model has the smallest MAPE value of 5.323%, dropout of 0.01, timesteps of 10, and the number of LSTM units of 30.

Table 3. LSTM best model results

Activation	Dropout	Timesteps	LSTM Unit	MAPE
Tanh	0.01	10	30	5.323%
Sigmoid	0.01	20	10	5.536%
ReLU	0.2	10	30	5.326%

The following is a visualization of the LSTM method with the best model using the Tanh activation function. Figure 5 aims to show the length to which the model is able to capture data patterns and minimize prediction errors.

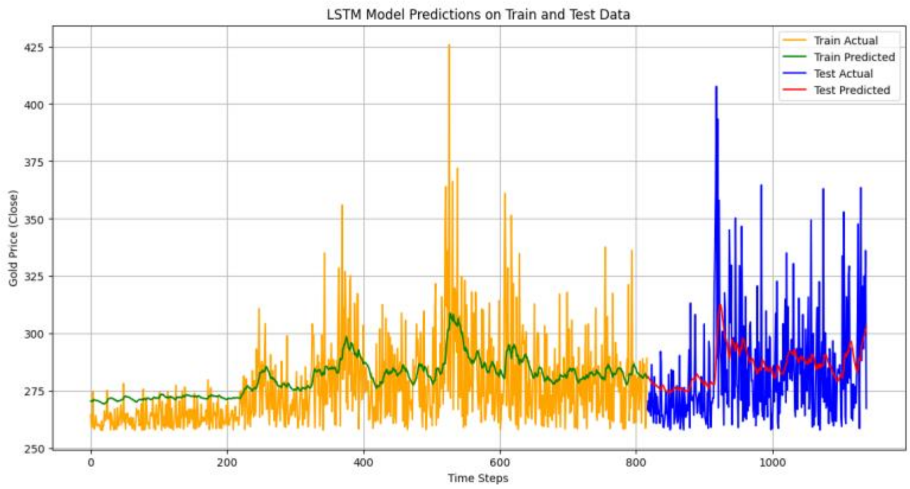

Figure 5. Plot of prediction results on training and testing data of LSTM model

3.5. ARIMAX Model

The application of the ARIMAX method involves various important stages designed to ensure the resulting model has high accuracy and validity. The stages begin with the identification of the ARIMA model to determine the optimal (p, d, q) parameter values using ACF and PACF plots, which help in understanding the basic structure of the data, such as AR, MA components, and the need for differencing (d) for the data to become stationary.

Unlike the standard ARIMA model which relies solely on past values of the variable itself, the ARIMAX model incorporates an exogenous variable (X_t) at time t to improve forecasting accuracy. In this research, the open price is introduced as X_t to explain the variation in the close price (Y_t)

https://doi.org/10.13057/ijas.v8i2.97739
E-ISSN: 2621-086X

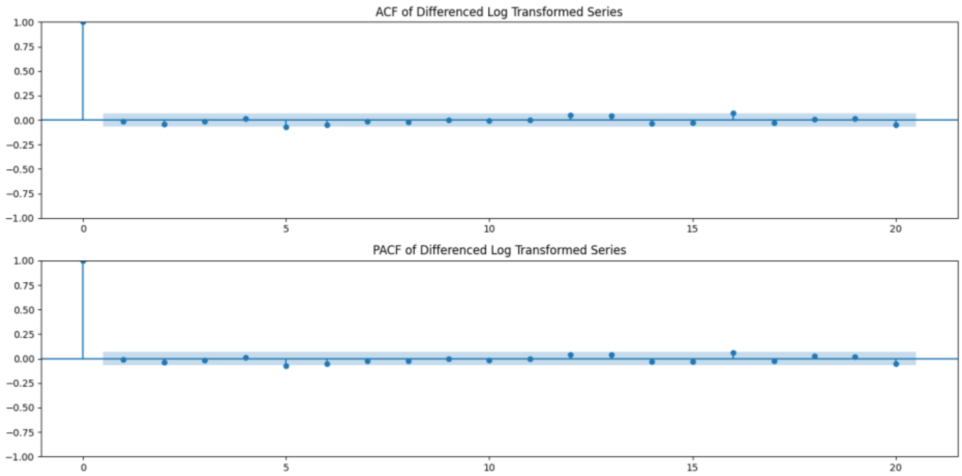

Figure 6. ACF and PACF plot

Based on Figure 6, the model formed from the order (p,d,q) is the ARIMA (2,1,2) model, then the combination of the ARIMA (2,1,2) model will be checked, namely the ARIMA (2,1,2), ARIMA (2,1,1), ARIMA (2,1, 0), ARIMA (1,1,2), ARIMA (1,1,1), ARIMA (1,1,0), ARIMA (0,1,2) and ARIMA (0,1,1) which then from these models will be selected by looking at the parameter significance test against the ARIMAX model.

Table 4. ARIMAX model parameter significance test results

ARIMAX Model	*P-value*			
	AR (1)	AR (2)	MA (1)	MA (2)
(2,1,2)	0.000*	0.117	0.684	0.000*
(2,1,1)	0.307	0.489	0.000*	-
(2,1,0)	0.000*	0.000*	-	-
(1,1,2)	0.330	-	0.954	0.336
(1,1,1)	0.617	-	0.000*	-
(1,1,0)	0.000*	-	-	-
(0,1,2)	-	-	0.000*	0.285
(0,1,1)	-	-	0.000*	-

Based on Table 4, the results of the parameter significance test for ARIMAX models with various combinations show the p-value for each AR and MA parameter. Based on the p-value of each ARIMAX model, there are model parameters with a p-value smaller than the 5% significance level, namely the ARIMAX (2,1,0), (1,1,0) and (0,1,1) models. Other ARIMAX models have parameters with p-values greater than the 5% significance level. The best model selection is done by evaluating ARIMAX parameters based on statistical significance and validation of AIC and BIC values. Based on Table 5, the ARIMAX (0,1,1) model is chosen as the best model because it has the lowest AIC and BIC values. Next, we will estimate the ARIMAX (0,1,1) model. In the gold price ARIMAX model, the parameter significance test is conducted using because of the large sample size. Based on Python output, a summary of the estimation results for the ARIMAX (0,1,1) model parameters is presented in Table 6.

Table 5. AIC and BIC value of significant ARIMAX model

ARIMAX Model	AIC	BIC
(2,1,0)	-5270.134	-5251.106
(1,1,0)	-5204.327	-5190.056
(0,1,1)	-5461.753	-5447.482

https://doi.org/10.13057/ijas.v8i2.97739
E-ISSN: 2621-086X

Table 6. Parameter estimation results of ARIMAX (0,1,1) model

Parameter	Coefficient	Standard Error	*z*-count	*p-value*
Open	0.9996	0.001	1537.749	0.000
ma.L1	-0.9999	0.220	-4.545	0.000
sigma2	9.984e-05	2.17e-05	4.595	0.000

Parameter testing is conducted to ensure all parameters in the model are statistically significant. With a significance level of 5%, parameters such as exogenous variables (Open), ma.L1 and sigma2 show a p-value below 0.05. This result confirms that the parameters used contribute significantly to the prediction.

The accuracy of the ARIMAX (0,1,1) model is measured using MAPE. Based on Table 7, the model performs very well with a MAPE value of 0.61% for the training data and 0.55% for the test data. Based on Figure 7, the visualization of the comparison of predictions and actual data also corroborates the accuracy of the model, where the forecasting pattern follows the trend of the actual data consistently.

Table 7. MAPE result ARIMAX (0,1,1) model

Data	MAPE Result
Training	0.61%
Testing	0.55%

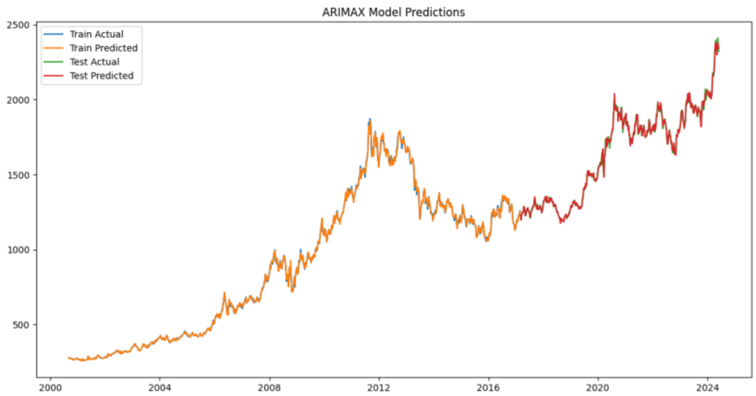

Figure 7. Plot of prediction results on training and testing data ARIMAX (0,1,1)

After the ARIMAX model is formed, residual assumption testing is carried out to make sure the validity of the model. The assumption tests performed are white noise test and residual normality test. The white noise test purpose to make sure that the residuals do not have significant autocorrelation, so that the model has captured the data pattern well. Based on Ljung-Box test, the test statistic is 13.801 with a p-value of 0.182. Since the p-value is greater than 0.05, the null hypothesis (H_0) is not rejected. This indicates that the residuals are random without significant autocorrelation. Thus, the model has met the white noise assumption, which indicates that the pattern in the data has been fully captured by the model, with no missing information in the residuals.

The next assumption test is the residual normality test to check whether the model residuals follow a normal distribution. Based on Jarque-Bera test, the test statistic is 7387.10 with a p-value of 0.00. Due to the p-value, the null hypothesis (H_0) is rejected, which indicates that the residuals are not normally distributed. However, in the context of time series prediction, the assumption of residual normality is not always considered a critical requirement. In the book The Analysis of Time Series, it is emphasized that the main purpose of time series models is to produce accurate predictions, not to ensure the distribution of residuals [18]. This opinion is supported that the assumption of residual normality is not always

https://doi.org/10.13057/ijas.v8i2.97739
E-ISSN: 2621-086X

important in prediction models, as long as the model can handle autocorrelation well [9]. The white noise assumption is more important than normality in models such as ARIMA or ARIMAX [19].

Based on the test results, although the residuals are not normally distributed, the ARIMAX model is still valid for use in prediction. The fulfillment of the white noise assumption ensures that the model has captured the data pattern effectively, and the resulting prediction results are reliable even though the residual normality assumption is not fully met.

3.6. Comparative Analysis of LSTM and ARIMAX Models

The LSTM and ARIMAX methods both perform well in predicting global gold prices, but with different advantages. Both use data that has been stationary through differencing and logarithmic transformation to overcome nonstationarity. The LSTM model is more effective in capturing linear patterns with a configuration involving Tanh activation function, 30 LSTM units, 10 timesteps and 0.01 dropout resulting in a MAPE of 5.323%. Although the accuracy is slightly lower than ARIMAX, LSTM is still very good at modeling linear patterns with gold price data.

The ARIMAX (0,1,1) model using the opening price as an exogenous variable yielded a MAPE of 0.55%, demonstrating its ability to handle linear patterns more simply and efficiently. ARIMAX is faster and easier to compute, making it suitable for data with stable linear relationships.

A comparison of the two shows that ARIMAX is superior in terms of accuracy and computational efficiency, while LSTM, although requiring longer computation time, can handle more complex patterns.

4. CONCLUSION

Based on the results of the analysis, it can be concluded that the best LSTM model uses data that has been stationary through differencing and logarithmic transformation with an optimal configuration that includes the Tanh activation function, 30 LSTM units, 10 timesteps and a dropout of 0.01 which produces a MAPE value of 5.323%, showing its ability to handle linear patterns and improve prediction accuracy for complex data. Meanwhile, the best ARIMAX model using stationary data by differencing and logarithmic transformation, as well as using the exogenous variable of opening price (Open), resulted in a MAPE value of 0.55%, showing high accuracy in modeling linear patterns and stable data. A comparison of the two methods shows that while ARIMAX has higher accuracy, LSTM with accuracy below 10% still performs well for modeling characteristics with more complex data patterns, making it an appropriate choice for data analysis with more dynamic characteristics.

REFERENCES

[1] M. F. A. Azis, F. Darari, and M. R. Septyandy, “Time series analysis on earthquakes using EDA and machine learning,” *in Proc. Int. Conf. Adv. Comput. Sci. Inf. Syst.*, Indonesia, 2020, doi: https://doi.org/10.1109/ICACSIS51025.2020.9263188.

[2] Z. Berradi and M. Lazzar, “Integration of principal component analysis and recurrent neural network to forecast the stock price of Casablanca Stock Exchange,” *Procedia Comput. Sci.*, vol. 148, pp. 55–61, 2019, doi: https://doi.org/10.1016/j.procs.2019.01.008.

[3] D. Rosadi, *Analisis Runtun Waktu dan Aplikasinya dengan R*. Yogyakarta: UGM Press, 2014.

[4] K. P. Danukusumo, *Implementasi Deep Learning Menggunakan Convolutional Neural Network untuk Klasifikasi Citra Candi Berbasis GPU*, Universitas Atma Jaya, Yogyakarta, 2017, https://e-journal.uajy.ac.id/12425/.

[5] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow*, 2nd ed., O'Reilly Media, 2019.

https://doi.org/10.13057/ijas.v8i2.97739
E-ISSN: 2621-086X

[6] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, 3rd ed., Morgan Kaufmann, USA, 2011, doi: https://doi.org/10.1016/C2009-0-61819-5.

[7] L. Deng, "Deep learning: methods and applications", *Foundations and Trends in Signal Processing*, vol. 7, no. 3, pp. 197–387, 2014, doi: https://doi.org/10.1561/2000000039.

[8] S. Kusumadewi, *Artificial Intelligence (Teknik dan Aplikasinya)*, Yogyakarta: Graha Ilmu, 2003.

[9] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting: Methods and Applications*, 3rd ed., Wiley, 1998, doi: https://doi.org/10.2307/2287014.

[10] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Networks*, vol. 5, no. 2, pp. 157-166, 1994, doi: https://doi.org/10.1109/72.279181.

[11] R. K. Paul, "ARIMAX-GARCH-WAVELET model for forecasting volatile data," *Model Assisted Statistics and Applications*, vol. 10, no. 3, pp. 243-252, 2015, doi: https://doi.org/10.3233/MAS-150328.

[12] D. Rosadi, *Analisis Ekonometrika dan Runtun Waktu Terapan dengan R*. Yogyakarta: ANDI, 2011.

[13] W. W. Wei, *Time Series Analysis*, Addison Wesley, New York, 2006.

[14] D. Lewis, *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*, Butterworth Scientific, London, 1982.

[15] E. J. Levin, A. Montagnoli, and R. E. Wright, *Short-run and Long-run Determinants of The Price of Gold*, Research Study 32. World Gold Council, 2006, https://strathprints.strath.ac.uk/7215/.

[16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016, doi: https://doi.org/10.4258/hir.2016.22.4.351.

[17] J. Brownlee, *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery, 2017.

[18] C. Chatfield, *The Analysis of Time Series: An Introduction*, 6th ed. Boca Raton, FL: CRC Press, 2003, doi: https://doi.org/10.4324/9780203491683.

[19] J. D. Hamilton, *Time Series Analysis*, Princeton University Press, 1994, doi: https://doi.org/10.1515/9780691218632.

https://doi.org/10.13057/ijas.v8i2.97739
E-ISSN: 2621-086X