Journal of Artificial Intelligence and Technology Information (JAITI)
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, June 2026
https://doi.org/10.58602/jaiti.v4i2.260
Page 203-216

Comparative Analysis of Hybrid ARIMA-LSTM against Statistical and Machine Learning Benchmarks for Commodity Stock

Muhammad Iszul Wilsa[1*], Heru Purnomo Kurniawan[2], Rizki Dewantara[3], Dinda Febrihastatiwi[4], Indri Setiawati[5]
[1,2,3,5] Informatika, Universitas Islam Negeri Siber Syekh Nurjati Cirebon, Cirebon, Indonesia
[4] Master's Program in Science Education, Universitas PGRI Semarang, Semarang, Indonesia
[1*]iszulwilsa@uinssc.ac.id, [2]herupurnomo@uinssc.ac.id, [3]dewantararizki@uinssc.ac.id, [4]dinda8235@guru.smp.belajar.id, [5]indrisetiawati0830@uinssc.ac.id

Abstract: Predicting stock prices in Indonesia's commodities and energy sectors is a complex challenge due to high volatility influenced by global market dynamics and macroeconomic factors. This study aims to test the robustness of the ARIMA-LSTM hybrid model in predicting closing stock prices for six major issuers: ADRO, PTBA, MEDC, ANTM, MDKA, and AALI. The proposed approach employs a dual-input strategy that integrates 27 technical indicators with the linear residuals from the ARIMA model. The research methodology begins with data decomposition using the ARIMA model to capture linear components, followed by modeling the residuals using Long Short-Term Memory (LSTM) to capture complex non-linear patterns. The experimental results show that the hybrid model consistently delivers the best performance compared to single models such as ARIMA, Random Forest, and Single LSTM across all test datasets. In the 1-step-ahead scenario, the hybrid model achieved the lowest average MAPE of 2.10%, while in the 5-step-ahead scenario, the error rate remained at 4.00%. The statistical validity of these improvements was confirmed via the Wilcoxon signed-rank test, demonstrating statistically significant performance gains in four out of six stocks across both forecasting horizons. A key finding of this research is the hybrid architecture's ability to mitigate the extreme overfitting experienced by the Single LSTM model, while providing better prediction stability against variations in issuer characteristics. This study concludes that the integration of statistical decomposition and deep learning provides a reliable framework for investors and analysts to make data-driven decisions amid the volatile fluctuations of the Indonesian capital market.
Keywords: ARIMA-LSTM; Stock Prices; Commodities; Deep Learning; Forecasting

1. INTRODUCTION

Stock price forecasting is the most complex challenge in the field of finance, due to the complexity of its time-series characteristics and the volume of financial data [1]. Stock price fluctuations such as rises and falls are driven in part by supply and demand in the market, which often makes it difficult for investors to predict the future prospects of stock

Muhammad Iszul Wilsa: *Corresponding Author

Journal of Artificial Intelligence and Technology Information (JAITI)
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, June 2026
https://doi.org/10.58602/jaiti.v4i2.260
Page 203-216

investments [2]. In the financial markets, accurate stock price predictions are particularly helpful for investors [3].

In recent years, time series forecasting in economics has become a widely studied topic [4]. The issue of prediction accuracy will remain an ongoing challenge, as economic time-series data itself contains a great deal of noise and is non-stationary, making it difficult to predict stock prices [3]. One source of noise in economic time series data is volatility; stock price volatility is an important measure of financial risk and serves as an indicator in decision-making regarding risk management and asset allocation [4]. Movements in the Indonesian stock index are influenced by several factors, both internal such as corporate fundamentals and external such as macroeconomic conditions and commodity prices [5]. The volatility of Indonesian stocks, particularly in the energy and commodities sectors, is influenced by commodity price dynamics, changes in trading volume, and macroeconomic events such as pandemics, geopolitical conflicts, and trade policies [6], [7].

The Autoregressive Integrated Moving Average (ARIMA) model is one of the most popular and widely used methods for forecasting financial time series [8], [9], [10]. The Arima model is an efficient and robust econometric model that is widely used to forecast short-term financial time series data [8]. A study of the Philippine Stock Exchange Index found that the ARIMA model demonstrated strong accuracy in short-term and daily forecasts [9]. ARIMA performs well when dealing with linear and clear time series data, but its performance is less effective when dealing with nonlinear relationships or complex patterns [10].

Over time, machine learning methods such as random forests have been used for stock prediction [11], [12]. Random forests are capable of handling complex relationships, reduce overfitting compared to single decision trees, and produce reliable predictions through ensemble learning [11], [13]. However, the characteristics of financial time-series data pose a challenge for random forests, which have inherent limitations in capturing temporal dependencies, adapting to market changes, and competing with deep learning methods [11], [14].

Long Short-Term Memory (LSTM), a deep learning method frequently used in stock price forecasting, is one of the deep learning methods employed for stock price prediction due to its ability to handle sequential data that is uncertain, noisy, and nonlinear [8], [15]. With its unique architecture, LSTM has the ability to retain important information over long periods of time and selectively discard irrelevant information, making it ideal for use in financial data analysis [16], [17]. It is this ability to remember that makes LSTM so useful, as it has a significant impact on future values [10]. One of the most critical challenges in using LSTM for stock price prediction is its high sensitivity to hyperparameter selection [3]. In addition, there is a risk of overfitting; although neural network models can achieve good generalization, they are prone to overfitting due to their high capacity [18]. The characteristics of financial time-series data make LSTMs prone to substantial prediction errors, as the model must simultaneously capture linear and nonlinear patterns, long-term trends and short-term fluctuations, as well as meaningful signals and random noise from raw, undecomposed data [19], [20].

Hybrid forecasting models have become a common approach to forecasting financial time series; this approach aims to address the nonlinearity, volatility, and disturbances inherent in financial data. These models combine traditional statistical techniques (such as ARIMA and GARCH) with machine learning and deep learning methods (such as LSTM, CNN, SVM, and XGBoost), often incorporating decomposition or ensemble strategies to capture both linear and nonlinear patterns [21], [22]. Based on that, [10] tested the hybrid ARIMA-LSTM model using maritime stock data such as CPLP, ESEA, GOGL, SHIP, PXS, and GASS over a five-year period from 2017 to 2022. This hybrid model combines the strengths of each model ARIMA's ability to handle linear data and LSTM's ability to handle nonlinear data as a solution to the noise and volatility in stock price time series.

Muhammad Iszul Wilsa: *Corresponding Author

Journal of Artificial Intelligence and Technology Information (JAITI)
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, June 2026
https://doi.org/10.58602/jaiti.v4i2.260
Page 203-216

In this article, the ARIMA-LSTM hybrid model proposed by [10] is used to test the robustness of this hybrid architecture specifically on highly volatile stocks in Indonesia, namely those in the commodities sector. The test results consist of a performance comparison between statistical models such as ARIMA, machine learning models such as random forests, deep learning models such as LSTM, and the ARIMA-LSTM hybrid model.

2. RESEARCH METHODOLOGY

This study was conducted in several stages, beginning with data collection, data preprocessing, feature extraction, statistical testing, model development, and the evaluation of results. To test the model's robustness, this study utilized six datasets from companies in the commodities sector in Indonesia, as described in Table 1, covering a 10-year period from January 2015 to January 2025 from Yahoo Finance API. To evaluate these models without data leakage, a 10-year historical dataset was divided chronologically into a training dataset (80%) and a test dataset (20%). The first eight years were used to train the models and optimize their weights, whilst the final two years served specifically as a previously unseen test dataset to validate the forecasts' robustness across various market conditions.

Table 1. Commodity Stock Dataset

Ticker	Company	Main Sector
ADRO	Adaro Energy Indonesia Tbk.	Coal Mining
PTBA	Bukit Asam Tbk.	Coal Mining
MEDC	Medco Energi Internasional Tbk	Oil and Gas
ANTM	Aneka Tambang Tbk.	Mineral Mining
MDKA	Merdeka Copper Gold Tbk.	Gold and Copper Mining
AALI	Astra Agro Lestari Tbk.	Agriculture

During data preprocessing and feature extraction, to improve the model's predictive capabilities, 27 technical variables were extracted, including lag features (historical prices from t-1 to t-5), moving averages (SMA and EMA with periods of 5, 10, and 20), volatility indicators (standard deviation and Bollinger Bands), and momentum indicators such as the Rate of Change (ROC) and Relative Strength Index (RSI). The data was then normalized using the MinMaxScaler to a range of 0 to 1 to accelerate the convergence process in the deep learning model, with the target variable being the daily closing price. In the testing conducted, fixed parameters were used, meaning the entire dataset used the same parameters. This was done to test the model's robustness against the data. Testing was performed with 1-step-ahead predictions (one day later) and 5-step-ahead predictions (one week in the stock market).

Brock Dechert Scheinkman(BDS)

The Brock-Dechert-Scheinkman (BDS) test is the most widely used nonparametric statistical test for detecting nonlinear dependencies and deterministic structures in data [23]. In this case, the BDS test is conducted to determine whether the residuals from the linear model (ARIMA) still exhibit a nonlinear structure. If the p-value is <0.05, the data contains a nonlinear structure, which supports the use of a hybrid model.

ARIMA

The Auto Regressive Integrated Moving Average (ARIMA) is a statistical model capable of handling time series data by better understanding the data and processing it to generate future predictions. The ARIMA model itself is an extension of the Autoregressive Moving Average (ARMA) model, which can only model stationary time series data [24].

$$y'_t = c + \emptyset_1 y'_{t-1} + \cdots + \emptyset_1 y'_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_1 \epsilon_{t-p} + \epsilon_t \quad (1)$$

Muhammad Iszul Wilsa: *Corresponding Author

Journal of Artificial Intelligence and Technology Information (JAITI)
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, June 2026
https://doi.org/10.58602/jaiti.v4i2.260
Page 203-216

In a non-seasonal ARIMA model, typically denoted as ARIMA (p,d,q), where p represents the order or number of time lags, d represents the order of differentiation, and q represents the order of the moving average model [24]. In this study, the ARIMA (p,d,q) parameters were determined using the automatic parameter selection method from Auto-ARIMA based on the lowest Akaike Information Criterion (AIC) value.

Random Forest

Random Forest is an ensemble learning algorithm that combines predictions from several independent decision trees to produce more accurate and stable predictions [11], [13]. The prediction process using Random Forest is illustrated in Fig. 1.

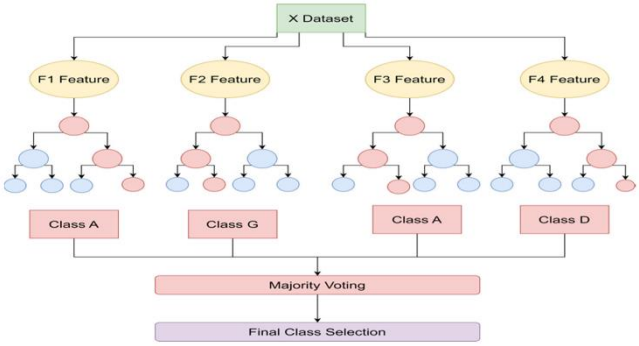


Figure 1. Random Forest [13]

In general, random forests can produce good results on large datasets; however, this increases the computational load, which can impact performance due to the large number of trees [13]. In this case, the random forest builds 100 decision trees, and the final prediction is the average of all these trees to improve stability and reduce variance. The inputs used are the 27 technical indicators described earlier.

LSTM

LSTM is an improved version of RNN, designed to address the vanishing gradient and exploding gradient problems, which are the main weaknesses of RNNs in processing long-term time series data [11], [25], unlike RNNs, which cannot capture high-order dependencies and are prone to the vanishing gradient problem [26].

$$f_t = \sigma(W_f X_t + U_f h_{t-1} + b_f) \quad (2)$$

In the first stage, the Forget Gate (f_t) is responsible for selecting information and determining whether information from the previous cell state should be discarded or retained using a sigmoid activation function (σ), Then, using the weights learned (W_f) is combined with the new input in the current step (X_t), during training to identify which features are irrelevant, the hidden state from the previous step (h_{t-1}) with using hidden-to-hidden weight (U_f) and a bias (b_f)is added to shift the activation function so that the model becomes more flexible as shown in equation (2) [25].

$$i_t = \sigma(W_i X_t + U_i h_{t-1} + b_i) \quad (3)$$

$$\hat{C}_t = tanh(W_c X_t + U_c h_{t-1} + b_c) \quad (4)$$

The information is then passed to the Input Gate (i_t), which determines which new information from the current input (X_t) and the previous hidden state(h_{t-1}) should be stored in the cell state. As shown in equation (3), this gate uses the sigmoid activation function (σ) to produce a value between 0 and 1, which acts as a filter for that information. At the same time, a new candidate value vector ($\hat{C}_t$) is generated using the ($tanh$) activation

Muhammad Iszul Wilsa: *Corresponding Author

Journal of Artificial Intelligence and Technology Information (JAITI)
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, June 2026
https://doi.org/10.58602/jaiti.v4i2.260
Page 203-216

function as described in equation (4) [25]. This vector represents new information that may be added to long-term memory (the cell state). In these equations, (W_i) and (W_c) represent the weight matrices for the inputs, while (U_i) and (U_c) (referred to as (U_f) in some variants) are the recurrent weight matrices for the previous hidden states. The bias terms (b_i) and (b_c) allow the model to learn thresholds for updating information, ensuring the system selectively retains only the temporal patterns most relevant for stock price prediction.

$$O_t = \sigma(W_O X_t + U_O h_{t-1} + b_O) \quad (5)$$

$$h_t = O_t \times tanh\,(C_t) \quad (6)$$

The final step is the output gate (O_t) which determines which information from the current cell state is selected to be sent as the final output. As described in Equation (5) the output gate uses a sigmoid activation function to filter the current input (X_t) and the previous hidden state (h_{t-1}). Once the output is determined, the cell state (C_t) is passed through the ($tanh$) function to be normalized, and the result is multiplied by the output gate result as shown in equation (6) [25] to produce the new hidden state (h_t).

In this study, the model was built with an architectural specification optimized to handle 27 technical features as inputs. The data was transformed into a three-dimensional space comprising samples, time steps, and features, with a one-day lookback period. The configuration used consisted of 64 hidden layers with a $softsign$ activation function. Regularization was performed using Dropout with a rate of 0.25, a dense layer with a fully connected layer containing 32 units, and $ReLu$ activation. The model was trained using Adam optimization with a learning rate of 0.001. The training process was conducted for 100 epochs with an early stopping mechanism (patience = 15) to ensure efficiency and prevent overfitting.

The LSTM model architecture was determined through a rigorous series of preliminary empirical tests to balance computational efficiency and prevent overfitting. The selection of 64 hidden units provides sufficient capacity to learn complex temporal patterns from the 27 technical features. The $softsign$ activation function was chosen for the hidden layers because it helps reduce vanishing gradients while providing a smoother clipping effect than $tanh$. To ensure generalization, a Dropout rate of 0.25 was introduced. Additionally, an early stopping mechanism with a patience of 15 epochs was implemented as a regularization strategy, halting the training process if the validation loss fails to improve, thereby avoiding overfitting on noisy financial data.

Hybrid ARIMA-LSTM

Essentially, this study adopts the hybrid ARIMA-LSTM framework from [10]. However, one key difference from that study lies in the use of stock data: while [10] utilized stock data from the maritime sector, this study incorporates a slight modification by adding a dual-input strategy. In addition to using linear residuals, the LSTM model in this study is combined with 27 technical indicators to capture the more complex volatility dynamics in the Indonesian capital market.

In the first stage, the ARIMA model is used to extract linear patterns and the main trend from stock prices. The parameters (p, d, q) are determined automatically based on the lowest AIC value. The output of this stage consists of residuals. The residuals (e_t) represent the difference between the actual values (yt) and the ARIMA-predicted values ($\widehat{L_t}$), as shown in equation (7).

$$e_t = yt - \widehat{L_t} \quad (7)$$

Residual values often still exhibit nonlinear patterns; to verify this, a BDS test is performed to validate the nonlinear structure. If a nonlinear structure is present, it is modeled using an LSTM. Thus, this model employs a dual-input strategy, where the LSTM

Muhammad Iszul Wilsa: *Corresponding Author

Journal of Artificial Intelligence and Technology Information (JAITI)
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, June 2026
https://doi.org/10.58602/jaiti.v4i2.260
Page 203-216

model not only accepts the residuals as a single input but also accepts 27 original technical features (X_t). This allows the LSTM to learn the correlation between market indicators and the prediction errors generated by the ARIMA model. The nonlinear relationship in the residuals is expressed in equation (8).

$$\hat{e}_t = f_{LSTM}(e_{t-1}, e_{t-2}, \dots, e_{t-n}, X_t) \quad (8)$$

The final step is to combine the predictions from both models to obtain a comprehensive price estimate ($\hat{y}_t$), as shown in equation (9). By combining these two components, the hybrid model overcomes the ARIMA model's weakness in capturing volatility and helps mitigate the tendency toward overfitting in the LSTM model.

$$\hat{y}_t = \hat{L}_t + \hat{e}_t \quad (9)$$

Evaluation

This study uses two evaluation metrics: Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). MAE measures the absolute difference between the actual value and the predicted value, as shown in equation (10). MAPE, on the other hand, measures the average percentage error between the predicted value and the actual value, as described in equation (11) [27].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|^2 \quad (10)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times 100 \quad (11)$$

To evaluate the exact error reduction achieved by the proposed ARIMA-LSTM hybrid model compared to the traditional linear baseline model, the Relative Improvement (RI) metric is introduced. Calculated based on the Mean Absolute Percentage Error (MAPE), this relative improvement measures the percentage reduction in error and is formulated as follows equation (12)[28]:

$$Relative\ Improvement\ (\%) = \frac{MAPE_{Baseline} - MAPE_{Hybrid}}{MAPE_{Baseline}} \times 100 \quad (12)$$

where $MAPE_{Baseline}$ represents the error metric generated by the traditional ARIMA model, and $MAPE_{Hybrid}$ represents the error generated by the proposed ARIMA-LSTM hybrid architecture. A positive RI percentage indicates success in reducing error, confirming that the integration of the two-input LSTM framework improves forecasting accuracy.

Furthermore, relying solely on descriptive error metrics such as MAE and MAPE is insufficient to prove that the proposed model's superiority is not due to random variance or market noise. Therefore, the Wilcoxon Signed-Rank Test is employed as a rigorous non-parametric statistical hypothesis test to determine whether the differences in predictive performance are statistically significant [29]. This test evaluates the absolute prediction errors generated by both models on a previously unseen test dataset, on a daily basis. Equation (13) is the difference between the absolute errors of the baseline model and the hybrid model at time t.

$$d_t = |e_t ARIMA| - |e_t Hybrid| \quad (13)$$

The null hypothesis (H_0) is defined as there being no significant difference between the prediction errors of the basic ARIMA model and the ARIMA-LSTM hybrid model (the median d_t is equal to 0), while the alternative hypothesis (H_1) is defined as the prediction error of the hybrid ARIMA-LSTM model being significantly lower than that of the basic ARIMA model (median d_t greater than 0).

The Wilcoxon test statistic (W) is calculated by sorting the absolute differences d_t from smallest to largest, assigning a sign based on whether d_t is positive or negative, and summing these ranks. A calculated $p - value$ lower than the standard significance threshold (α = 0.05) rejects the null hypothesis, which mathematically confirms that the proposed hybrid architecture provides a statistically significant improvement in commodity stock price forecasting[29].

Muhammad Iszul Wilsa: *Corresponding Author

Journal of Artificial Intelligence and Technology Information (JAITI)
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, June 2026
https://doi.org/10.58602/jaiti.v4i2.260
Page 203-216

3. RESULT AND DISCUSSIONS

BDS Validation Test

Before performing hybrid modeling, the ARIMA residuals were tested to confirm the presence of nonlinear patterns or components. The testing utilized the BDS (Brock-Dechert-Scheinkman) test across various dimensions.

Table 2. BDS Test Results for ARIMA Residuals

Ticker	Dimension 2 (p-value)	Dimension 3 (p-value)	Notes
ADRO	0.000	0.000	Non-linear detected
PTBA	0.000	0.000	Non-linear detected
MEDC	0.000	0.000	Non-linear detected
ANTM	0.000	0.000	Non-linear detected
MDKA	0.000	0.000	Non-linear detected
AALI	0.000	0.000	Non-linear detected

Based on Table 2, all p-values are less than 0.05, indicating that the null hypothesis that the data are identically distributed and independent is rejected. This confirms that the linear ARIMA model is unable to capture all the information in the data. Therefore, a hybrid LSTM model is required to incorporate a nonlinear component to account for these residual errors.

Performance Comparison of the 1-Step-Ahead Scenario Model

In the scenario of predicting the closing price of a stock for the next day, the ARIMA-LSTM hybrid model was compared with baseline models such as ARIMA, random forest, and LSTM. The average evaluation results on the test data using the MAE and MAPE metrics are shown in Table 3.

Table 3. 1-Step Ahead Test Results

Ticker	ARIMA		RF		LSTM		ARIMA-LSTM	
	MAE	MAPE%	MAE	MAPE%	MAE	MAPE%	MAE	MAPE%
ADRO	31.24	2.4	102.95	6.22	490.71	36.66	29.56	2.27
PTBA	43.91	2.06	71.11	3.22	378.73	17.2	43.4	2.03
MEDC	36.3	3.25	140.49	11.32	246.9	21.52	28.92	2.58
ANTM	32.87	2.2	48.35	3.49	53.72	3.67	24.32	1.63
MDKA	83.59	2.99	99.11	3.51	173.39	5.75	79.97	2.87
AALI	75.93	1.24	98.53	1.65	291.79	4.87	75.48	1.23
AVG	50.64	2.36	93.42	4.90	272.54	14.95	46.94	2.10

In this scenario, the ARIMA-LSTM hybrid model performed best among the other models, with an average MAPE of 2.10%. Although the difference from the ARIMA model was only slight approximately 2.36% the ARIMA-LSTM hybrid model successfully handled the LSTM residuals, thereby improving predictions in the 1-step-ahead scenario. A visualization of the predicted results compared with the actual data is shown in Fig. 2.

Muhammad Iszul Wilsa: *Corresponding Author

Journal of Artificial Intelligence and Technology Information (JAITI)
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, June 2026
https://doi.org/10.58602/jaiti.v4i2.260
Page 203-216

Table 4. Statistical Significance and Relative Improvement Analysis for 1-Step-Ahead Scenario

Ticker	ARIMA MAPE%	HYBRID MAPE%	Relative Improvement (%)	Wilcoxon Statistic	p-value	Status
ADRO	2.4	2.27	5.2709	49714.0	0.000878	Significant
PTBA	2.06	2.03	1.5696	58261.0	0.547458	Not Significant
MEDC	3.25	2.58	20.6887	33895.0	0.000000	Significant
ANTM	2.2	1.63	25.9827	29552.0	0.000000	Significant
MDKA	2.99	2.87	3.8232	48643.0	0.047571	Significant
AALI	1.24	1.23	0.8631	58163.0	0.478361	Not Significant

To rigorously verify the significance of these performance differences as suggested, a Wilcoxon signed-rank test was conducted alongside calculations of relative returns for each stock code, as shown in Table 4. The hybrid model achieved positive relative increases across all six assets, led by ANTM (25.98%), MEDC (20.69%), ADRO (5.27%), and MDKA (3.82%), followed by small increases in PTBA (1.57%) and AALI (0.86%). The Wilcoxon test yielded statistically significant p-values for ANTM (p = 0.00000), MEDC (p = 0.00000), ADRO (p = 0.00088), and MDKA (p = 0.04757), all of which fell below the significance threshold (p = 0.05). Conversely, for PTBA (p = 0.54746) and AALI (p = 0.47836), the performance improvements were not statistically significant, indicating that the baseline ARIMA model remains highly competitive for these specific assets due to their behavior being dominated by linearity during this period. This statistical validation confirms that the hybrid model's ability to minimize short-term errors is very strong and statistically significant for four of the six commodity stocks.

Muhammad Iszul Wilsa: *Corresponding Author

Journal of Artificial Intelligence and Technology Information (JAITI)
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, June 2026
https://doi.org/10.58602/jaiti.v4i2.260
Page 203-216

Figure 2. Visualization of prediction results compared to actual data for the 1-Step Ahead scenario

Comparison of the Performance of the 5-Step-Ahead Scenario Model

The second scenario involves testing the model's robustness over a longer time frame of 5 trading days. Theoretically, accuracy will decline over time due to the accumulation of errors in the model. The average results of the evaluation metrics are shown in Table 5.

Table 5. 5-Step Ahead Test Results

Ticker	ARIMA		RF		LSTM		ARIMA-LSTM	
	MAE	**MAPE%**	**MAE**	**MAPE%**	**MAE**	**MAPE%**	**MAE**	**MAPE%**
ADRO	51.55	3.98	141.93	9.49	421.46	31.11	51.6	3.97
PTBA	74.51	3.53	116.1	5.34	471.6	21.47	73.93	3.48
MEDC	64.15	5.74	173.72	14.28	249.5	21.66	61.28	5.48
ANTM	59.4	3.98	115.52	8.13	86.94	6.1	54.12	3.62
MDKA	159.16	5.77	245.07	8.81	452.38	18.02	151.23	5.45
AALI	123.89	2.05	283.7	4.77	435.88	7.4	120.98	2
AVG	88.78	4.18	179.34	8.47	352.96	17.63	85.52	4.00

The results in Table 5 show that all models experienced an increase in error values. However, the ARIMA-LSTM hybrid model maintained its position as the model with the lowest error value, with a MAPE of 4.00%, outperforming the ARIMA model, which had a MAPE of 4.18%. This also indicates that the nonlinear patterns learned by the LSTM contribute positively to maintaining the stability of the forecast results when faced with weekly price volatility. A visualization of the predicted results compared with the actual data is shown in Fig. 3.

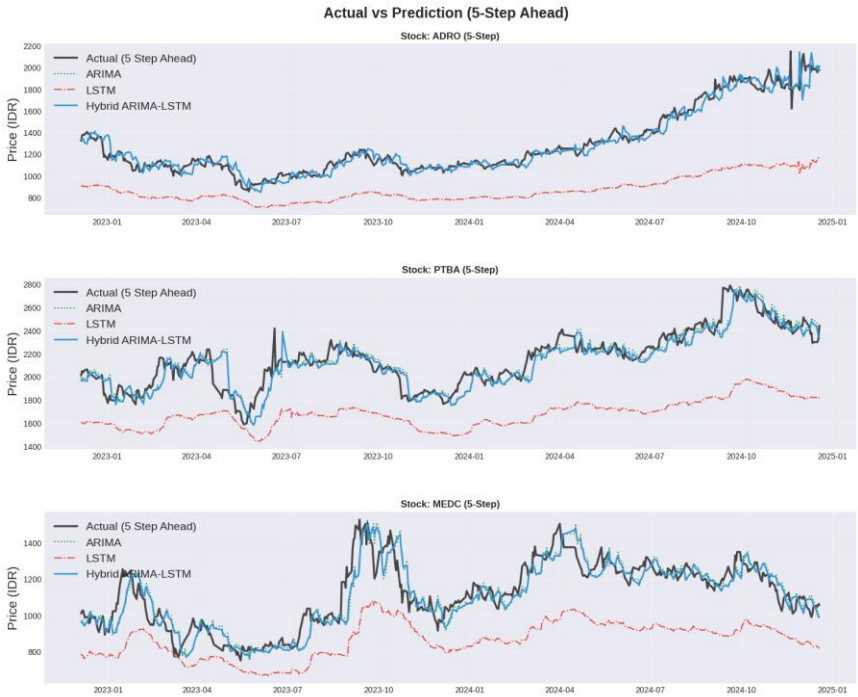
Muhammad Iszul Wilsa: *Corresponding Author

Journal of Artificial Intelligence and Technology Information (JAITI)
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, June 2026
https://doi.org/10.58602/jaiti.v4i2.260
Page 203-216

Table 6. Statistical Significance and Relative Improvement Analysis for 5-Step-Ahead Scenario

Ticker	ARIMA MAPE%	HYBRID MAPE%	Relative Improvement (%)	Wilcoxon Statistic	p-value	Status
ADRO	3.98	3.97	0.2151	59134.0	0.746552	Not Significant
PTBA	3.53	3.48	1.3249	58559.0	0.612473	Not Significant
MEDC	5.74	5.48	4.5688	49462.0	0.000511	Significant
ANTM	3.98	3.62	8.9584	42444.0	0.000000	Significant
MDKA	5.77	5.45	5.4893	44038.0	0.000365	Significant
AALI	2.05	2.00	2.4808	47974.0	0.000079	Significant

This evaluation was further supported by testing its statistical significance using the Wilcoxon signed-rank test, as shown in Table 6. At the individual asset level, the non-linear error correction generated by the LSTM proved to be statistically significant for four assets: ANTM (Relative Increase = 8.96%, p = 0.00000), MDKA (Relative Increase = 5.49%, p = 0.00037), AALI (Relative Increase = 2.48%, p = 0.00008), and MEDC (Relative Increase = 4.57%, p = 0.00051). However, for ADRO (Relative Improvement = 0.22%, p = 0.74655) and PTBA (Relative Improvement = 1.32%, p = 0.61247), these performance improvements were not statistically significant (p > 0.05), despite a small decrease in MAPE. These findings suggest that as the forecasting horizon is extended to a weekly framework, the accumulation of errors in the deep learning framework limits its ability to provide a statistically significant advantage over the conservative linear baseline for certain commodity assets, while remaining highly effective for others.



Muhammad Iszul Wilsa: *Corresponding Author

Journal of Artificial Intelligence and Technology Information (JAITI)
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, June 2026
https://doi.org/10.58602/jaiti.v4i2.260
Page 203-216

Figure 3. Visualization of prediction results compared to actual data for the 5-Step Ahead scenario

Discussion

One of the main objectives of this study is to test the robustness of the ARIMA-LSTM hybrid model against the volatility of various stocks in Indonesia's commodities sector. Although the improvement in accuracy of the hybrid model compared to baseline models such as ARIMA appears relatively small, the hybrid model demonstrates consistent superiority across nearly all test datasets. A significant advantage is clearly evident when comparing the hybrid model with a single deep learning model (LSTM). The hybrid model is able to overcome the fundamental weakness of deep learning related to high sensitivity to hyperparameter selection. In this test, even though uniform parameters were applied to all datasets, the hybrid model still demonstrated high performance stability. This performance stability, along with its limits, has now been rigorously verified through Wilcoxon signed-rank test results. In the 1-step-ahead scenario, the hybrid model's precision advantage has been statistically verified ($p < 0.05$) for 4 out of 6 stocks, proving that the integration of 27 technical features with ARIMA residuals provides a real predictive advantage, not random variation. Interestingly, when shifting to a 5-step-ahead scenario, statistical significance changes, becoming highly pronounced for assets such as ANTM, MDKA, MEDC, and AALI ($p < 0.001$), while remaining statistically indistinguishable from ARIMA for highly trend-following assets like ADRO and PTBA. This clear significance threshold directly addresses the issue of hyperparameter dependence, indicating that the hybrid framework retains its general-purpose capabilities without requiring specific tuning for each stock, even though its performance is significantly and dynamically influenced by the forecasting horizon and the volatility conditions of specific assets.

This stability is supported by the data in Table 3, where, in the 1-step-ahead scenario, the single LSTM model produced an average MAPE of 14.95%, while the hybrid model successfully reduced it to 2.10%. A similar pattern is observed in Table 4 for the 5-step-ahead scenario, where the LSTM's MAPE of 17.63% drops drastically to 4.00% in the hybrid model. In terms of stability, the ARIMA and hybrid models demonstrate excellent generalization capabilities with minimal gaps between training and testing values. In the

Muhammad Iszul Wilsa: *Corresponding Author

Journal of Artificial Intelligence and Technology Information (JAITI)
Website : https://ejournal.techcart-press.com/index.php/jaiti
E-ISSN 2985-6396
P-ISSN 2985-5306
Volume 4, Number 2, June 2026
https://doi.org/10.58602/jaiti.v4i2.260
Page 203-216

1-step-ahead scenario, the hybrid model recorded an average training MAPE of 2.88% and a testing MAPE of 2.10%. This proves that the hybrid architecture successfully extracts fundamental features from stock price movements without being trapped by market noise. In contrast, the single LSTM model showed strong signs of overfitting. For example, for ADRO stock, the LSTM produced a training MAPE of 10.51%, but this spiked sharply to 36.66% during the testing phase. This surge in error confirms that the LSTM model tends to "memorize" fluctuations in the training data and fails to adapt when faced with new trends in the market.

The hybrid model's consistent outperformance of Random Forest and ARIMA confirms that the use of 27 technical indicators combined with ARIMA residuals provides a more comprehensive insight for the model. The Random Forest model was found to struggle to capture meaningful signals when faced with high volatility in Indonesia's commodities sector, which tends to exhibit more dynamic trend patterns than other sectors.

4. CONCLUSION

This study concludes that the ARIMA-LSTM model demonstrates robust performance in predicting stock prices in Indonesia's commodities sector compared to ARIMA, Random Forest, and LSTM. This superiority is evidenced by an average MAPE of 2.10% in the 1-step-ahead scenario and 4.00% in the 5-step-ahead scenario. The Wilcoxon signed-rank test mathematically solidifies this framework, proving that the hybrid model yields statistically significant improvements for ANTM, MEDC, ADRO, and MDKA in short-term daily horizons, and retains its significant edge for MEDC, ANTM, MDKA, and AALI as the horizon extends to a weekly frame. This success lies in the model's ability to accurately decompose data, where linear trends are handled by ARIMA while non-linear fluctuations are managed by LSTM, supported by 27 technical indicators. This approach has proven effective in addressing issues such as extreme overfitting a common challenge for deep learning models while delivering stable performance in adapting to the characteristics of Indonesia's capital market.

For future research, it is recommended to use optimization algorithms such as Particle Swarm Optimization (PSO) or Genetic Algorithms (GA) to automatically optimize LSTM parameters based on the data. Additionally, other variables such as real-time exchange rates, global commodity prices, and NLP-based sentiment analysis could be incorporated to enhance the predictive results.

5. REFERENCES

[1] H. Hamzah and S. Winardi, "Effective Stock Prediction Model Using MACD Method," *International Journal of Informatics and Computation*, vol. 4, no. 2, p. 1, Dec. 2022, doi: 10.35842/ijicom.v4i2.51.

[2] F. Rihaadatul Aisyi and O. Rohaeni, "Perbandingan Simple Linear Regressiondan Double Exponential Smoothing dalam Memprediksi Harga Saham," *Jurnal Riset Matematika*, pp. 167–176, Dec. 2023, doi: 10.29313/jrm.v3i2.2836.

[3] S. Latif, N. Javaid, F. Aslam, A. Aldegheishem, N. Alrajeh, and S. H. Bouk, "Enhanced prediction of stock markets using a novel deep learning model PLSTM-TAL in urbanized smart cities," *Heliyon*, vol. 10, no. 6, p. e27747, Mar. 2024, doi: 10.1016/j.heliyon.2024.e27747.

[4] Z. Shi, Z. Wu, S. Shi, C. Mao, Y. Wang, and L. Zhao, "High-Frequency Forecasting of Stock Volatility Based on Model Fusion and a Feature Reconstruction Neural Network," *Electronics (Basel).*, vol. 11, no. 23, p. 4057, Dec. 2022, doi: 10.3390/electronics11234057.

[5] E. Abnaina and F. Swandari, "Pengaruh Variabel Makroekonomi Dan Variabel Global Index Tehadap Indeks Harga Saham Gabungan (IHSG)," *Jurnal Manajemen dan*

Muhammad Iszul Wilsa: *Corresponding Author

Bisnis Performa, vol. 19, no. 01, pp. 83–92, Mar. 2022, doi: 10.29313/performa.v19i01.9724.

[6] A. Firnanda and B. Budiasih, "Performa Saham Sektor Energi selama Periode Covid-19 Delta dan Omicron di Indonesia," *Seminar Nasional Official Statistics*, vol. 2023, no. 1, pp. 331–342, Oct. 2023, doi: 10.34123/semnasoffstat.v2023i1.1627.

[7] N. P. G. Darmayanti and I. K. Yadnyana, "Does firm size can moderate the effect of dividend pay-out ratio and leverage on stock volatility?," *International journal of business, economics & management*, vol. 6, no. 2, pp. 148–154, Jun. 2023, doi: 10.21744/ijbem.v6n2.2138.

[8] R. K. Si, S. K. Padhan, and Dr. B. Bishi, "Application of Box – Jenkins ARIMA (p, d, q) Model for Stock Price Forecasting and Detect Trend of S&P BSE Stock Index: An Evidence from Bombay Stock Exchange," *Scholars Journal of Physics, Mathematics and Statistics*, vol. 7, no. 7, pp. 110–125, Jul. 2020, doi: 10.36347/sjpms.2020.v07i07.006.

[9] John Mark Limel Papag, Mary Grace Puma, and Gabriela Nichole Lim, "Forecasting the Daily Stock Prices of the Philippine Stock Exchange Index During the Opening and Closing of the Market," *International Journal For Multidisciplinary Research*, vol. 5, no. 3, May 2023, doi: 10.36948/ijfmr.2023.v05i03.3246.

[10] K. Gerakoudi, D. Georgoulas, and P. J. Stavroulakis, "A novel hybrid ARIMA-LSTM model for maritime shipping stock forecasting: comparative evidence against statistical and machine learning benchmarks," *Maritime Business Review*, vol. 11, no. 1, pp. 2–20, Mar. 2026, doi: 10.1108/MABR-04-2024-0035.

[11] N. Li, "Literature Review: Machine Learning in Stock Predictions," *Highlights in Business, Economics and Management*, vol. 24, pp. 853–859, Jan. 2024, doi: 10.54097/81x6z947.

[12] L. Farhatuaini, H. P. Kurniawan, and I. Muslihah, "Hybrid Model of Artificial Neural Networks and Flower Pollination Algorithm for Stock Price Prediction," *Jurnal Sistem Cerdas*, vol. 7, no. 3, pp. 356–365, Dec. 2024, doi: 10.37396/jsc.v7i3.433.

[13] G. Sonkavde, D. S. Dharrao, A. M. Bongale, S. T. Deokate, D. Doreswamy, and S. K. Bhat, "Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications," *International Journal of Financial Studies*, vol. 11, no. 3, p. 94, Jul. 2023, doi: 10.3390/ijfs11030094.

[14] J. Zhang, K. Cai, and J. Wen, "A survey of deep learning applications in cryptocurrency," *iScience*, vol. 27, no. 1, p. 108509, Jan. 2024, doi: 10.1016/j.isci.2023.108509.

[15] D. Xiao and J. Su, "Research on Stock Price Time Series Prediction Based on Deep Learning and Autoregressive Integrated Moving Average," *Sci. Program.*, vol. 2022, pp. 1–12, Mar. 2022, doi: 10.1155/2022/4758698.

[16] S. A. Alokley, S. Araichi, and G. Alomair, "Exploring the Relationship and Predictive Accuracy for the Tadawul All Share Index, Oil Prices, and Bitcoin Using Copulas and Machine Learning," *Energies (Basel).*, vol. 17, no. 13, p. 3241, Jul. 2024, doi: 10.3390/en17133241.

[17] K. Li, "BILSTM Model for Machine Learning Stock Prediction and Portfolio Construction in the Context of Covid-19 Pandemic," *Finance & Economics*, vol. 1, no. 8, Aug. 2024, doi: 10.61173/q8zeqe61.

[18] A. Kulaglic and B. Berk Ustundag, "Stock Price Prediction Using Predictive Error Compensation Wavelet Neural Networks," *Computers, Materials & Continua*, vol. 68, no. 3, pp. 3577–3593, 2021, doi: 10.32604/cmc.2021.014768.

[19] P. Lv, Q. Wu, J. Xu, and Y. Shu, "Stock Index Prediction Based on Time Series Decomposition and Hybrid Model," *Entropy*, vol. 24, no. 2, p. 146, Jan. 2022, doi: 10.3390/e24020146.

Muhammad Iszul Wilsa: *Corresponding Author

[20] Z. D. Akşehir and E. Kılıç, "2LE-BO-DeepTrade: an integrated deep learning framework for stock price prediction," *PeerJ Comput. Sci.*, vol. 11, p. e3107, Aug. 2025, doi: 10.7717/peerj-cs.3107.
[21] S. Han, "Financial Time Series Forecasting: A Hybrid Approach Combining AR-GARCH and Machine Learning Models," *Transactions on Computer Science and Intelligent Systems Research*, vol. 10, pp. 72–77, Aug. 2025, doi: 10.62051/pg9aec47.
[22] M. R. Kabir, D. Bhadra, M. Ridoy, and M. Milanova, "LSTM–Transformer-Based Robust Hybrid Deep Learning Model for Financial Time Series Forecasting," *Sci*, vol. 7, no. 1, p. 7, Jan. 2025, doi: 10.3390/sci7010007.
[23] D. A. HSIEH, "Chaos and Nonlinear Dynamics: Application to Financial Markets," *J. Finance*, vol. 46, no. 5, pp. 1839–1877, Dec. 1991, doi: 10.1111/j.1540-6261.1991.tb04646.x.
[24] J. Liu, "Navigating the Financial Landscape: The Power and Limitations of the ARIMA Model," *Highlights in Science, Engineering and Technology*, vol. 88, pp. 747–752, Mar. 2024, doi: 10.54097/9zf6kd91.
[25] G. Kumar, U. P. Singh, and S. Jain, "An adaptive particle swarm optimization-based hybrid long short-term memory model for stock price time series forecasting," *Soft comput.*, vol. 26, no. 22, pp. 12115–12135, Nov. 2022, doi: 10.1007/s00500-022-07451-8.
[26] K. Mahboob, M. H. Shahbaz, F. Ali1, and R. Qamar, "Predicting the Karachi Stock Price index with an Enhanced multi-layered Sequential Stacked Long-Short-Term Memory Model," *VFAST Transactions on Software Engineering*, vol. 11, no. 2, pp. 249–255, Jun. 2023, doi: 10.21015/vtse.v11i2.1571.
[27] A. M. Priyatno, L. S. Tanjung, W. F. Ramadhan, P. Cholidhazia, P. Z. Jati, and F. I. Firmananda, "Comparison Random Forest Regression and Linear Regression For Forecasting BBCA Stock Price," *Jurnal Teknik Industri Terintegrasi*, vol. 6, no. 3, pp. 718–732, Jul. 2023, doi: 10.31004/jutin.v6i3.16933.
[28] V. H. Ho, H. Morita, F. Bachofer, and T. H. Ho, "Random forest regression kriging modeling for soil organic carbon density estimation using multi-source environmental data in central Vietnamese forests," *Model. Earth Syst. Environ.*, vol. 10, no. 6, pp. 7137–7158, Dec. 2024, doi: 10.1007/s40808-024-02158-1.
[29] R. F. Woolson, "Wilcoxon Signed-Rank Test," in *Encyclopedia of Biostatistics*, Wiley, 2005. doi: 10.1002/0470011815.b2a15177.

Muhammad Iszul Wilsa: *Corresponding Author