Jurnal Matematika UNAND
Vol. **15** No. **3** Hal. 274 – 290
Edisi Juli 2026
ISSN : 2303–291X
e-ISSN : 2721–9410

https://doi.org/10.25077/jmua.15.3.274-290.2026

HYBRID AUTOREGRESSIVE INTEGRATED MOVING AVERAGE-LONG SHORT-TERM MEMORY (ARIMA-LSTM) MODEL FOR FORECASTING BRI BANK STOCK PRICES

DIAN KURNIASARI*, SILVI FITRIANI, WARSONO, MUSTOFA USMAN

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Lampung, Indonesia
email: dian.kurniasari@fmipa.unila.ac.id, silvi.fitriani105419@students.unila.ac.id, warsono.1963@fmipa.unila.ac.id, mustofa.usman@fmipa.unila.ac.id

Received October 18, 2024, Received in revised form January 18, 2026
Accepted July 9, 2026 Available online July 31, 2026

Abstract. Investors often consider BRI stock as one of the most attractive investment options, which makes it essential to develop reliable methods to forecast its price. The AutoRegressive Integrated Moving Average (ARIMA) model has been widely used to predict stock prices, but it has limitations due to its assumption of linearity. As an alternative, Long Short-Term Memory (LSTM) models are better suited for capturing both linear and non-linear patterns in time series data. This study explores the use of an ARIMA-LSTM hybrid model to forecast BRI stock prices, combining the strengths of both models. The hybrid approach involves splitting the data into two subsets for training and testing. In the first model, 80% of the data is used for training and 20% for testing, yielding an RMSE of 169.4864 and an MAPE of 0.0298%. The second model, using a split of 90% training and 10% tests, demonstrates improved performance with an RMSE of 169.0244 and an MAPE of 0.0297%. The results indicate that a higher proportion of the training data leads to better accuracy compared to the split of the training data 80%. Additionally, the model forecasts a continued upward trend in BRI's stock price for the next 60 periods, from October 31, 2022, to December 9, 2022, aligning with the latest data trends. These findings suggest that the ARIMA-LSTM hybrid model, especially when using 90% training data, is a highly effective approach for predicting BRI stock prices.

Keywords: Forecasting, Hybrid ARIMA-LSTM, Stocks.

1. Introduction

Stocks are financial instruments that represent ownership or a stake in a company and are a popular investment choice for both individuals and institutions [1] [2]. Based on data from the Indonesia Stock Exchange, Bank Rakyat Indonesia (BRI)

*Corresponding author

shares are in demand among domestic and foreign investors due to their large market capitalization and significant weighting in the IDX Composite and LQ45 indices, where these market-cap-weighted indices are dominated by large-cap stocks. BRI shares are highly liquid with high trading volume, so prices quickly reflect information within the framework of the efficient market hypothesis. The focus on SME financing demonstrates a direct link to the real economy in Indonesia, which contributes significantly to GDP. Share price movements are sensitive to interest rates, inflation, and economic growth, making them relevant as a proxy in predictive models with exogenous variables. As a blue-chip stock, the time series data is relatively stable and exhibits minimal anomalies. In April 2024, BRI's stock price is projected to reach 5,700 IDR per share, with a trading volume of 178,817,800 shares, reflecting significant investor interest.

Generally, stocks exhibit significant volatility, and their value is prone to fluctuation in both upward and downward directions. The volatile nature of the market makes it dynamic and difficult to predict, resulting in this predicament. Hence, accurately predicting stock prices is a significant hurdle due to the substantial volatility in a company's stock value, which can cause financial losses for numerous stakeholders in various sectors, including education, employment, technology and the general economy [3][4].

Predicting stock prices is crucial for maximizing profits and minimizing the risk of losses. Forecasting typically employs statistical approaches such as the ARIMA model, which is effective at capturing linear patterns in time series data [5]. However, long-term forecasting remains challenging due to the complexity and nonstationary of the data. ARIMA models are also limited in capturing nonlinear patterns, which can be demonstrated through nonlinearity tests or residual analysis; residuals that still contain autocorrelation or nonlinear structures indicate that the patterns have not been fully accommodated. Therefore, a nonlinear approach or hybrid model is needed to improve prediction accuracy.

Recently, artificial intelligence-based time series forecasting has shown a significant improvement in accuracy. This approach generally utilizes sequence models such as Long Short-Term Memory (LSTM), a variant of Recurrent Neural Networks (RNNs) that is effective for temporal data, including text and time series [6]. LSTM excels at modeling nonlinear relationships through gating mechanisms and is capable of capturing long-term dependencies via cell memory [7][8]. Conceptually, ARIMA, LSTM, and hybrid models have distinct differences: ARIMA is based on statistical assumptions to capture linear patterns such as autocorrelation and trends; LSTM is based on data learning to model complex nonlinear patterns; whereas hybrid models combine both to represent linear and nonlinear patterns simultaneously. However, in practice, using a single model such as LSTM alone may not necessarily be able to capture both patterns in a balanced manner, potentially increasing prediction errors and leading to underfitting or overfitting [9].

Dave et al. [10] state that ARIMA and LSTM have complementary strengths and limitations. The ARIMA model is effective at capturing linear patterns but is less capable of representing nonlinear relationships. Conversely, LSTM is capable of modeling both linear and nonlinear patterns, but requires a longer training time and lacks established standards for determining optimal hyperparameters. Therefore, this study integrates both models into a hybrid framework to optimize their respective strengths while mitigating their limitations. The contributions of this study include the application of the hybrid model to the BRI stock dataset, which exhibits complex characteristics with a combination of linear and nonlinear patterns. The methodological novelty is demonstrated through the integration strategy, where the linear component is modeled first using ARIMA, and then the residuals containing nonlinear patterns are processed using LSTM. This approach enables more structured and comprehensive modeling compared to previous methods, and is thus expected to produce more accurate and robust predictions than single models.

2. Related Work

ARIMA has been extensively utilized to predict and forecast time series data. In their study, Dai and Chen [11] utilized the ARIMA model to predict population data in Zhejiang Province from 2018 through 2022. Ewere and Eke [12] utilized the ARIMA model to forecast infant mortality rates in Nigeria until 2030. They used the United Nations's Inter-Agency Group data for Childhood Mortality Estimation (UN-IGME).

Subsequent studies conducted by other scholars have demonstrated that LSTM models exhibit superior aptitude in accurately predicting long-term trends in time series data. Elsaraiti and Merabet [13] performed a comparative investigation using LSTM and ARIMA models for wind speed prediction. The analysis results demonstrated that the LSTM approach exhibited superior accuracy to ARIMA. In their study, Sirisha et al. [14] conducted a profit analysis on financial data by comparing the performance of three models: ARIMA, Seasonal ARIMA (SARIMA), and LSTM. The results demonstrated that the LSTM model surpassed other models, with an accuracy rate of 97.01%.

Regrettably, the training procedure utilizing the LSTM model is time-consuming and necessitates a substantial number of parameters to attain optimal outcomes. Hybrid models, including ARIMA and LSTM, have demonstrated superior performance in diverse domains, particularly time series forecasting. An illustration of this is the study conducted by Jin et al. [15], which used the ARIMA-LSTM hybrid model to forecast COVID-19 data.

Farouk et al. [16] forecast stock prices using the ARIMA-LSTM hybrid model. The hybrid model's prediction results are derived from aggregating the outcomes of a superior single approach, surpassing both the ARIMA and LSTM single models.

This study employs a novel methodology combining predicted outcomes from linear models (ARIMA) with nonlinear models (LSTM). The methodology consists of three distinct stages: (a) utilizing the ARIMA model for forecasting, (b) modelling the residuals acquired from the ARIMA model using the LSTM model, and (c) combining the forecasts generated from both the ARIMA and LSTM models through a hybrid technique. Therefore, modelling linear and nonlinear patterns in data can be enhanced and more effective. This study aims to construct a hybrid ARIMA-LSTM model for predicting the closing price of Bank BRI's stock for the upcoming 60 periods. The acquired forecasting findings are anticipated to assist investors in making investment decisions regarding BRI Bank stock.

3. Methodology

3.1. *Data*

The study uses secondary data obtained from the website `https://finance.yahoo.com`, specifically a time series of daily BRI stock prices for the period from January 2017 to October 2022. The dataset includes six variables: opening price, high price, low price, closing price, adjusted closing price, and volume. However, the analysis focuses on the closing price variable, as it represents the final price that has accommodated all market information within a single trading period. Furthermore, the selection of this variable aligns with the univariate nature of the ARIMA model; thus, the use of a single primary variable is considered more appropriate, simpler, and capable of reducing the complexity of other variables that tend to be redundant.

3.2. *Time Series Forecasting*

A time series is a compilation of past data that illustrates the progression of a phenomenon throughout time. The period in question may vary from hours to years. Time series forecasting is predicting the future values of a time series. This prediction helps make informed decisions based on the forecasted results [17]. Time series forecasting models are categorized based on the number of variables they utilize. There are two distinct categories: univariate time series models, which solely utilize a single variable, and multivariate time series models, which employ multiple variables [18].

3.3. *Autoregressive Integrated Moving Average (ARIMA)*

The study utilized univariate time series data for analysis. The ARIMA model is a traditional statistical tool for analyzing univariate time series data. The ARIMA model is a composite of Autoregressive (AR) and Moving Average (MA) models and is most suitable for modelling univariate time series data [19]. Geurts et al. [20] study popularized this particular model. The ARIMA model can be mathematically expressed as Equation 3.1:

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)\varepsilon_t, \tag{3.1}$$

where $\phi_p(B)$ represents the parameter of the autoregressive (AR) process, whereas $\theta_q(B)$ represents the parameter of the moving average (MA) process.

The estimation of the ARIMA model parameters involves three key steps. First, the values of p, d, and q are determined during the model identification phase, typically using diagnostic tools such as the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Next, the parameters of the AR and MA components are estimated using statistical techniques such as Maximum Likelihood Estimation (MLE) or the Least Squares method. Finally, diagnostic checks are performed to ensure the model's adequacy, where the residuals e_t are evaluated to confirm that they resemble white noise. These steps ensure that the ARIMA model is appropriately fitted and capable of providing reliable forecasts for univariate time series data.

3.4. *Long Short-Term Memory (LSTM)*

The LSTM model is a recurrent neural network that incorporates multiple gates to create a network of memory blocks. That allows the model to effectively process data, retain information over extended periods, and address the vanishing gradients that often occur when analyzing long-term dependencies [21][22]. Generally, the gate being referred to comprises input gates, forget gates, and output gates [23].

(1) Forget Gate (f_t): The forget gate determines which information will be retained in the cell state. It receives the input from the previous hidden state (h_{t-1}) and the current input (x_t), and outputs a value between 0 and 1, dictating whether to keep or discard the information from C_{t-1}. The forget gate is represented mathematically as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{3.2}$$

where σ denotes the sigmoid activation function, and W_f and b_f represent the weights and biases for the forget gate, respectively.

(2) Input Gate (i_t): The input gate controls the information to be added to the cell state. It consists of two activation functions: the sigmoid function, which determines which values will be updated, and the tanh function, which generates the candidate new cell state vector C'_t. The mathematical expressions for the input gate and the candidate cell state are:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{3.3}$$

$$C'_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{3.4}$$

where tanh represents the hyperbolic tangent activation function, and W_C and b_C are the weights and biases for the cell state candidate.

(3) Cell State Update: The cell state C_t is updated by multiplying the previous cell state C_{t-1} by f_t to forget old information, and adding the result of the input gate, which is multiplied by the candidate cell state C'_t. This process is represented as:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C'_t. \tag{3.5}$$

(4) Output Gate (o_t): The output gate decides which part of the cell state will be output. The sigmoid activation function is used to select the relevant portion of the cell state, which is then passed through the tanh function and multiplied by the output gate. The equations for the output gate and the final output h_t are:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{3.6}$$

$$h_t = \tanh(C_t) \cdot o_t, \tag{3.7}$$

where W_o and b_o are the weights and biases for the output gate, respectively.

3.5. *Hyperparameter Tuning*

Hyperparameter tuning is crucial in constructing an optimal prediction model architecture by selecting the most suitable hyperparameter configurations. Optimizing hyperparameters can be a laborious process, mainly when there are a large number of parameters that require adjustment [24]. This study's variables that require modification are the number of neurons, epoch duration, and batch size.

3.6. *Hybrid ARIMA-LSTM*

The ARIMA-LSTM hybrid model combines the ARIMA and LSTM models, leveraging the strengths of each model based on their respective specializations. ARIMA is utilized to address the linear aspects of data, whilst LSTM is employed to address the nonlinear aspects of data. The integration of time series models with linear and nonlinear autocorrelation patterns can be mathematically represented. That assumes that the linear and nonlinear forecasted values can be combined to produce an overall output, representing the forecasting process's outcome [25].

$$y_t = L_t + N_t, \tag{3.8}$$

with:

$$\begin{aligned} y_t &: T - th \text{ period data}, \\ L_t &: \text{ linear component}, \\ N_t &: \text{ nonlinear component}. \end{aligned}$$

The hybrid model begins by applying the ARIMA method to identify and forecast the linear component L_t. ARIMA is well-suited for this task as it models linear relationships through its AR, differencing (I), and MA processes. Once the ARIMA model generates a forecast for the linear component, denoted as $\hat{L}_t$, the residuals e_t are calculated by subtracting $\hat{L}_t$ from the original data y_t. These residuals, defined as:

$$e_t = y_t - \hat{L}_t,$$

primarily represent the nonlinear patterns and noise that ARIMA is unable to capture.

The next step is to feed the residuals into the LSTM model to capture the remaining nonlinear patterns in the time-series data. This process requires careful preprocessing due to the differences in characteristics between the ARIMA and LSTM models. The initial step involves identifying significant lags in the residuals through ACF and PACF analysis, thereby determining the extent to which temporal dependence persists after ARIMA modeling.

The analysis results indicate the presence of significant lags up to the p lag (e.g., lag 1 through lag k), indicating that the residuals are not yet fully white noise and still contain patterns not captured by ARIMA. This finding forms the basis for using LSTM to model the remaining nonlinear components. Based on this, the input sequence length for the LSTM is set to p, which is the number of significant lags identified. Each input to the LSTM is organized as a sequence of residuals at times $t-1, t-2, \cdots, t-p$, with the target output being the residual at time t. This approach allows the LSTM to capture the remaining temporal dependencies more optimally, thereby improving the prediction accuracy of the hybrid model.

Given that LSTMs are sensitive to the scale of input data, the residuals are also normalized to ensure consistent ranges across features. Normalization, such as Min-Max scaling or standardization, ensures that the input data remain suitable for training the LSTM, preventing issues such as exploding or vanishing gradients. The prepared sequences are then divided into training and testing datasets to validate the performance of the LSTM model and ensure it generalizes well to unseen data.

The LSTM model processes the residuals and generates a forecast for the nonlinear component, denoted as $\hat{N}_t$. By focusing on the residuals, the LSTM effectively isolates the nonlinear dynamics of the time series that are not addressed by ARIMA.

Finally, the forecasts from the ARIMA and LSTM models are combined to produce the overall forecast for the time series. Mathematically, the final prediction y_t is expressed as:

$$y_t = \hat{L}_t + \hat{N}_t.$$

This integration ensures that the linear patterns are captured by ARIMA, while the nonlinear aspects are modeled by LSTM, resulting in a more accurate and comprehensive forecast compared to using either model independently. This hybrid approach leverages the strengths of both ARIMA and LSTM, with ARIMA addressing the linear dependencies and LSTM capturing the nonlinear complexities, thereby providing a robust solution for time series forecasting.

3.7. *Model Evaluation*

The precise choice of performance criteria is crucial for evaluating the model [26]. Soy Temür and Yıldız [27] propose that the evaluation of forecasting models can be accomplished by utilizing two accuracy criteria, specifically:

(a) Mean Absolute Percentage Error (MAPE)
Kim and Kim [28] identified one evaluation metric extensively utilized because of its scale independence and interpretability benefits. MAPE, also known as Mean Absolute Percentage Error, is a metric used to quantify the accuracy of a forecast. Calculated by finding the average percentage difference between the predicted and actual numbers. The formulation denotes as Equation 3.9:

$$\text{MAPE} = \left(\frac{\sum_{t=1}^{n} \frac{|Y_t - \hat{Y}_t|}{Y_t}}{n} \right) \times 100\%. \tag{3.9}$$

(b) Root Mean Square Error (RMSE)
RMSE, short for Root Mean Square Error, is calculated by taking the square root of the total squared differences between the actual and forecast values. Then, this result is divided by the number of data points. The mathematical

description is Equation 3.10:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}{n}}. \tag{3.10}$$

4. Results

4.1. *Data Visualization*

The graph shows the fluctuations in the BRI's stock price from January 2017 to October 2022. Figure 1 illustrates the volatility of closing prices over time, while also showing some missing data. Quantitatively, the proportion of missing data is relatively small (less than 5% of the total observation), so it does not disrupt the main pattern of the time series; however, it still needs to be addressed to prevent it from affecting the analysis results.

To address this, the linear interpolation imputation method was used because it maintains temporal continuity and follows the local pattern of data changes. This method is considered more appropriate than replacement with the median, especially for dynamic financial data, as it utilizes information from the values before and after the missing data, thereby producing estimates that are more logical and consistent with the data movement.

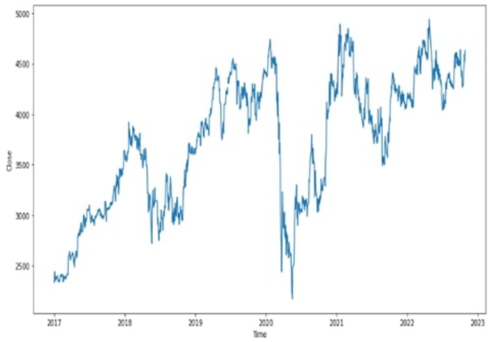

Figure. 1: Historical Closing Price of BRI Stock

4.2. *Stationary Test*

The p-value obtained from the Augmented Dickey–Fuller (ADF) test was 0.159, whereas the significance level employed in this study was 5%. Since the p-value exceeded the significance level α, the null hypothesis H_0 could not be rejected, indicating that the data were non-stationary and required first-order differencing. After the differencing process was applied, the resulting p-value became lower than the significance level α. Consequently, the null hypothesis H_0 was rejected, indicating that the differenced data were stationary. Therefore, the differencing parameter in the ARIMA model, denoted as d, was determined to be 1, indicating that only a single differencing process was required to achieve stationarity.

4.3. *Model Identification*

Subsequently, the model will be determined by examining the ACF and PACF plots in Figure 2.

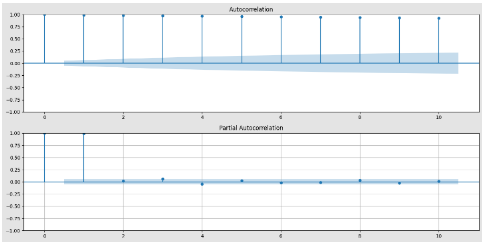

Figure. 2: ACF and PACF Plot

Based on the ACF and PACF plots, it can be seen that the second and third lags exceed the significance threshold, indicating the presence of autocorrelation at those lags. In the Box–Jenkins approach, significant spikes in the ACF are used as an indication of candidate Moving Average (MA) orders (q), while spikes in the PACF indicate candidate Autoregressive (AR) orders (p). Additionally, the cut-off and tailing patterns in both plots help distinguish the AR and MA components more clearly.

The findings of significance at the second and third lags are then used as the basis for determining the candidate model parameters. Therefore, the search space is focused on $p \in \{0, 2, 3\}$ and $q \in \{0, 2, 3\}$, with $d = 1$ based on the results of the differencing process to achieve stationarity.

4.4. *Estimation of Model Parameters and Diagnostic Testing of the Model*

Among the options obtained for the ARIMA(p,d,q) model, the ARIMA (0,1,2) satisfies the assumptions of significant parameters and residual white noise and has the lowest AIC value. The ARIMA(0,1,2) model will be utilized for making predictions.

4.5. *ARIMA Prediction*

Figure 3 demonstrates that the forecasted outcomes produced by the ARIMA(0,1,2) model closely align with the actual data pattern. The effectiveness of the ARIMA(0,1,2) model is evident from the obtained MAPE and RMSE values. The root mean square error (RMSE) value obtained is 73.7809, and the mean absolute percentage error (MAPE) value, which represents the absolute value of the residual percentage of data compared to the average, is 0.0141 percent. Therefore, the accuracy achieved by the ARIMA approach is 99.9859 percent. Therefore, it can be inferred that the ARIMA model with parameters (0,1,2) is highly effective in forecasting the closing price of the BRI stock.

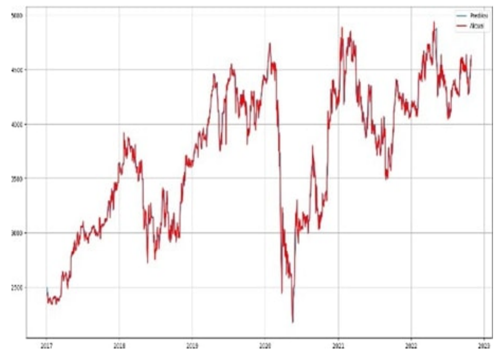

Figure. 3: Comparative Plot of Actual Data vs ARIMA Predictions

4.6. *ARIMA Forecasting*

Figure 4 indicates that the forecasting results generated by the ARIMA(0,1,2) model exhibit a strong correspondence with the observed data pattern. The reliability of the ARIMA(0,1,2) approach is reflected in the resulting RMSE and MAPE metrics. The model achieved an RMSE value of 170.3508, while the MAPE value, representing the average absolute percentage deviation between predicted and actual values, reached 0.0305%. These results correspond to a forecasting accuracy of 99.9695%. Accordingly, the ARIMA(0,1,2) model can be considered highly capable of predicting the closing prices of tBRI stock.

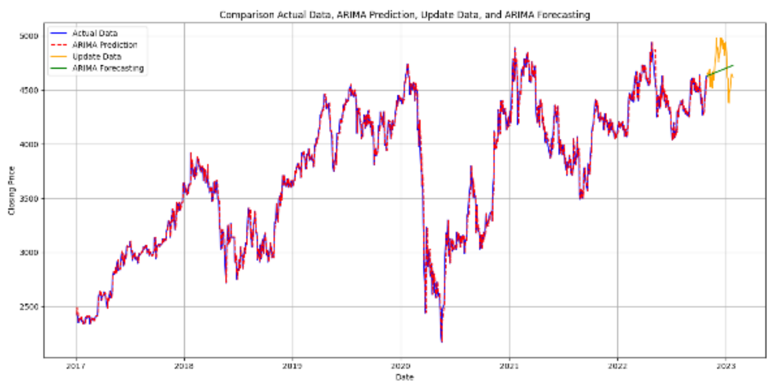

Figure. 4: Comparative Plot of Actual Data, Update Data, ARIMA Predictions, and ARIMA Forecasting

4.7. *Residual ARIMA*

Figure 5 depicts a graphical representation of the residual value obtained from the ARIMA model. The Residuals are derived by subtracting the predicted data from the actual data. The ARIMA residual value will be utilized as an input in the LSTM model and a nonlinear element in the hybrid model.

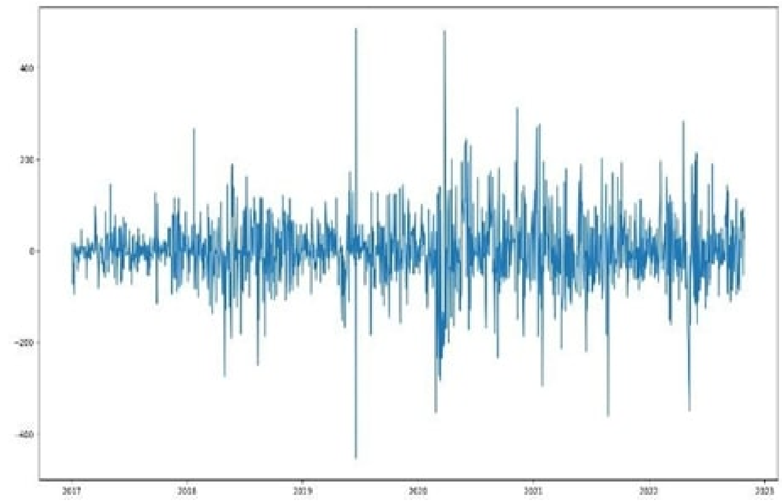

Figure. 5: Residual Plot

4.8. *Hybrid ARIMA-LSTM Process*

The initial stage in implementing the ARIMA-LSTM hybrid process involves partitioning the ARIMA prediction data and ARIMA residuals into separate sets for training and testing purposes. The data distribution for this study consists of 80% training data and 20% testing data, as well as 90% training data and 10% testing data. Following data sharing, the subsequent stage involves data transformation using a min-max scaler. That is necessary due to the comprehensive ARIMA prediction and residual data range.

The LSTM models were built with two hidden layers and one output layer. The LSTM model used ARIMA residual results as input. The use of residuals focuses on modeling error components that still contain pattern information. With this approach, linear patterns are modeled by ARIMA and nonlinear patterns are refined by LSTM more effectively.

The variables analyzed include the number of neurons in the hidden layer, the number of epochs, and the batch size. The hidden layer uses 64 and 128 neurons, with 100 epochs and batch sizes of 16, and 32. The parameter combinations were obtained through a hyperparameter tuning process. The results of the optimal combinations, the prediction evaluation for the LSTM model, and the prediction evaluation for the ARIMA-LSTM model are presented in Table 1, Table 2, and Table 3, while the LSTM and Hybrid ARIMA-LSTM model prediction evaluation are shown in Figure 6 through Figure 9.

Table 1: Hyperparameter tuning of the first LSTM model

Data Splitting	LSTM Unit	Batch Size	Dropout	Epoch
80:20	64	32	0.1	100
90:10	64	32	0.1	100

Table 2: LSTM Prediction Model Evaluation

Metrics	First Scheme (80:20)	Second Scheme (90:10)
RMSE	72.9708	61.5799
MAPE	1.2543	1.1746
Accuracy	98.7457%	98.8254%

Table 3: Hybrid ARIMA-LSTM Prediction Model Evaluation

Metrics	First Scheme (80:20)	Second Scheme (90:10)
RMSE	72.9708	61.5799
MAPE	0.0120	0.0108
Accuracy	99.9880%	99.9892%

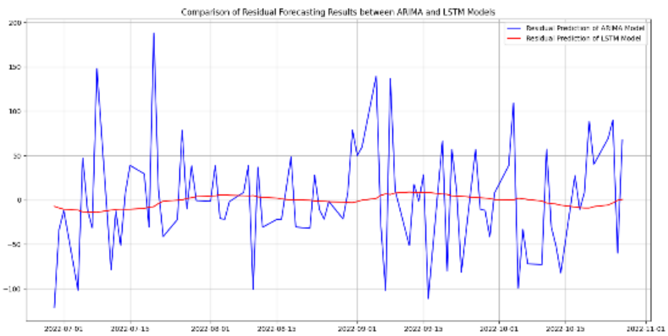

Figure. 6: LSTM Model Prediction Plot for First Scheme

4.9. *Model Hybrid ARIMA-LSTM Forecasting Results*

Upon assessing the model, it has been determined that the developed model exhibits a commendable level of accuracy. Consequently, the subsequent task entails predicting the outcomes for the upcoming 30 periods, specifically from October 31, 2022, to December 9, 2022. The forecasting process utilizes the most recent 60 historical data points of BRI stock's closing price to generate predictions for the upcoming period. The forecasting evaluation for the LSTM and hybrid ARIMA-LSTM model

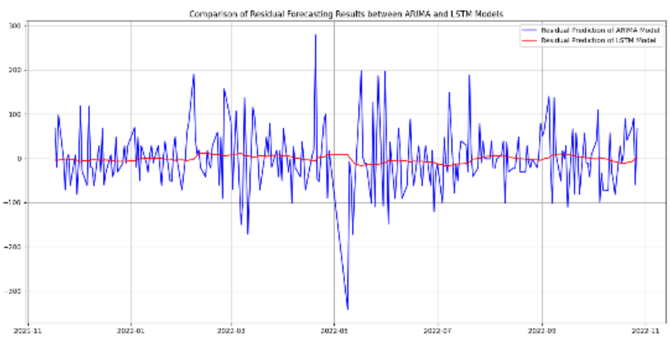

Figure. 7: LSTM Model Prediction Plot for Second Scheme

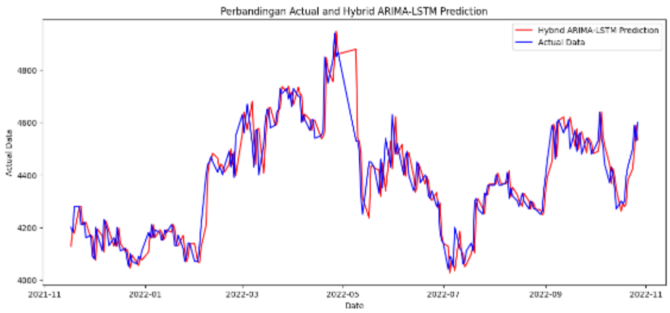

Figure. 8: Hybrid ARIMA-LSTM Model Prediction Plot for First Scheme

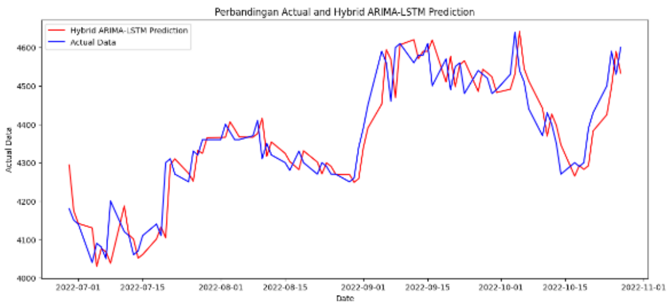

Figure. 9: Hybrid ARIMA-LSTM Model Prediction Plot for Second Scheme

are presented in Table 4 and Table 5, while the differences in forecasting results achieved by utilizing the LSTM model and the ARIMA-LSTM hybrid model are shown in Figure 10 to Figure 13.

Table 4: LSTM Model Forecasting Evaluation

Metrics	First Scheme (80:20)	Second Scheme (90:10)
RMSE	176.3081	173.2573
MAPE	1.1088	1.0450
Accuracy	98.8912%	98.9550%

Table 5: Hybrid ARIMA-LSTM Model Forecasting Evaluation

Metrics	First Scheme (80:20)	Second Scheme (90:10)
RMSE	169.4864	169.0204
MAPE	0.0298	0.0297
Accuracy	99.9703%	99.9702%

5. Conclusion

The ARIMA–LSTM hybrid approach is recommended for forecasting the closing price of Bank Rakyat Indonesia shares because it yields relatively low RMSE and MAPE values. With an 80% training and 20% testing data split, the model produced

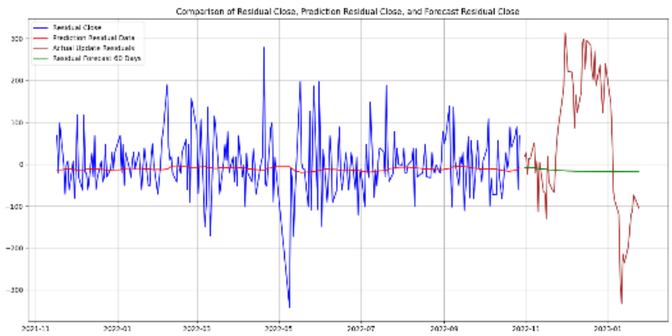

Figure. 10: LSTM Model Forecasting Plot for First Scheme

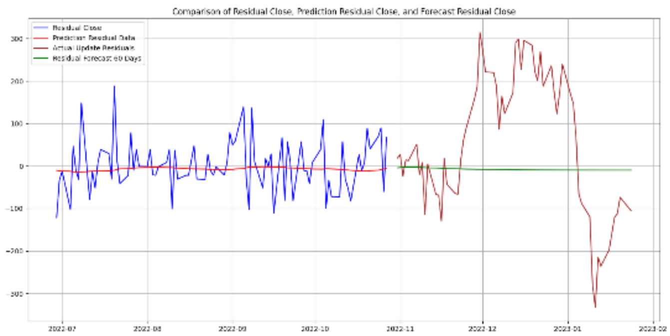

Figure. 11: LSTM Model Forecasting Plot for Second Scheme

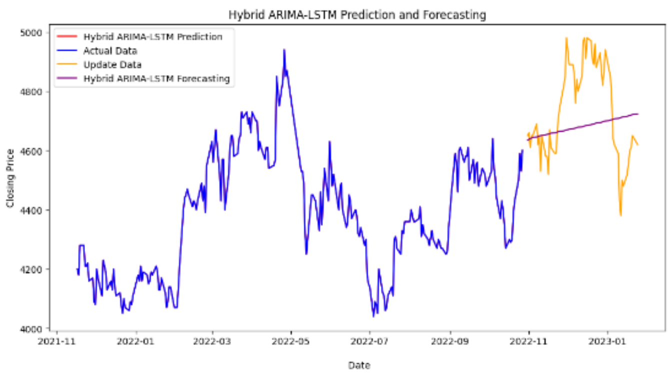

Figure. 12: Hybrid ARIMA-LSTM Model Forecasting Plot for First Scheme

an RMSE of 169.4865 and a MAPE of 0.0298%, while with a 90% training and 10% testing split, an RMSE of 169.0204 and a MAPE of 0.0297% were obtained. These results indicate that the 90:10 split yields a higher level of accuracy.

However, under certain conditions, the ARIMA model can produce a smaller RMSE compared to the hybrid model. This is due to the ARIMA model's predictive characteristics, which tend to be smoother and more stable in following linear trends, resulting in a relatively small deviation from the actual data. Conversely,

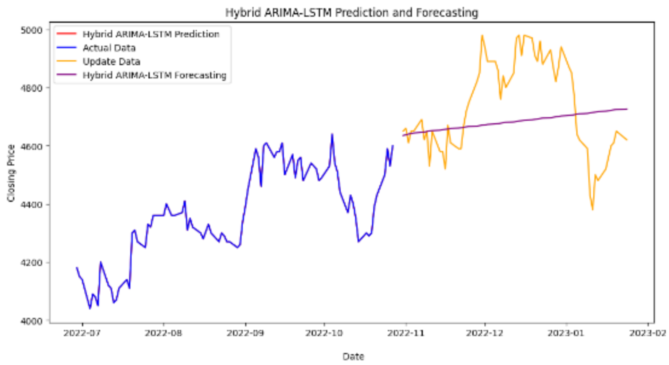

Figure. 13: Hybrid ARIMA-LSTM Model Forecasting Plot for Second Scheme

the hybrid model involving LSTM produces more fluctuating prediction patterns as it attempts to capture complex nonlinear dynamics. These fluctuations enhance the model's ability to represent data patterns more.

Bibliography

[1] Ruhani, F., Islam, M. A. I., & Tunku Ahmad, T. S., 2018, Theories Explaining Stock Price Behavior: A Review of the Literature, *International Journal of Islamic Banking and Finance Research*, Vol. **2**(2). `https://doi.org/10.46281/ijibfr.v2i2.215`

[2] Utomo, W. C., 2023, Prediksi Pergerakan Saham BBRI ditengah Issue Ancaman Resesi 2023 dengan Pendekatan Machine Learning, *J. Teknol. dan Manaj. Inform*, Vol. **9**(1). `https://doi.org/10.26905/jtmi.v9i1.9135`

[3] Setiani, I., Tentua, M. N., & Oyama, S., 2021, Prediction of Banking Stock Prices Using Naïve Bayes Method, *Journal of Physics: Conference Series*, Vol **1823** (1). `https://doi.org/10.1088/1742-6596/1823/1/012059`

[4] Singh, S., Gutta, S., and Hadaegh, A., 2021, Stock Prediction Using Machine Learning, *WSEAS Transactions on Computer Research*, Vol. **9**. `https://doi.org/10.37394/232018.2021.9.17`

[5] Khan, S., & Alghulaiakh, H., 2020, ARIMA model for accurate time series stocks forecasting, *International Journal of Advanced Computer Science and Applications*, Vol **11**(7). `https://doi.org/10.14569/IJACSA.2020.0110765`

[6] Arifin, S., Wijaya, A., Nariswari, R., Yudistira, I. G. A. A., Suwarno, S., Faisal, F., & Wihardini, D., 2023, Long Short-Term Memory (LSTM): Trends and Future Research Potential, *International Journal of Emerging Technology and Advanced Engineering*, Vol **13**(5): 24–34. `https://doi.org/10.46338/ijetae0523\_04`

[7] Wen, T., Liu, Y., Bai, Y. he, & Liu, H., 2023, Modeling and forecasting CO2 emissions in China and its regions using a novel ARIMA-LSTM model, *Heliyon*, Vol **9**(11). `https://doi.org/10.1016/j.heliyon.2023.e21241`

[8] Wiranda, L., & Sadikin, M., 2019, Penerapan Long Short Term Memory pada Data Time Series untuk Memprediksi Penjualan Produk PT. Metiska Farma, *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, Vol **8**(3).

[9] Domingos, D. S., de Oliveira, J. F. L., & de Mattos Neto, P. S. G., 2019, An intelligent hybridization of ARIMA with machine learning models for time

series forecasting, *Knowledge-Based Systems*, Vol **175**. https://doi.org/10.1016/j.knosys.2019.03.011

[10] Dave, E., Leonardo, A., Jeanice, M., & Hanafiah, N., 2021, Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM, *Procedia Computer Science*, Vol **179**: 480–487. https://doi.org/10.1016/j.procs.2021.01.031

[11] Dai, J., & Chen, S., 2019, The application of ARIMA model in forecasting population data, *Journal of Physics: Conference Series*, Vol **1324**(1). https://doi.org/10.1088/1742-6596/1324/1/012100

[12] Ewere, F., & Eke, D. O., 2020, Time Series Analysis and Forecast of Infant Mortality Rate in Nigeria: An ARIMA Modeling Approach, *Canadian Journal of Pure and Applied Sciences*, Vol **14**(2).

[13] Elsaraiti, M., & Merabet, A., 2021, A comparative analysis of the arima and lstm predictive models and their effectiveness for predicting wind speed, *Energies*, Vol **14**(20). https://doi.org/10.3390/en14206782

[14] Sirisha, U. M., Belavagi, M. C., & Attigeri, G., 2022, Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison, *IEEE Access*, Vol **10**. https://doi.org/10.1109/ACCESS.2022.3224938

[15] Jin, Y. C., Cao, Q., Wang, K. N., Zhou, Y., Cao, Y. P., & Wang, X. Y., 2023, Prediction of COVID-19 Data Using Improved ARIMA-LSTM Hybrid Forecast Models, *IEEE Access*, Vol **11**. https://doi.org/10.1109/ACCESS.2023.3291999

[16] Farouk, U., Abdulrahman, I., Ussiph, N., & Hayfron-Acquah3, B., 2020, A Hybrid Arima-Lstm Model for Stock Price Prediction, *International Journal of Computer Engineering and Information Technology*, Vol **12** (8).

[17] Casolaro, A., Capone, V., Iannuzzo, G., & Camastra, F., 2023, Deep Learning for Time Series Forecasting: Advances and Open Problems, In *Information (Switzerland)*, Vol **14**(11). https://doi.org/10.3390/info14110598

[18] Hao, J., & Liu, F., 2024, Improving long-term multivariate time series forecasting with a seasonal-trend decomposition-based 2-dimensional temporal convolution dense network, *Scientific Reports*, Vol **14**(1). https://doi.org/10.1038/s41598-024-52240-y

[19] Sahai, A. K., Rath, N., Sood, V., & Singh, M. P., 2020, ARIMA modelling & forecasting of COVID-19 in top five affected countries, *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, Vol **14**(5). https://doi.org/10.1016/j.dsx.2020.07.042

[20] Geurts, M., Box, G. E. P., & Jenkins, G. M., 1977, Time Series Analysis: Forecasting and Control, *Journal of Marketing Research*, Vol **14**(2): 269. https://doi.org/10.2307/3150485

[21] Van Houdt, G., Mosquera, C., & Nápoles, G., 2020, A review on the long short-term memory model, *Artificial Intelligence Review*, Vol **53** (8). https://doi.org/10.1007/s10462-020-09838-1

[22] Karno, A. S. B., 2020, Prediksi Data Time Series Saham Bank BRI Dengan Mesin Belajar LSTM (Long ShortTerm Memory), *Journal of Informatic and Information Security*, Vol **1**(1). https://doi.org/10.31599/jiforty.v1i1.133

[23] Wang, W., Shao, J., & Jumahong, H., 2023, Fuzzy inference-based LSTM for long-term time series prediction, *Scientific Reports*, Vol **13**(1). https://doi.org/10.1038/s41598-023-47812-3

[24] Hossain, R., & Timmer, D. D., 2021, Machine learning model optimization with hyper parameter tuning approach, *Global Journal of Computer Science and Technology*, Vol **21**(2).

[25] Deng, Y., Fan, H., & Wu, S., 2023, A hybrid ARIMA-LSTM model optimized by BP in the forecast of outpatient visits, *Journal of Ambient Intelligence and Humanized Computing*, Vol **14**(5). `https://doi.org/10.1007/s12652-020-02602-x`

[26] St-Aubin, P., & Agard, B., 2022, Precision and Reliability of Forecasts Performance Metrics, *Forecasting*, Vol **4**(4). `https://doi.org/10.3390/forecast4040048`

[27] Soy Temür, A., & Yıldız, Ş., 2021, Comparison of Forecasting Performance of ARIMA LSTM and HYBRID Models for The Sales Volume Budget of a Manufacturing Enterprise, *Istanbul Business Research*, Vol **50**(1). `https://doi.org/10.26650/ibr.2021.51.0117`

[28] Kim, S., & Kim, H., 2016, A new metric of absolute percentage error for intermittent demand forecasts, *International Journal of Forecasting*, Vol **32**(3). `https://doi.org/10.1016/j.ijforecast.2015.12.003`