JURNAL ILMIAH MANAJEMEN BISNIS DAN INOVASI
UNIVERSITAS SAM RATULANGI (JMBI UNSRAT)

A COMPARATIVE ANALYSIS OF FINANCIAL PERFORMANCE FORECASTING MODELS: ARIMA, ARIMA-GARCH & LSTM IN INDONESIAN BANKING STOCKS

Zabrina Raissa
Universitas Pelita Harapan

A R T I C L E I N F O

Keyword: *Stock price prediction, Banking, ARIMA, ARIMA-GARCH, LSTM*

Kata Kunci: Prediksi harga saham, Perbankan, ARIMA, ARIMA-GARCH, LSTM

Corresponding author:

Zabrina Raissa
01617210008@student.uph.edu

Abstract. The banking sector is a crucial generator of economic activity and financial stability in Indonesia's stock market, so precise forecasting of bank stock prices is critical for making informed investment decisions. This study evaluates the forecasting performance of ARIMA, ARIMA-GARCH, and Long Short-Term Memory (LSTM) models for predicting daily closing prices of five key Indonesian banking stocks: BBCA.JK, BBNI.JK, BBRI.JK, BMRI.JK, and BNGA.JK. Using historical data from January 2022 to December 2024, the models are evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results show that the ARIMA model consistently outperforms ARIMA-GARCH and LSTM across all equities and measures. ARIMA has the lowest average MAE, RMSE, and MAPE values, at 74.46, 93.01, and 1.297%, respectively, demonstrating its reliability for static short-term price projections, surpassing both ARIMA-GARCH and LSTM in static forecasting. While ARIMA- GARCH incorporates volatility modelling, it provides only marginal improvements, and LSTM exhibits the weakest performance with higher error rates. These findings indicate that traditional econometric models, particularly ARIMA, are still excellent tools for projecting stock values in Indonesia's banking sector, allowing investors and analysts to make more informed decisions.

Abstrak. Sektor perbankan merupakan generator penting aktivitas ekonomi dan stabilitas keuangan di pasar saham Indonesia, sehingga peramalan harga saham bank yang tepat sangat penting untuk membuat keputusan investasi yang tepat. Studi ini mengevaluasi kinerja peramalan model ARIMA, ARIMA-GARCH, dan Long Short-Term Memory (LSTM) untuk memprediksi harga penutupan harian dari lima saham perbankan utama Indonesia: BBCA.JK, BBNI.JK, BBRI.JK, BMRI.JK, dan BNGA.JK. Dengan menggunakan data historis dari Januari 2022 hingga Desember 2024, model dievaluasi menggunakan Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan Mean Absolute Percentage Error (MAPE). Hasilnya menunjukkan bahwa model ARIMA secara konsisten mengungguli ARIMA-GARCH dan LSTM di semua ekuitas dan ukuran. ARIMA memiliki nilai MAE, RMSE, dan MAPE rata-rata terendah, masing-masing sebesar 74,46, 93,01, dan 1,297%, yang menunjukkan keandalannya untuk proyeksi harga jangka pendek statis, melampaui ARIMA-GARCH dan LSTM dalam peramalan statis. Sementara ARIMA-GARCH menggabungkan pemodelan volatilitas, ia hanya memberikan peningkatan marjinal, dan LSTM menunjukkan kinerja terlemah dengan tingkat kesalahan yang lebih tinggi. Temuan ini menunjukkan bahwa model ekonometrik tradisional, khususnya ARIMA, masih merupakan alat yang sangat baik untuk memproyeksikan nilai saham di sektor perbankan Indonesia, yang memungkinkan investor dan analis untuk membuat keputusan yang lebih tepat.

INTRODUCTION

Investing in stocks is very popular because it offers high returns through dividends and capital gains. However, these benefits also come with risks. Stock prices often move up and down, especially in Indonesia where the economy is still developing. Predicting stock prices is difficult because prices are influenced by many factors, such as economic conditions, company news, and investor behaviour. Many researchers (Appel, 2005; Brown et al., 1998; El-Nagar et al., 2022; Fromlet, 2001) have tried different methods to improve prediction accuracy. Because of this, investors and analysts need tools that can help them predict price movements and make better investment decisions.

In the context of Indonesia, the banking sector is one of the most influential components of the stock market, playing a vital role in supporting economic growth and maintaining financial stability (Huliselan, 2024). Given the sector's systemic importance for the country's economy, understanding its stock performance is useful for investors, analysts, and policymakers. With markets becoming more complex and uncertain, choosing the right forecasting model can help people make smarter investment choices.

ARIMA and ARIMA-GARCH are older models often used in time series forecasting (Kaur et al., 2023; Kontopoulou et al., 2023; Ospina et al., 2023), but LSTM is a newer deep learning method that can find patterns in data more effectively (Ahmed et al., 2022). Even though LSTM has shown good results in predicting stock prices in other countries, very few studies have tested it on Indonesian bank stocks. Most research focuses on global markets or different sectors, which leaves a gap in the local context.

This study aims to close this gap by comparing the ARIMA, ARIMA-GARCH, and Long Short-Term Memory (LSTM) techniques used in stock performance forecasting models for Indonesia's banking industry. Stock data collection, model training, and accuracy comparison using common error metrics like RMSE and MAE are all steps in the process. The objective is to determine which model best suits the Indonesian market in order to enhance future research and investment choices.

LITERATURE REVIEW

Several studies have been conducted to compare the accuracy of stock price prediction models. According to Jin et al. (2022), ARIMA greatly outperforms the prophet model in predicting Google stock closing prices, demonstrating its capacity to forecast the impact of events like as the Covid-19 pandemic. Previous research (Huang et al., 2023) found that the forecasting methods of the models influence their forecasting effects, with the ARIMA-GARCH model having the highest average forecasting accuracy in static forecasting and the LSTM model having the most accurate forecasting effect in dynamic forecasting, with an RMSE value of only 6.32%. In another study, Xiang (2022) compared ARIMA and ARIMA-GARCH models to analyze international oil price volatility. The findings revealed that the ARIMA(1,1,0)-GARCH(1,1) model provided the most accurate short-term forecasts, reducing the MAPE from 1.549% to 0.045% and the RMSE from 1.032 to 0.071. Collectively, these studies underscore the robustness and effectiveness of traditional econometric models such as ARIMA and ARIMA-GARCH for short-term time series forecasting, thereby motivating their use in this study.

With the advancement of computational technologies, more complex deep learning models have gained prominence in time series forecasting. Models such as Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), and Long Short-Term Memory (LSTM) are particularly adept at handling sequential data due to their ability to capture long-term dependencies (Benidis et al., 2023). For instance, Abbasimehr and Paki (2022) proposed a hybrid LSTM model augmented with multi-head attention mechanisms for time series forecasting. Their findings show that the proposed model consistently outperforms benchmark methods across multiple datasets in terms of symmetric mean absolute percentage error (SMAPE) and approximation ratio (AR). Similarly, Sirisha et al. (2022) compared traditional econometric models, including ARIMA and SARIMA, with an LSTM-based approach for profit forecasting. The LSTM model outperformed its counterparts, achieving an accuracy of 97.01%, compared to 93.84% and 94.38% for ARIMA and SARIMA, respectively. However, contrasting results are reported by Albeladi et al. (2023), who found ARIMA to outperform LSTM in forecasting property sector prices within the Mulkia Gulf Real Estate market. These inconsistencies highlight the importance of context-specific model evaluation, as the performance of forecasting models can vary significantly depending on the nature of the dataset and domain of application.

In the Indonesian banking sector, the application of time series forecasting models has received increasing attention in recent years. Sudipa et al. (2023) applied the ARIMA model to forecast the stock price trends of three major banks, BBCA.JK, BBRI.JK, and BMRI.JK, and reported low MAPE values of 4%, 5%, and 7%, respectively, indicating strong predictive performance. Jayaswara et al. (2023) employed Support Vector Regression (SVR) with linear and radial basis function (RBF) kernels to predict BBCA.JK stock prices, achieving a MAPE of 0.1021 and RMSE of 0.0499. Noviandy et al. (2024) explored the use of the Neural Prophet model to forecast bank stock prices and found it capable of accurately capturing both trends and minor fluctuations. These studies collectively demonstrate the growing interest and potential of time series forecasting models within the Indonesian banking sector. However, there is still a noticeable gap in the application of LSTM models specifically for predicting Indonesian banking stock prices, which presents a valuable research opportunity.

To address this gap, the present study applies the LSTM model to forecast daily closing prices of Indonesian banking stocks using a static prediction approach. Its performance is then compared to that of traditional econometric models, namely ARIMA and ARIMA-GARCH. To ensure optimal performance, rigorous model specification is conducted for the ARIMA and ARIMA-GARCH models, while hyperparameter tuning using a grid search strategy is implemented for the LSTM model. Furthermore, a series of diagnostic tests, including the Augmented Dickey-Fuller (ADF) test, Ljung-Box test, and Lagrange Multiplier (LM) test, are employed to validate the assumptions underlying the econometric models. This comprehensive approach allows for a robust and contextually relevant comparison of forecasting performance across modeling paradigms.

DATA AND METHODS

Data

The dataset used in this study comprises the historical daily closing prices of five of Indonesia's largest publicly listed banks: Bank Central Asia (BBCA.JK), Bank Negara Indonesia (BBNI.JK), Bank Rakyat Indonesia (BBRI.JK), Bank Mandiri (BMRI.JK), and Bank CIMB Niaga (BNGA.JK). The data spans from January 2022 to December 2024, resulting in 722 observations for each stock, after accounting for weekends and market holidays. All price data were retrieved from Yahoo Finance, a widely used and reliable source for financial market data. Figure 1 visualizes the historical price trends for each stock over the selected period, showing the fluctuations and overall movements that form the basis for the forecasting models evaluated in this study.

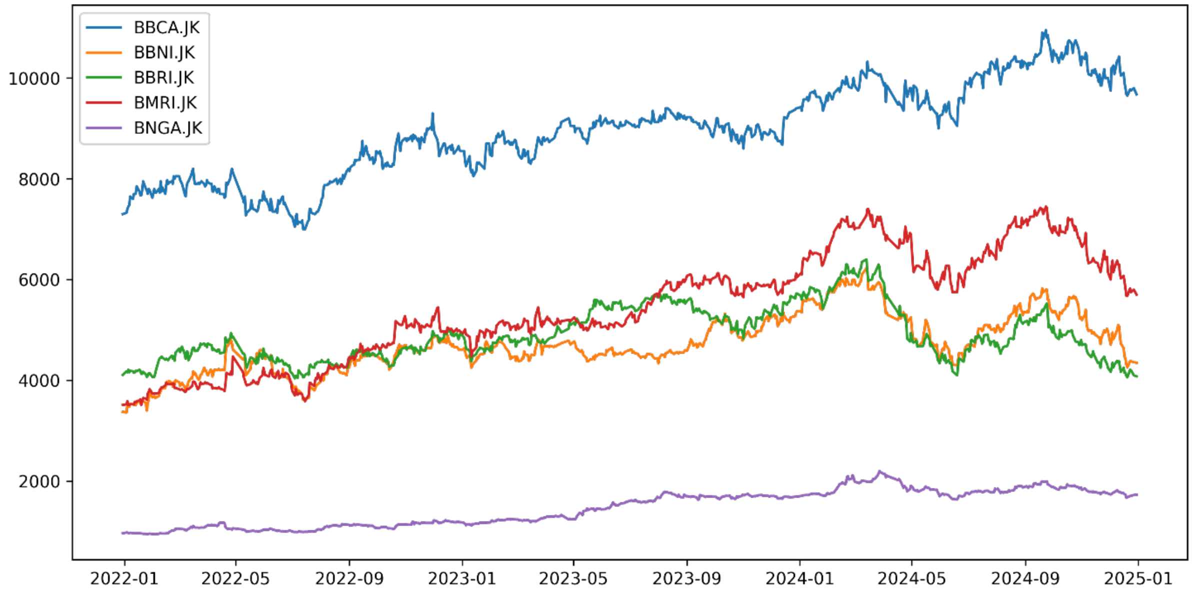


Figure 1. Historical Closing Prices of BBCA.JK, BBNI.JK, BBRI.JK, BMRI.JK, and BNGA.JK

Autoregressive Integrated Moving Average

The Autoregressive Integrated Moving Average (ARIMA) model is a foundational statistical approach widely applied in financial time series analysis, including stock price forecasting. It combines two main elements: Autoregressive (AR) terms, which reflect the dependency between current and past observations, and Moving Average (MA) terms, which account for the influence of previous prediction errors on the current observation.

The "Integrated" aspect allows ARIMA to handle non-stationary data by differencing the series d times to induce stationarity. Let Y_t represent the original time series, after applying differencing d times, the resulting stationary time series is denoted by,

$$W_t = \Delta^d Y_t.$$

The ARIMA(p,d,q) model can then be expressed as an ARMA(p,q) model as shown below,

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_q e_{t-q},$$

Where ϕ_i and θ_j denote the autoregressive and moving average coefficients, respectively, while e_t represents the residual errors, which are assumed to be white noise and uncorrelated with past values of the series.

Generalized Autoregressive Conditional Heteroskedasticity

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is widely used for modeling and forecasting time-varying volatility in time series data. Unlike traditional models that assume constant variance, GARCH captures the phenomenon of conditional heteroskedasticity, where the variance of the error terms evolves over time based on past information. This makes it particularly suitable for time series that exhibit volatility clustering, where periods of high and low variance alternate irregularly.

In the context of this study, GARCH is used in conjunction with ARIMA to form an ARIMA-GARCH framework. While ARIMA models the linear structure and trend in the conditional mean of the series, GARCH models the time-varying conditional variance of the residuals produced by ARIMA. This combination enhances the model's ability to capture both mean dynamics and volatility patterns in stock price movements. Let Y_t denote the observed time series, a traditional econometric model such as ARIMA estimates the conditional mean as follows,

$$Y_t = \mu(t, Y_{t-1}, \epsilon_{t-1}, Y_{t-2}, \ldots) + \epsilon_t,$$

where ϵ_t represent the error terms. This error term is further modeled as,

$$\epsilon_t = \sigma_t Z_t,$$

where Z_t ~ Normal IID (0,1), and σ_t^2 is the conditional variance at time t. Then the GARCH(p,q) model can be used to calculate the conditional variance σ_t^2 using the following formula,

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-1}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2.$$

with the assumptions that $\alpha_0 > 0$ and $\alpha_i, \beta_j \geq 0$ for every $i = 1, \ldots, p$ and $j = 1, \ldots, q$. The α_i coefficients capture the influence of past shocks (ARCH terms), while the β_j coefficients account for the persistence in volatility (GARCH terms). By jointly modeling the conditional mean with ARIMA and the conditional variance with GARCH, the ARIMA-GARCH framework provides a more robust and realistic approach to forecasting time series that exhibit both trend and volatility.

Long Short-Term Memory Neural Networks

Long Short-Term Memory (LSTM) networks are an advanced form of neural network architecture tailored for processing and learning from sequence-based data. They enhance Recurrent Neural Networks (RNNs) by overcoming common issues like vanishing and exploding gradients, which can make it difficult to train standard RNNs on long sequences.

The LSTM architecture introduces a gating mechanism to control the flow of information through time. It consists of three key gates: the input gate, the forget gate, and the output gate. These gates regulate what information is added, removed, or passed to the next time step. The corresponding equations are given below,

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i),$$
$$f_t = \sigma\left(w_f[h_{t-1}, x_t] + b_f\right),$$
$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o),$$

where i_t, f_t, and o_t denote the input gate, forget gate, and output gate, respectively. The terms w_k and b_k refer to the weight matrices and the biases corresponding to each gate. The variable h_{t-1} denotes the output from the previous LSTM unit at time step $t-1$, and x_t represents the current input at time t. Finally, σ denotes the sigmoid function.

While h_t captures the short-term memory in an LSTM unit, the long-term memory is maintained through the cell state, denoted by C_t. At each time step t g a candidate cell state, denoted as $\hat{C}_t$, is produced and regulated by the input gate. This candidate state is calculated using the following formula,

$$\hat{C}_t = \tanh\left(w_c[h_{t-1}, x_t] + b_c\right).$$

Next, the cell state is modified by merging the past cell state with the new candidate cell state, each influenced by the forget gate and input gate, respectively.

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t.$$

Finally, the output for the LSTM unit at time step t, denoted by h_t, can then be formulated as,

$$h_t = o_t * \tanh(C_t).$$

Research Flow

Figure 2 presents the overall research flow adopted in this study. The process begins with data preprocessing, which involves two primary steps: scaling and train-test splitting. Among common scaling techniques, which are normalization via min-max scaling and standardization, this study adopts standardization. This choice is based on the fact that stock prices do not have fixed upper or lower bounds, making min-max scaling less suitable. Following scaling, the dataset is split into training and testing sets using a 90:10 ratio, with the last three months of the data, October to December 2024, reserved for testing.

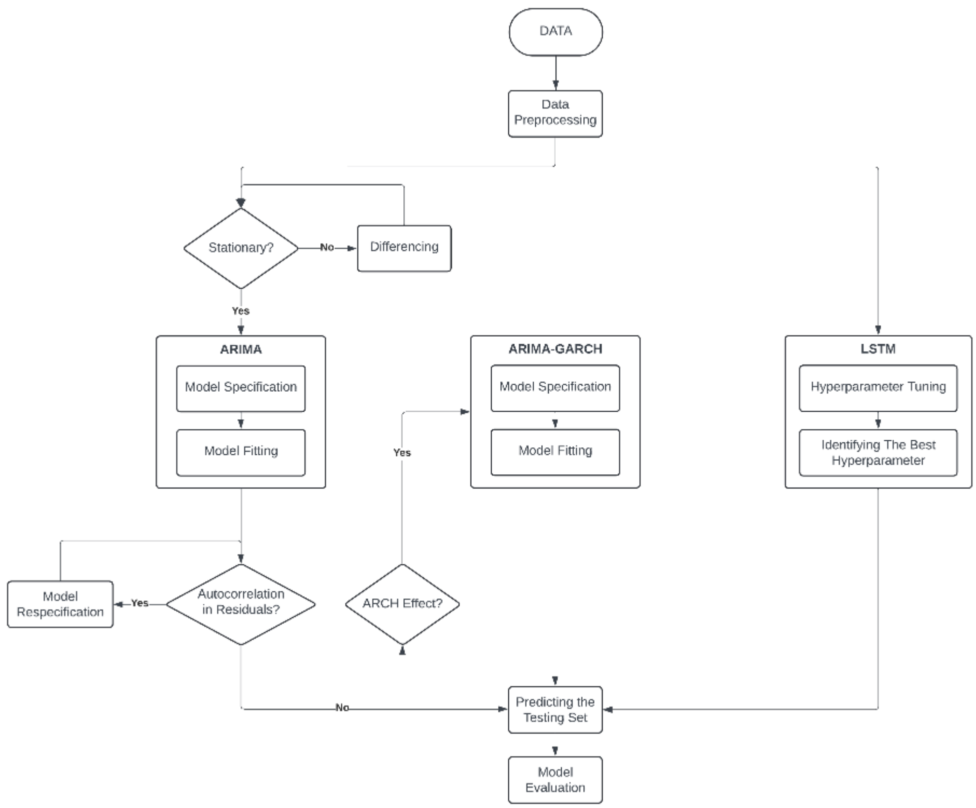


Figure 2. Research Flowchart

Once pre-processing is complete, the data is passed through the models, beginning with ARIMA. Since ARIMA models rely on the assumption of stationarity—where statistical characteristics like mean and variance remain constant over time—the Augmented Dickey-Fuller (ADF) test is applied to determine whether the time series exhibits unit roots, thereby validating the stationarity assumption. The hypotheses for the ADF test are as follows:

- $\boldsymbol{H_0}$: The time series contains a unit root, indicating non-stationarity.
- $\boldsymbol{H_a}$: The time series does not contain a unit root, indicating stationarity.

If stationarity is confirmed, the ARIMA model can be applied for forecasting. Model selection for ARIMA is guided by the Akaike Information Criterion (AIC), which serves as a metric to evaluate the trade-off between the model's goodness of fit and its complexity. Although the Bayesian Information

Criterion (BIC) imposes a stronger penalty for model complexity, this study prioritizes forecasting accuracy over model simplicity, making AIC the preferred metric. A grid search is used to explore values of p and q ranging from 0 to 9.

Once the model is defined, it is applied to the dataset, followed by diagnostic evaluations to assess how well it fits. A key part of this evaluation is checking for autocorrelation in the residuals, as the presence of autocorrelation suggests that the model has not fully captured the time-based patterns in the data. To test for this, the Ljung-Box test is used, which involves the following hypotheses:

- $\boldsymbol{H_0}$: The residuals are independent (no autocorrelation).
- $\boldsymbol{H_a}$: The residuals are autocorrelated.

If autocorrelation is detected, the model must be re-specified. Otherwise, the ARIMA model is deemed appropriate and can be used to generate forecasts on the test set. To evaluate the applicability of the ARIMA-GARCH model, the residuals of the ARIMA model are further examined for conditional heteroscedasticity (time-varying variance), also referred to as the ARCH effect. The Lagrange Multiplier (LM) test is used to detect this phenomenon, with hypotheses as follows:

- $\boldsymbol{H_0}$: The residuals have constant variance (homoscedastic).
- $\boldsymbol{H_a}$: The residuals have time varying variance (heteroscedastic).

If no heteroscedasticity is found, the ARIMA-GARCH model is not applicable. However, if the ARCH effect is present, the ARIMA-GARCH model is implemented. The ARIMA component retains the previously determined order, and a GARCH model is applied to the ARIMA residuals to model their conditional variance.

The LSTM modeling process in this study involves tuning several hyperparameters to determine the most suitable configuration for each stock. Before feeding the data into the LSTM model, the time series must be reshaped using a rolling window approach. This method converts the series into a tabular format, where a window of size *n* defines the number of past observations used as input features for the model.

The hyperparameters considered in this study include the number of LSTM units, the number of Dense layers, and the window size. Table 1 outlines the different LSTM architectures evaluated. A total of five model architectures are tested, each with three window sizes: 5, 10, and 21. These correspond to roughly one week, two weeks, and one month of trading days, respectively. In total, 15 different LSTM models are trained and assessed for each stock.

Table 1. LSTM Architectures for Hyperparameter Tuning

Architecture	LSTM layers	Dense layers
1	1	1
2	2	1
3	1	2
4	2	2
5	3	3

To select the best-performing configuration, each model is evaluated on a validation set, which is split from the original training set. To reduce the risk of overfitting, early stopping is applied with a minimum improvement threshold of 0.0001 and a patience of 5 epochs. All models are trained using the ADAM optimizer.

RESULTS AND DISCUSSION

This section outlines the key results of the research. The effectiveness of the forecasting models is evaluated using three metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).
Let m represent the total number of observations, and let y_i and $\hat{y}_i$ denote the actual and predicted closing prices of the i-th observation, respectively. The formulas for each metric are given below,

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i|,$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2},$$

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\frac{|y_i - \hat{y}_i|}{y_i},$$

ARIMA

As mentioned earlier, the use of econometric models such as ARIMA and ARIMA-GARCH requires the data to be stationary. Table 2 shows the results of the Augmented Dickey-Fuller (ADF) test on the stock closing prices, both in their original form and after first differencing. It can be seen that the p-values for the raw prices are all above 0.05, meaning the null hypothesis of non-stationarity cannot be rejected. However, after applying one round of differencing, the p-values drop to 0 or close to 0, indicating that the differenced data is now stationary. Since only one differencing step is needed, the *d* parameter in the ARIMA(*p, d, q*) model is set to 1.

Table 2. ADF Test Results Before and After Differencing

Stock	Before Differencing		After Differencing	
	ADF Test Statistics	P-value	ADF Test Statistics	P-value
BBCA.JK	-1.31007	0.62455	-20.62370	0
BBNI.JK	-2.33446	0.16110	-9.43712	5.01E-16
BBRI.JK	-2.05186	0.26431	-27.78778	0
BMRI.JK	-1.06224	0.72989	-20.77265	0
BNGA.JK	-0.88051	0.79432	-19.13740	0

With the data now stationary, the next step is to determine the best values for *p* and *q*. As previously mentioned, this study employs the Akaike Information Criterion (AIC) to determine the optimal model, testing both p and q values within the range of 0 to 9. Table 3 presents the selected ARIMA(*p, d, q*) model for each stock and their corresponding AIC values. Among the stocks, BNGA.JK has the lowest AIC with a value of 5971.168, indicating that ARIMA(2,1,2) fits the data best. In contrast, BBCA.JK has the highest AIC with a value of 8116.223, suggesting that the ARIMA model does not capture the pattern in the data as effectively.

Table 3. Model Specifications for ARIMA

Stock	ARIMA(p,d,q)	AIC
BBCA.JK	ARIMA(2,1,4)	8116.223338
BBNI.JK	ARIMA(2,1,3)	7618.638456
BBRI.JK	ARIMA(3,1,2)	7643.985974
BMRI.JK	ARIMA(0,1,2)	7843.950972
BNGA.JK	ARIMA(2,1,2)	5971.167512

Once the models are fitted, a diagnostic check is performed to see how well they represent the time series. One key step is to examine whether the residuals are autocorrelated. Table 4 displays the results of the Ljung-Box test on the residuals. All p-values are above 0.05, so the null hypothesis that the residuals are not autocorrelated cannot be rejected. This means the ARIMA models have successfully captured the time-based patterns in the data.

Table 4. Ljung-Box Test Results On ARIMA Residuals

Stock	Ljung-Box Test Statistics	P-value
BBCA.JK	3.99668	0.94750
BBNI.JK	13.14917	0.21545
BBRI.JK	4.59160	0.91674
BMRI.JK	4.15800	0.93994
BNGA.JK	2.64150	0.98865

ARIMA-GARCH

To support the use of GARCH, the residuals from the ARIMA models must show signs of time-varying volatility, also known as the ARCH effect. As mentioned in Section 3, this study uses the Lagrange-Multiplier (LM) test to check for its presence. Table 5 shows the LM test results on the ARIMA residuals. Since all p-values are lower than the 0.05 significance level, the null hypothesis of constant variance is rejected. This confirms that the ARCH effect is present and justifies the application of ARIMA-GARCH models for all stocks.

Table 5. Lagrange Multiplier Test Results on ARIMA Residuals

Stock	LM Test Statistics	P-value
BBCA.JK	31.04492	0.00058
BBNI.JK	23.00505	0.01073
BBRI.JK	27.31910	0.00232
BMRI.JK	32.67713	0.00031
BNGA.JK	77.84834	0

The GARCH model is then fitted to the residuals from the ARIMA models. The best combination of *p* and *q* values is selected using the same method as with ARIMA. Table 6 lists the selected GARCH (*p, q*) model for each stock, along with their AIC values. A similar trend to the ARIMA model selection is seen here, where BBCA.JK has the highest AIC at 8077.774, while BNGA.JK has the lowest at 5813.452, indicating stronger model fit for BNGA.JK.

Table 6. Model Specifications For GARCH

Stock	GARCH(p,q)	AIC
BBCA.JK	GARCH(6,0)	8077.774
BBNI.JK	GARCH(3,5)	7578.291
BBRI.JK	GARCH(1,1)	7595.273
BMRI.JK	GARCH(1,4)	7758.787
BNGA.JK	GARCH(7,8)	5813.452

LSTM

To determine the best hyperparameters for the LSTM models, hyperparameter tuning using a grid search approach is performed. Table 7 presents the selected architecture and window size for each stock, along with their corresponding MAE, RMSE, and MAPE scores on the validation set. A notable finding is that simpler architectures tend to yield better performance. This is reflected in the results, where four out of five stocks perform best with the 1 LSTM layer and 2 Dense layers structure. BBCA.JK is the only exception, performing best with the simpler 1 LSTM layer and 1 Dense layer configuration. This observation aligns with the findings of Bhandari et al. (2022), who showed that a single-layer LSTM outperforms more complex multi-layer LSTM models in stock market index prediction tasks, reinforcing the effectiveness of simpler LSTM structures in this context.

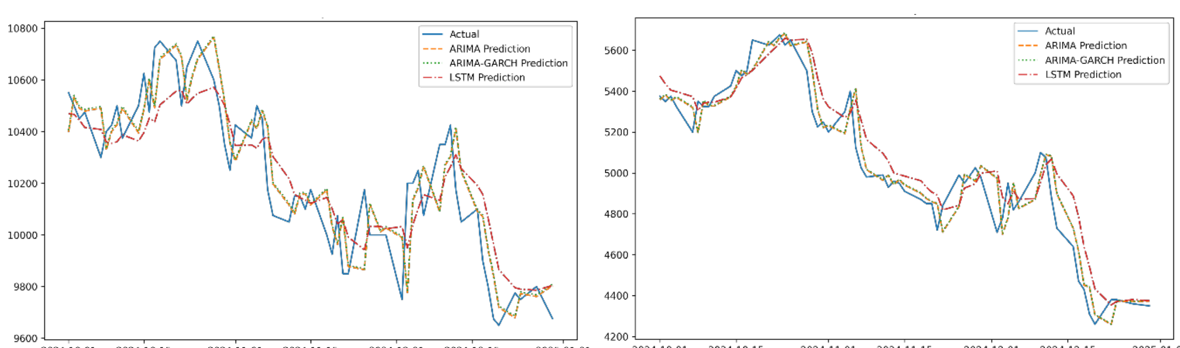
Table 7. LSTM Hyperparameter Tuning Results

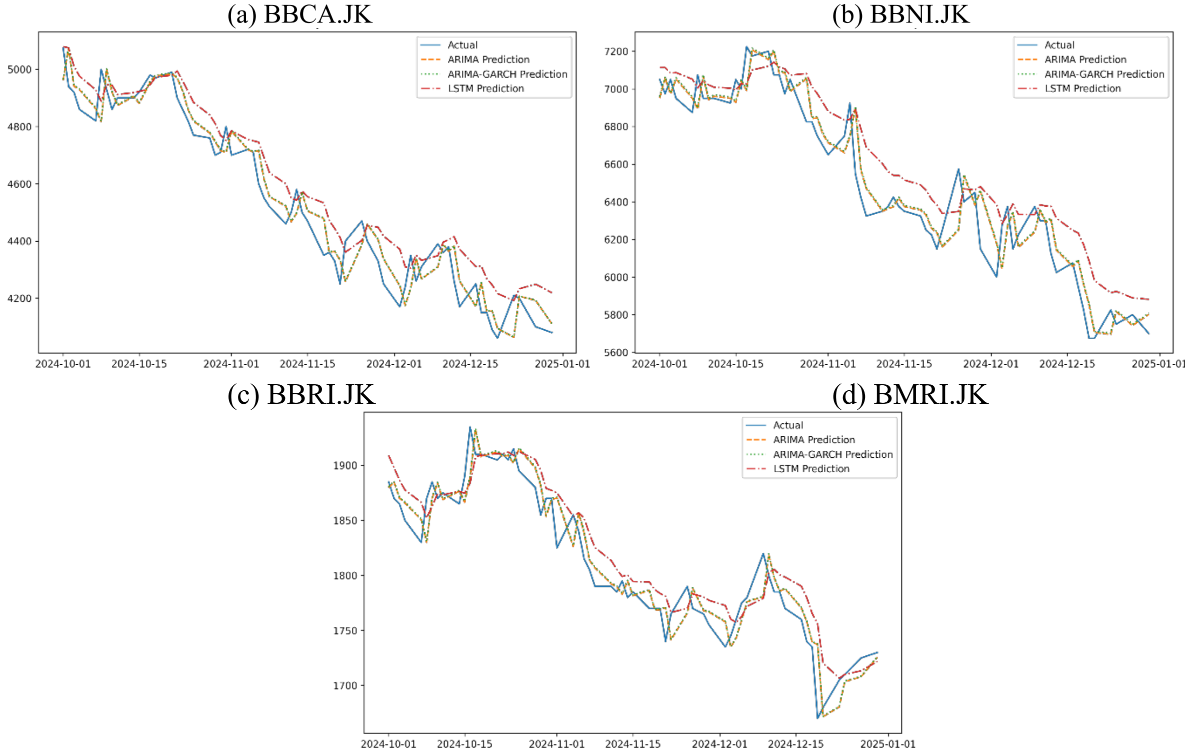
Stock	LSTM Structure	Window Size	MAE	RMSE	MAPE
BBCA.JK	1 LSTM, 1 Dense	5	123.36755	160.03611	0.01180
BBNI.JK	1 LSTM, 2 Dense	5	77.06643	98.32380	0.01419
BBRI.JK	1 LSTM, 2 Dense	5	76.97877	107.10213	0.01518
BMRI.JK	1 LSTM, 2 Dense	10	94.95347	112.60717	0.01338
BNGA.JK	1 LSTM, 2 Dense	5	21.54737	28.63837	0.01144

A smaller window size also tends to deliver better performance than larger ones. This is shown by the fact that four out of five stocks perform best with a window size of 5, while only BMRI.JK achieves the best results with a window size of 10. Notably, the largest window size of 21 is not selected by any of the stocks. This suggests that more recent closing prices play a more important role in predicting the next day's closing price than longer historical data. This finding aligns with previous studies such as (Baek, 2024), (Alam et al., 2024), and (Shen & Shafiq, 2020), that also highlight the effectiveness of utilizing more recent data in financial modelling.

Forecasting and Evaluations

The final step of this study, as outlined in the research flow in Section 3, is to use the trained models to forecast on the testing set and evaluate their performance using MAE, RMSE, and MAPE. Figure 3 displays the forecasted outcomes of the ARIMA, ARIMA-GARCH, and LSTM models for each stock, alongside the actual closing prices for comparison. It can be observed that the ARIMA and ARIMA-GARCH forecasts are nearly identical, with ARIMA-GARCH generally producing slightly higher predicted values. Interestingly, both ARIMA-based models more closely follow the actual price trends compared to the LSTM model. The LSTM predictions frequently overestimate or underestimate the closing prices, indicating that traditional econometric models may be better suited for forecasting static closing prices in Indonesian banking stocks. This tendency reflects a potential limitation in the LSTM's ability to generalize well in the context of relatively stable, short-term price movements—common in blue-chip banking stocks in emerging markets like Indonesia. From a financial perspective, these deviations could translate into suboptimal decisions, particularly in applications such as short-term trading or risk management, where accuracy in directional trends and volatility is crucial.





(a) BBCA.JK (b) BBNI.JK

(c) BBRI.JK (d) BMRI.JK

(e) BNGA.JK

Figure 3. Model Forecasting Results

Tables 8 through 10 present the evaluation results across all models using MAE, RMSE, and MAPE, respectively, with the lowest errors for each stock highlighted in orange. The results show that the ARIMA model consistently outperforms both ARIMA-GARCH and LSTM across all evaluation metrics and all stocks. Specifically, ARIMA achieves an average MAE of 74.46, outperforming ARIMA-GARCH (75.25) and LSTM (99.19). Similarly, ARIMA records the lowest average RMSE at 93.01, followed by ARIMA-GARCH at 93.69 and LSTM at 120.76. For MAPE, ARIMA again leads with an average of 1.297%, compared to 1.31% for ARIMA-GARCH and 1.784% for LSTM. This consistency is particularly important for institutional investors or analysts seeking low-variance predictive tools for conservative forecasting strategies. The addition of GARCH terms, while theoretically beneficial for modeling volatility, does not significantly improve forecast accuracy in this static closing price prediction setting, as evidenced by the slightly higher error rates in ARIMA-GARCH models.

Table 8. Model Evaluations Based on MAE

Stock	ARIMA	ARIMA-GARCH	LSTM
BBCA.JK	117.27022	118.12722	124.65589
BBNI.JK	76.09393	76.48107	104.13578
BBRI.JK	60.68905	61.48652	85.99196
BMRI.JK	102.50851	104.33033	161.25147
BNGA.JK	15.75651	15.84385	19.92337
AVERAGE	74.46365	75.25380	99.19169

Table 9. Model Evaluations Based on RMSE

Stock	ARIMA	ARIMA-GARCH	LSTM
BBCA.JK	142.47402	143.26421	150.70642

BBNI.JK	100.10157	100.59285	133.61291
BBRI.JK	73.40485	73.88804	101.56475
BMRI.JK	128.94002	130.49406	192.45071
BNGA.JK	20.10717	20.22714	25.47421
AVERAGE	93.00553	93.69326	120.76180

Table 10. Model Evaluations Based on MAPE

Stock	ARIMA	ARIMA-GARCH	LSTM
BBCA.JK	0.01148	0.01157	0.01222
BBNI.JK	0.01525	0.01533	0.02108
BBRI.JK	0.01348	0.01366	0.01936
BMRI.JK	0.01589	0.01618	0.02546
BNGA.JK	0.00872	0.00877	0.01110
AVERAGE	0.01297	0.01310	0.01784

On the other hand, the LSTM model consistently yields the highest error values across all metrics and stocks, indicating the weakest performance among the three. This outcome suggests that despite being a more modern and complex approach, LSTM does not necessarily outperform traditional models like ARIMA for this specific task. Supporting this observation, Zhang et al. (2022) found that ARIMA outperforms LSTM for monthly and weekly stock price forecasts. Similarly, Kobiela et al. (2022) reported that ARIMA delivers better results than LSTM when relying solely on historical price data as input. Overall, this study demonstrates that in the context of forecasting static closing prices for Indonesian banking stocks, traditional econometric approaches remain highly competitive, offering both accuracy and interpretability for practical financial applications.

CONCLUSION

This study set out to compare the performance of ARIMA, ARIMA-GARCH, and LSTM models in forecasting the closing prices of major Indonesian banking stocks. By applying rigorous model specifications, hyperparameter tuning, and standard error metrics such as MAE, RMSE, and MAPE, the analysis revealed that traditional econometric models, particularly ARIMA, consistently outperformed the more complex LSTM model in static forecasting scenarios. These results highlight that, in the context of short-term static predictions for Indonesian banking stocks, simpler and well-established models remain more effective than advanced deep learning techniques. For investors, analysts, and policymakers operating in emerging markets like Indonesia, this suggests that leveraging traditional models may offer more reliable guidance in portfolio decisions and financial planning.

By addressing the current research gap in the application of LSTM and ARIMA-based models to Indonesian bank stocks, this study provides a valuable benchmark for future work. Further research could explore hybrid approaches or dynamic forecasting horizons, and incorporate macroeconomic or sentiment-based variables to enhance model performance in more volatile or information-sensitive environments.

REFERENCES

Abbasimehr, H., & Paki, R. (2022). Improving time series forecasting using LSTM and attention models. *Journal of Ambient Intelligence and Humanized Computing*, *13*(1), 673–691. https://doi.org/10.1007/s12652-020-02761-x

Ahmed, D. M., Hassan, M. M., & Mstafa, R. J. (2022). A Review on Deep Sequential Models for Forecasting Time Series Data. *Applied Computational Intelligence and Soft Computing*, *2022*, 1–19. https://doi.org/10.1155/2022/6596397

Alam, K., Bhuiyan, M. H., Haque, I. U., Monir, M. F., & Ahmed, T. (2024). Enhancing Stock Market Prediction: A Robust LSTM-DNN Model Analysis on 26 Real-Life Datasets. *IEEE Access*, *12*, 122757–122768. https://doi.org/10.1109/ACCESS.2024.3434524

Albeladi, K., Zafar, B., & Mueen, A. (2023). Time Series Forecasting using LSTM and ARIMA. *International Journal of Advanced Computer Science and Applications*, *14*(1). https://doi.org/10.14569/IJACSA.2023.0140133

Baek, H. (2024). A CNN-LSTM Stock Prediction Model Based on Genetic Algorithm Optimization. *Asia-Pacific Financial Markets*, *31*(2), 205–220. https://doi.org/10.1007/s10690-023-09412-z

Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, Y., Maddix, D., Turkmen, C., Gasthaus, J., Bohlke-Schneider, M., Salinas, D., Stella, L., Aubet, F.-X., Callot, L., & Januschowski, T. (2023). Deep Learning for Time Series Forecasting: Tutorial and Literature Survey. *ACM Computing Surveys*, *55*(6), 1–36. https://doi.org/10.1145/3533382

Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. C. (2022). Predicting stock market index using LSTM. *Machine Learning with Applications*, *9*, 100320. https://doi.org/10.1016/j.mlwa.2022.100320

Guna Jayaswara, D., Slamet, I., & Susanti, Y. (2023). Prediction of Central Asia Bank's Stock Price using Support Vector Regression Method. In *Science and Education* (Vol. 2).

Huliselan, M. (2024). THE The Analysis Of The Impact Of Return On Asset, Return On Equity, And Ratio Of Liquidity, On Overall Company Performance: Case Study In Banking Sector Listed Indonesia Stock Exchange. *Indo-Fintech Intellectuals: Journal of Economics and Business*, *4*(6), 3808–3816. https://doi.org/10.54373/ifijeb.v4i6.2761

Kaur, J., Parmar, K. S., & Singh, S. (2023). Autoregressive models in environmental forecasting time series: a theoretical and application review. *Environmental Science and Pollution Research*, *30*(8), 19617–19641. https://doi.org/10.1007/s11356-023-25148-9

Khater AA, El-Nagar AM, El-Bardini M, EL-Rabaie NM. (2000) Online learning based on adaptive learning rate for a class of recurrent fuzzy neural network. *Neural Comput Appl, 32,* 8691-8710. https://doi.org/10.1007/s00521-019-04372-w

Kobiela, D., Krefta, D., Król, W., & Weichbroth, P. (2022). ARIMA vs LSTM on NASDAQ stock exchange data. *Procedia Computer Science*, *207*, 3836–3845. https://doi.org/10.1016/j.procs.2022.09.445

Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. *Future Internet*, *15*(8), 255. https://doi.org/10.3390/fi15080255

Noviandy, T. R., Hardi, I., & Idroes, G. M. (2024). Forecasting Bank Stock Trends Using Artificial Intelligence: A Deep Dive into the Neural Prophet Approach. *The International Journal of Financial Systems*, *2*(1), 29–56. https://doi.org/10.61459/ijfs.v2i1.41

Ospina, R., Gondim, J. A. M., Leiva, V., & Castro, C. (2023). An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil. *Mathematics*, *11*(14), 3069. https://doi.org/10.3390/math11143069

Pan, S., Ji, S., Jin, D., Xia, F., & Yu, P. S. (2022). Guest Editorial: graph-powered machine learning in future-generation computing systems. *Future Generation Computer Systems*, 126, 88-90. https://doi.org/10.1016/j.future.2021.08.005

Shen, J., & Shafiq, M. O. (2020). Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of Big Data*, *7*(1), 66. https://doi.org/10.1186/s40537-020-00333-6

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*, *10*, 124715–124727. https://doi.org/10.1109/ACCESS.2022.3224938

Sudipa, I. G. I., Riana, R., Putra, I. N. T. A., Yanti, C. P., & Aristana, M. D. W. (2023). Trend Forecasting of the Top 3 Indonesian Bank Stocks Using the ARIMA Method. *SinkrOn*, *8*(3), 1883–1893. https://doi.org/10.33395/sinkron.v8i3.12773

Xiang, Y. (2022). Using ARIMA-GARCH Model to Analyze Fluctuation Law of International Oil Price. *Mathematical Problems in Engineering*, *2022*, 1–7. https://doi.org/10.1155/2022/3936414

Zhang, R., Song, H., Chen, Q., Wang, Y., Wang, S., & Li, Y. (2022). Comparison of ARIMA and LSTM for prediction of hemorrhagic fever at different time scales in China. *PLOS ONE*, *17*(1), e0262009. https://doi.org/10.1371/journal.pone.0262009