Comparison of ARIMA, LSTM, and Ensemble Averaging Models for Short-Term and Long-Term Forecasting of Non-Stationary Time Series Data

Windy Ayu Pratiwi[1]*, I Made Sumertajaya[1], Khairil Anwar Notodiputro[1]
[1]School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia
*Corresponding author: windyayu@apps.ipb.ac.id

Received: 19 February 2025
Revised: 18 June 2025
Accepted: 30 October 2025

ABSTRACT – This study aims to forecast the highest weekly selling rate of the Indonesian Rupiah (IDR) against the US Dollar (USD) and identify the most accurate model among ARIMA, LSTM, and Ensemble Averaging. The evaluation results indicate that ARIMA achieves an accuracy of 99.37%, demonstrating strong performance in short-term forecasting, while LSTM achieves an accuracy of 99.99%, excelling in capturing complex and dynamic patterns for long-term predictions. The Ensemble Averaging approach achieves an accuracy of 99.87%, proving to be the optimal solution by combining ARIMA's stability with LSTM's adaptability, resulting in relatively accurate and stable predictions. Although the Ensemble Averaging model has higher RMSE and MSE values compared to the individual models (ARIMA and LSTM), this approach remains quite effective in forecasting both short-term and long-term time series data. This shows that, despite larger prediction errors, Ensemble Averaging provides more stable and accurate results in the long term. The findings highlight that the ensemble approach is more effective than individual models, as it balances accuracy and prediction stability across various forecasting scenarios. This method serves as a reliable tool for addressing market volatility and contributes significantly to the advancement of more adaptive and accurate financial and economic forecasting techniques.

Keywords – Ensemble Averaging, Short-term Forecasting, ARIMA, LSTM, Long-term Forecasting,

I. INTRODUCTION

Time series data refers to a series of observations collected or measured at consecutive time intervals. These data are gathered at predetermined intervals, such as daily, weekly, monthly, or annually, and often exhibit patterns such as upward or downward trends, seasonal variations, or cyclical behaviors. Time series data frequently display non-stationary patterns, which can pose challenges in analysis, including forecasting. which is influenced by dynamic economic, political, and global factors. Forecasting of time series data can be performed for both short-term and long-term horizons. Short-term forecasting, which typically covers a few upcoming periods such as daily, weekly, or monthly intervals, faces challenges in capturing rapid fluctuations and seasonal patterns. On the other hand, long-term forecasting involves predictions for more distant periods, such as years or decades, and is often used for strategic planning, including investment strategies or economic policy development. Long-term forecasting faces the difficulty of maintaining accuracy despite the presence of complex trends and high levels of uncertainty [1].

A case study on time series data that demonstrates these challenges is the exchange rate of the Indonesian Rupiah (IDR) against the United States Dollar (USD). This data shows highly dynamic fluctuations, influenced by both domestic and global factors such as inflation, interest rates, trade balance, geopolitical tensions, and external shocks. The IDR/USD exchange rate is particularly sensitive to sudden economic policy shifts or global market volatility, making it a complex and important variable to forecast. In this context, choosing the right forecasting model is crucial not only for financial institutions and investors, but also for policy makers aiming to maintain currency stability and design effective macroeconomic strategies. Both domestic and global factors significantly contribute to the non-linear behavior of the data. These complex fluctuations make forecasting the IDR/USD exchange rate a challenging task, especially in identifying and capturing rapid and non-linear movements. Short-term forecasting of the IDR/USD exchange rate, which involves predictions within days, weeks, or months, faces challenges in capturing quick fluctuations that occur in a short time span. On the other hand, long-term forecasting, which involves predictions for longer periods, such as several years, tends to be influenced by macroeconomic trends and broader factors, which are often more difficult to predict due to uncertainty and the rapid changes in global conditions.

To address these challenges, more advanced forecasting models are needed, such as ARIMA and LSTM models. ARIMA, developed by George Box and Gwilym Jenkins in 1976 [2], is one of the most popular time series analysis models frequently used for forecasting purposes. This model is highly regarded for its simplicity, ease of interpretation, and effectiveness in handling linear patterns. However, ARIMA encounters challenges when working with data exhibiting non-linear behavior. A 2023 study on ARIMA identified it as the optimal model for forecasting global gold prices [3]. Considering ARIMA's strengths and limitations, it is necessary to explore alternative models for more effective time series forecasting. In recent years, researchers have developed various models to enhance accuracy in time series analysis. Among these, LSTM has gained considerable attention for handling time series data due to its reliability in addressing non-linear patterns [4]. However, LSTM models are prone to overfitting and are often considered "black-box" models, making their processes less transparent and more difficult to interpret [5], [6].

A study comparing Multilayer Perceptron (MLP) and LSTM models for forecasting rice prices found that LSTM produced more accurate predictions, as evidenced by lower RMSE values and closer alignment with actual price ranges [7].

To leverage the strengths of ARIMA and LSTM while mitigating their respective weaknesses, combining these models through ensemble approaches is a viable solution. Ensemble forecasting techniques are renowned for their accuracy and reliability in predictions [8]. One promising ensemble method for time series forecasting is ensemble averaging. Research on ensemble averaging has demonstrated its effectiveness, achieving high accuracy in forecasting rainfall data in 2022 [9]. This ensemble model is used to improve prediction accuracy by combining multiple forecasting models to reduce error rates and uncertainty in exchange rate fluctuations. In the context of the selling rate of the Indonesian Rupiah (IDR) against the US Dollar (USD), the ensemble approach has proven effective in capturing complex patterns influenced by global and domestic economic factors. Forecasting the highest exchange rate of the IDR against the USD within a one-week period offers benefits for market participants, such as investors, exporters, and importers, in planning their financial strategies more effectively. Accurate predictive information can assist in determining the optimal timing for foreign exchange transactions to minimize the risk of losses due to market volatility. Furthermore, the government and policymakers can leverage these forecasting results to formulate more adaptive economic policies in response to global market dynamics. Thus, the application of ensemble methods in exchange rate forecasting has the potential to become a strategic tool for financial risk management and data-driven decision-making.

This research aims to assess and compare the effectiveness of ARIMA, LSTM, and Ensemble Averaging models in forecasting non-stationary time series data. Additionally, it focuses on identifying the model that delivers optimal performance for short-term and long-term forecasts by utilizing appropriate metrics to evaluate forecasting accuracy. This research also develops an ensemble averaging model by combining ARIMA and LSTM to model the exchange rate of the Indonesian Rupiah (IDR) against the US Dollar (USD). Such an ensemble model has not been explored in previous literature, highlighting the novelty and potential contribution of this research to the field of time series forecasting. The findings are expected to provide valuable insights for selecting the most suitable forecasting model for various applications.

II. LITERATURE REVIEW

A. *Non-Linear Time Series*

Data characterized by complex relationships between observed variables, which cannot be explained by simple linear models, is referred to as non-linear time series data. Non-linearity in time series often arises from natural or artificial processes exhibiting chaotic or irregular dynamic behavior. Non-linear models offer a distinct advantage over linear models due to their capability to capture dynamic behaviors in data, such as limit cycles and bifurcations [10]. The characteristics of non-linear time series include instability, irregular cycles, and variance that changes over time (heteroskedasticity). Moreover, non-linear time series exhibit dependency on past data [11]. These features make non-linear models particularly valuable for accurately analyzing and predicting patterns in complex datasets where linear models fall short.

B. *ARIMA*

ARIMA consists of two primary components, namely Autoregressive (AR) and Moving Average (MA), which are frequently utilized in data analysis [2].

$$\Delta Y_t = \varphi_1 \Delta Y_{t-1} + \varphi_2 \Delta Y_{t-2} + \varphi_p \Delta y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_p e_{t-p} \tag{1}$$

The terms used in this explanation are defined as follows: t represents the time index, Y_t denotes the series value at time t, while Y_{t-1}, Y_{t-2} indicate the values of the series at previous time points. Similarly, e_{t-1}, e_{t-2} refer to the residuals from previous time steps. Finally, $\theta_1, \theta_p, \varphi_1, \varphi_p$ correspond to the coefficients of the model.

C. LSTM

LSTM, as a component of Deep Learning, represents an enhanced version of the RNN model and was specifically created to overcome the shortcomings of RNN in retaining information over extended periods, which often complicates accurate predictions. A defining feature of LSTM is its capability to efficiently preserve essential information while filtering out irrelevant data. This improvement allows LSTM to process and store information more effectively, making it particularly suitable for forecasting tasks involving formats such as text, video, and time series data [12], [13]. Below is the architecture of LSTM:

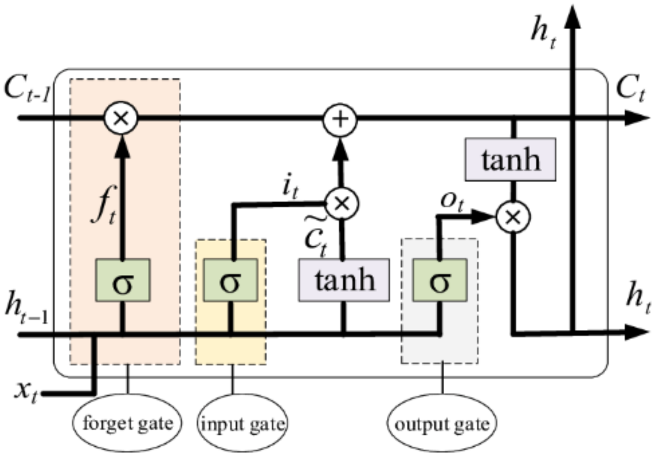

Figure 1 LSTM Architecture

In Figure 1, the addition operation is represented by ⊕, while the multiplication operation is represented by ⊗, and arrows indicate the direction of data flow. In this model, the memory cell (neuron) incorporates additional information by using the hidden state and cell state from the preceding time step. The process begins with the previous cell state (xs) and hidden state (hs−1) passing through the forget gate, utilizing weights (We) and bias (be) to filter out unnecessary information. Next, within the input gate, two operations are performed on (xs) and (hs−1) before the previous cell state (Cs−1) is transformed into the new cell state (Cs). In the input gate process, the input gate performs two operations. First, a sigmoid function (Is), combined with weights (Ws), and bias (bi), Identifies information that needs to be updated in the cell state. Second, a tan function (Cs), along with weights (Wb), and bias (bb), produces the output of this gate, which contains new information to be added to the cell state. Data then flows through the output gate, where weights (Wl), and bias (bl), are applied to compute the output, generating the updated cell state. Finally, the output value (hs) is calculated based on the modified cell state. The following equations describe the operation of LSTM.

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

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

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

$$C_t = \tanh(f_t \times C_{t-1} + i_t \times \acute{C}_t) \quad (5)$$

$$O_t = \sigma(W_o \times x_t + U_o \times h_{t-1} + b_o) \quad (6)$$

$$h_t = O_t \times \tanh(C_t) \quad (7)$$

D. Model Ensemble

An Ensemble Model is formed by combining the forecasting results from multiple single models. This approach is designed to address the accuracy limitations of individual forecasting models [8]. In this study, the single models used are ARIMA and LSTM. Once the forecasting results from ARIMA and LSTM are obtained, the next step is to combine these results to form the ensemble model. Two of the most commonly used ensemble approaches are averaging.

E. Model Ensemble averaging

The ensemble averaging model is a forecasting model that combines the results of individual models by averaging their outputs. The general formula for a simple average forecast is as follows [11]:

$$\hat{x}_i = (\hat{x}_i^p) =, i = 1,2,\dots,n \quad (8)$$

Where $\hat{x}_i$is the forecast result from the i-th ensemble model, and

$$\hat{x}_i = \frac{1}{N}\sum_{p=1}^{N} \hat{x}_i^p \quad (9)$$

The use of the averaging model has been demonstrated to effectively enhance the performance of individual models. [14].

III. METHODOLOGY

This research was conducted through several stages, starting with data collection sourced from the bi.go.id website. The data period covers weekly observations from January 2018 to December 2024, with a total of 365 data points. The variable analyzed is the highest selling rate of the Indonesian Rupiah (IDR) against the US Dollar (USD). The data exhibits a trend component and is non-stationary. This study employs an ensemble averaging model that combines the individual ARIMA and LSTM models.

A. Data Analysis Procedure

1) Explore the data using time series plots to analyze the overall pattern.
2) Divide the Data: The entire dataset (January 2018 – December 2024) will be divided into two parts: training data (January 2018 – December 2023) and the remaining portion (January 2024 – December 2024) for testing data.
3) ARIMA Model Analysis Procedure
   a) Stationarity Check: Evaluate the training data using ACF plots and the ADF test to determine if the data is stationary. Data with a consistent mean and variance over time is considered to exhibit a stationary pattern. [20].
   b) Non-stationary mean: Apply differencing.
   c) Non-stationary variance: Apply natural logarithm transformation to normalize variance [15], [16].
   d) Model Identification: Utilize ACF and PACF plots to identify provisional models and estimate the appropriate parameters for the model.
   e) Model Selection: Select the most optimal model based on the lowest AIC value.
   f) Diagnostic Testing: Validate the selected model for adequacy using residual analysis.
   g) Overfitting Test: Add parameters to check if overfitting improves the model.
   h) Final Model Selection: Compare tentative and overfitted models based on AIC and parameter significance.
   i) Forecasting: Use the best model to predict values (yt).
   j) Accuracy Metrics: Measure forecast accuracy using RMSE and MAPE.

4) LSTM Model Analysis Procedure
   a) Normalization: Normalize data using MinMaxScaler to scale values between 0 and 1.
   b) Supervised Learning Transformation: Convert normalized data into supervised learning format.
   c) Dimensional Reshaping: Reshape data into three dimensions [samples, time steps, features] using numpy.
   d) Hyperparameter Optimization: Optimize Dense Layer, neuron count, epochs, loss function (MSE), optimizers (SGD, RMSprop, ADAM), and activation functions (Sigmoid, Tanh).
   e) Grid Search with Cross-Validation: Perform GridSearchCV with k-fold cross-validation to find the best hyperparameters, evaluating model performance by averaging validation scores across folds [17].
   f) Final Model Training: Train the model with optimal hyperparameters.

5) Ensemble Averaging Model Analysis Procedure
   a) Forecasting with Single Models: Generate forecasts using ARIMA and LSTM models and calculate their RMSE.
   b) Weighted Averaging: Combine forecasts using weighted averages. Assign higher weights to models with smaller RMSE values:

$$\omega_{ARIMA} = \frac{\frac{1}{RMSE_{ARIMA}}}{\frac{1}{RMSE_{ARIMA}}+\frac{1}{RMSE_{LSTM}}} \tag{10}$$

$$\omega_{LSTM} = \frac{\frac{1}{RMSE_{LSTM}}}{\frac{1}{RMSE_{ARIMA}}+\frac{1}{RMSE_{LSTM}}} \tag{11}$$

$$\omega_{ARIMA} + \omega_{LSTM} = 1 \tag{12}$$

6) Ensemble Prediction: Calculate the ensemble forecast using:

$$\hat{y}_{ensemble} = \omega_{ARIMA} \times \hat{y}_{ARIMA} + \omega_{LSTM} \times \hat{y}_{LSTM} \tag{13}$$

7) Model Evaluation: Assess the ensemble model's performance using RMSE, MAPE, and MSE metrics. The following is the workflow of this study.

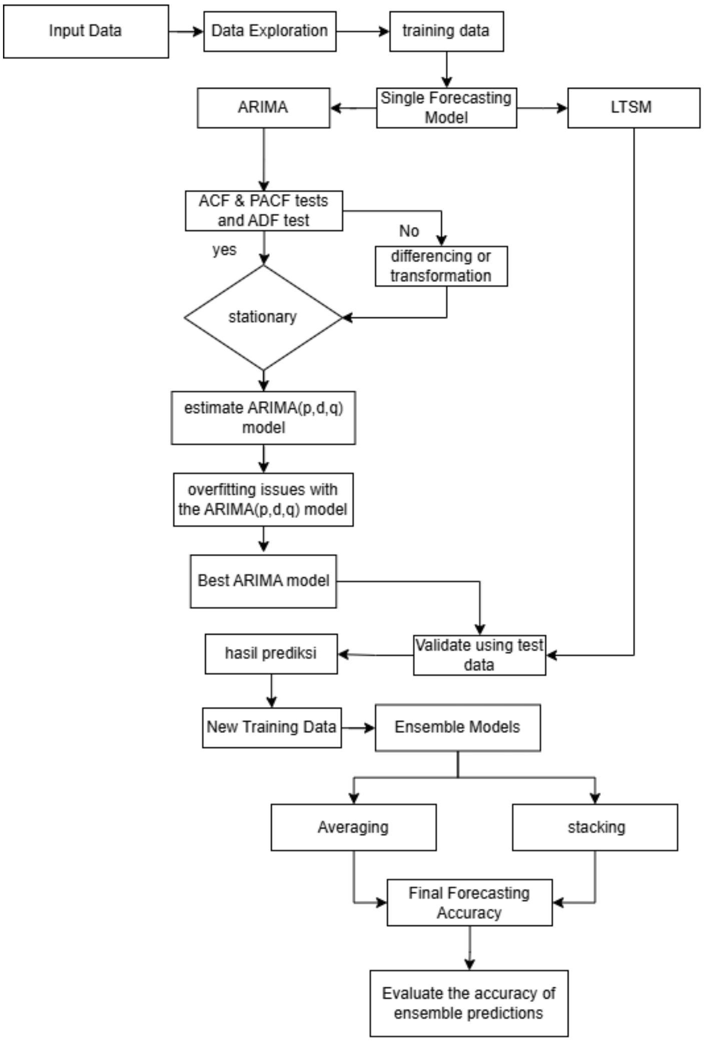

Figure 2 Workflow for Combining ARIMA and LSTM Models with Ensemble Techniques for Gold Price Forecasting

IV. RESULTS AND DISCUSSIONS

1) Data Exploration

This study utilizes the highest exchange rate data of the Indonesian Rupiah (IDR) against the US Dollar (USD), with the variable used being the selling rate. The data was collected from January 2018 to December 2024, totaling 365 observations. The data will be divided into training and testing datasets. The training data covers the period from January 2018 to February 2023, while the testing data spans from March 2023 to December 2024. Below is a graph illustrating the highest selling rate after dividing the data into training and testing datasets.

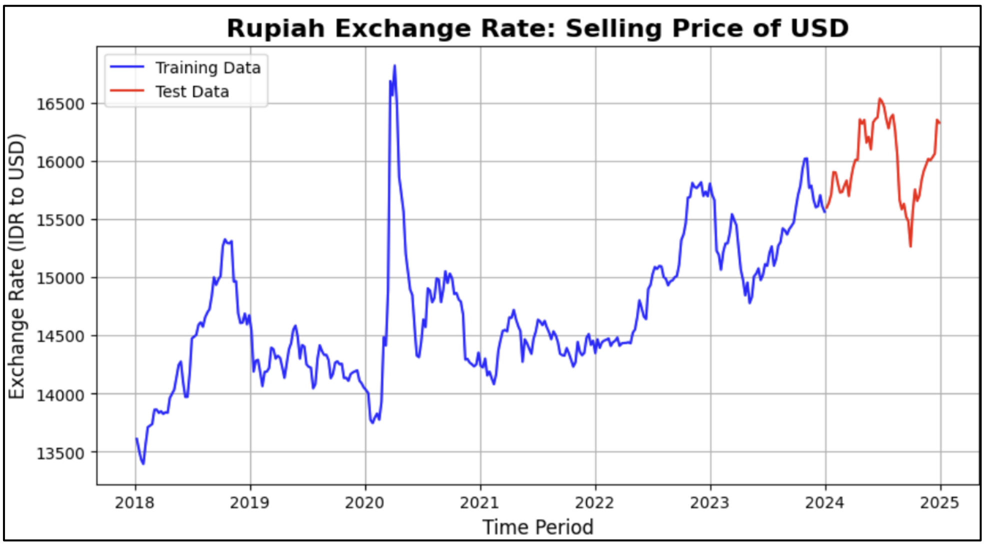

Figure 3 The highest exchange rate of the Indonesian Rupiah (IDR) against the US Dollar (USD)..

Based on Figure 3, an upward trend is observed, reflecting the depreciation of the Rupiah against the USD during this period. A significant spike occurred in early 2020, likely influenced by global economic factors such as the COVID-19 pandemic, which caused drastic changes in the exchange rate. Following this spike, the exchange rate exhibited fluctuating patterns with a long-term upward trend. There are indications that the data may not be stationary, as evidenced by changes in trends and variations in exchange rate fluctuations over the observed period. The training data (January 2018 - December 2023) shows movement patterns that tend to change over time, while the test data (January 2024 - December 2024) indicates a continuation of the trend with a higher potential for non-stationarity. To address the issue of non-stationarity and heteroscedasticity (unequal variance across time), differencing was applied. This method helps remove trends and stabilize the variance in the data, making it more suitable for time series modeling by ensuring more consistent fluctuations throughout the observation period.

These indications of non-stationarity suggest that the exchange rate of the Rupiah against the US Dollar may be influenced by various dynamic economic factors, such as inflation, monetary policies, and ever-changing global economic conditions. Therefore, selecting a forecasting model capable of handling potential non-stationarity in the data is crucial. Models such as ARIMA can capture differentiation components to address changing trends, while machine learning-based models like LSTM can recognize complex patterns in sequential data. The use of appropriate models is expected to enhance the accuracy of exchange rate predictions and provide better insights for stakeholders in developing more adaptive financial strategies and economic policies.

2) Data Stationarity

The weekly highest gold futures price data exhibits fluctuations, indicating a non-stationary pattern. Figure 3 reinforces this indication, showing a gradual decline in the ACF plot (tails off) and lag values that fall outside the standard error interval. To further verify the non-stationarity of the data, an ADF test was performed. The test results revealed that the data is non-stationary, as indicated by a p-value of 0.736 (0.736 > 0.05). This confirms that the mean is not constant over time. Figure 4 below presents the ACF and PACF plot of the training data.

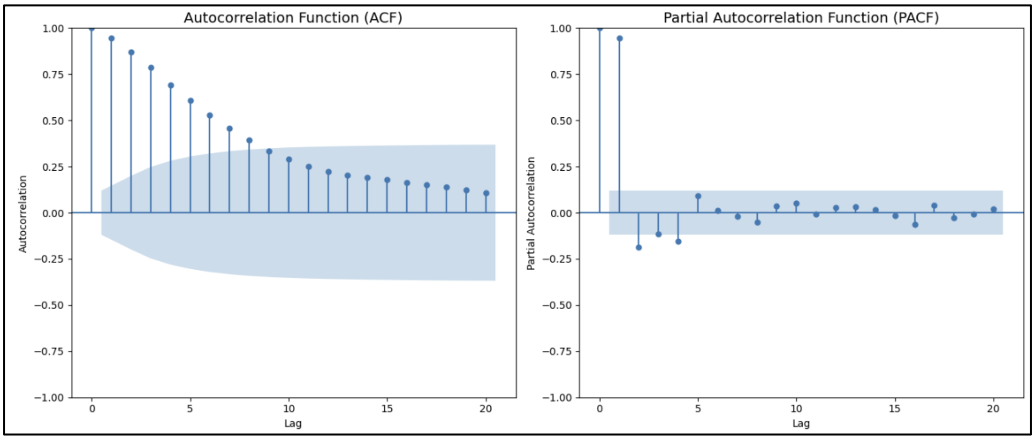

Figure 4 ACF and PACF plot based on training data.

Based Figure 4 presents the results of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analysis based on training data. These plots are used to evaluate the autocorrelation patterns in the data before applying the ARIMA model. In the ACF plot, the autocorrelation values gradually decrease as the lag increases, indicating the presence of serial dependence in the data. This suggests that the data still exhibits non-stationary patterns. Meanwhile, in the PACF plot, a sharp cutoff occurs after the first lag, suggesting the presence of an autoregressive (AR) component in the ARIMA model. These plots help determine the optimal parameters for the ARIMA model, specifically p (AR order) and q (MA order), by considering the significant lags in each plot. Subsequently, the ARIMA model will be applied to the transformed data to ensure stationarity, allowing for more accurate forecasting.

According to [12], [18], non-stationarity in data is not an issue when performing data analysis using deep learning models such as LSTM. These models enable more flexible analysis and can effectively handle the complexities of non-stationary data, eliminating the need for special steps to make the data stationary before modeling [12], [18]. This contrasts with statistical methods such as ARIMA, which heavily rely on stationarity to produce accurate predictions. In the ARIMA model, which analyzes time series data, the stability between mean and variance values is crucial, as data with a consistent mean indicates that it is stationary. To achieve stationarity, data often requires transformations such as differencing before applying the ARIMA model. Without these transformations, ARIMA may produce inaccurate and unstable predictions due to its inability to handle complex non-stationary patterns. The ARIMA model identification process also involves analyzing autocorrelation (ACF) and partial autocorrelation (PACF) patterns in the transformed data to determine the optimal parameters.

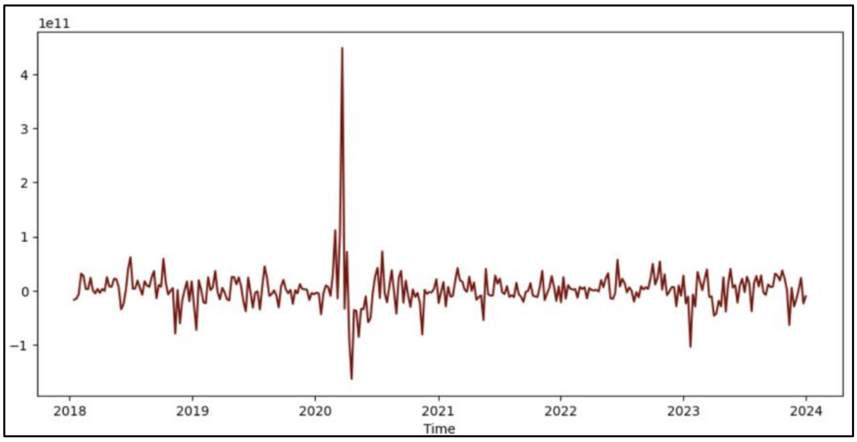

Figure 5 Plot of Training Data After Differencing

As shown in Figure 5, after differencing was applied, the Augmented Dickey-Fuller (ADF) test was re-evaluated to determine whether the data had become stationary with respect to its mean. The test results confirm that after differencing, the data is now stationary in terms of its mean. Differencing is a technique used to remove temporal dependence (serial correlation) and make the data stationary, especially when the data exhibits a clear trend. In this case, differencing successfully made the data stationary by eliminating the existing trend. This process involves subtracting the current observation from the previous observation, which helps reduce or eliminate trends in the data. By applying differencing, non-stationary patterns in the data can be addressed, allowing statistical models such as ARIMA to produce more accurate forecasts. Stationarity is crucial in time series modeling because many statistical methods, such as ARIMA, assume that the data is stationary to produce valid predictions.

3) ARIMA Model Identification and Parameter Estimation

The ARIMA model can be identified once the data is stationary. Stationarity is confirmed by analyzing the Partial Autocorrelation (PACF) and Autocorrelation (ACF) plots. The PACF plot is used to determine the optimal order for the AR(p) (Autoregressive) component, while the ACF plot helps identify the optimal order for the MA(q) (Moving Average) component. In this study, stationarity was achieved through a single differencing operation.

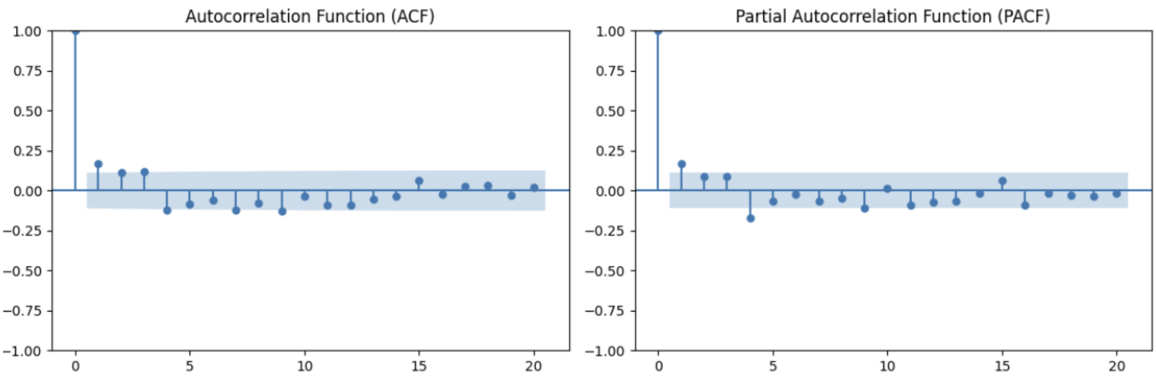

Figure 6 Plots of ACF and PACF after the differencing process.

Figure 6 presents the ACF and PACF plots after the differencing process was applied to the weekly highest exchange rate of the Indonesian Rupiah (IDR) against the US Dollar (USD). The ACF indicates that most autocorrelation values fall within the confidence interval after the first lag, suggesting that long-term autocorrelation components have been reduced after differencing. Meanwhile, the PACF shows a significant cutoff after the first lag, indicating that the appropriate model for this data is likely ARIMA (1,1,0). Based on these results, the ARIMA (1,1,0) model has been selected as a tentative model for further analysis.

4) Finding the definitive model Based on AIC and BIC

To find the definitive model, several different ARIMA models have been evaluated based on their AIC and BIC values to identify the best-fitting model. The following table compares the AIC and BIC values of various ARIMA models:

Table 1 Evaluation of ARIMA models based on AIC values

Model	AIC	BIC	HQIC
ARIMA (1,1,1)	4047,25	4058,4789	4051.7378
ARIMA (1,1,0)	4047.3021	4054,7881	4050,2939
ARIMA (0,1,1)	4048,3297	4055,8157	4051.3216
ARIMA (2,1,1)	4049,0779	4064.0499	4055.0618
ARIMA (1,1,2)	4048,8695	4063.8415	4054.853

Based on Table 1, the most appropriate model, identified by having the lowest AIC BIC and HQIC value of 4047.3021, 4054,7881, and 4050,2939, is ARIMA (1,1,0). To refine the model selection, the processes of overfitting and underfitting were conducted by increasing and decreasing the order of the AR (Autoregressive) or MA (Moving Average) components. Based on the models in Table 1, potential candidates were proposed for further evaluation based on their AIC and BIC values. As shown in Table 1. ARIMA (1,1,0) has the lowest AIC and BIC values, making it the temporarily selected model for further analysis. The next step is to evaluate the accuracy of this model to ensure its performance in forecasting data.

5) Evaluation of the selected ARIMA model.

Effectiveness of the ARIMA (1,1,0) on the test data can be evaluated using metrics such as the MAPE, RMSE, and MSE model's predictions achieved a 99,37% accuracy according to MAPE, with an MAPE value of 0.63 %, an RMSE of 133,58, and an MSE of 17844,65, indicating excellent prediction performance.

6) LSTM Model

The initial step in utilizing the LSTM model for forecasting is designing its architecture. The model structure consists of two LSTM layers with specified numbers of neurons, epochs, and learning rates, followed by a dense output layer. The Adam optimizer is used in this study as it is known to enhance training efficiency and is recognized by Mehmood et al. (2023) as one of the most effective optimization methods [17]. To prevent overfitting, an early stopping technique is applied by monitoring the validation loss and halting the training process when no further improvement is observed.

In this study, a weekly univariate time series of the highest IDR/USD exchange rate is used as the input. The LSTM model is configured to process the data with one time step and one feature per sequence, allowing the model to learn temporal dependencies across time periods. The final dense layer is used to map the learned representation into a single prediction output. This structure enables the LSTM to recognize both short-term memory and long-term dependencies, which are commonly present in currency exchange rate data.

The selection of neuron counts (20, 32, 56, 64, 100, 200), epochs (50, 100, 200, 500), learning rates, and optimizer types was conducted systematically using a Grid Search approach on a predefined parameter space. This structured process allows for the evaluation of multiple parameter combinations to identify the optimal configuration based on validation performance. Such an approach not only increases the likelihood of obtaining the best-performing configuration but also enhances the reproducibility of the experiment. The batch size was calculated based on the greatest common divisor of the training and testing dataset sizes, and in this study, it was set to 56. The following table presents the hyperparameters used in the implementation of the LSTM model:

Table 2 Hyperparameters Used in the LSTM Model

Hyperparameter	Value
Optimizer	Adam
Neuron	20, 32,56, 64, 100, 200
Batch size	56
Learning Rate	0,01, 0,001, 0,0001
Epoch	50, 100, 200, 500
Layer	1
Dropout	0,2

The lower part of the text explains that based on the results from Grid Search, the best hyperparameters were obtained. These optimal hyperparameters will be implemented to assess and validate the LSTM model's effectiveness using the test dataset. The use of Grid Search allows for systematically identifying the best-performing configuration of hyperparameters, which is critical for improving the model's predictive performance.

Table 3 Best Selected Hyperparameters

Model	Hyperparameter				RMSE	MAPE	MSE
	Optimizer	Learning Rate	Neuron	Epoch			
LSTM	Adam	0,0001	100	500	166,6910	0,0086	27785,8876

The study results indicate that the LSTM model successfully achieved a MAPE of 0,0086%, an RMSE of 166,6910 and an MSE of 27785,8876. The optimizer utilized and trained in this study was Adam, with a learning rate of 0,0001, 100 neurons, and 500 epochs. These findings confirm that the LSTM model effectively predicted gold futures prices with minimal errors, as evidenced by the forecasting accuracy metrics used in this study.

7) Ensemble Averaging Model

To combine the best models, ARIMA (1,1,0) and LSTM, an ensemble averaging approach is used, utilizing weighted averaging based on the RMSE of the individual models, The RMSE of the ARIMA (1,1,0) model is 133,545 while the RMSE of the LSTM model is 260,652, Since RMSE measures prediction error, a smaller RMSE indicates higher prediction accuracy, so a higher weight is assigned to the model with the lower RMSE, The resulting weights are ARIMA model weight: 0,66 and LSTM model weight: 0,34, This weighting suggests that the LSTM model is slightly more reliable in predicting compared to the ARIMA model, based on the RMSE accuracy test, After the weighting was determined, predictions were made, and the accuracy of the ensemble model was evaluated using MAPE, RMSE, and MSE metrics, The evaluation results are as follows:

Table 4 Evaluation of Ensemble Averaging Model Based on MAPE, RMSE, and MSE

Method	MAPE	RMSE	MSE
Ensemble Averaging	0,0130	248,6772	61840.368794

These results show that the ensemble model successfully provided predictions with very low error, The extremely small MAPE indicates that the average relative prediction error is very low, indicating high accuracy, The RMSE and MSE values also demonstrate excellent performance, being lower than the individual model predictions,

8) **Prediction Results of the Optimal Model,**

From the evaluation of all models—ARIMA, LSTM, and combinations of single models utilizing the averaging method—it was determined that each model produced relatively low MAPE, RMSE, and MSE values, However, when considering the ranking of the smallest values, the ensemble averaging model demonstrated the best performance, achieving the lowest MAPE, RMSE, and MSE values. These results indicate that the ensemble averaging approach effectively combines the strengths of each model, producing more accurate predictions compared to individual models. Below is a graph illustrating the comparison of predictions across all models.

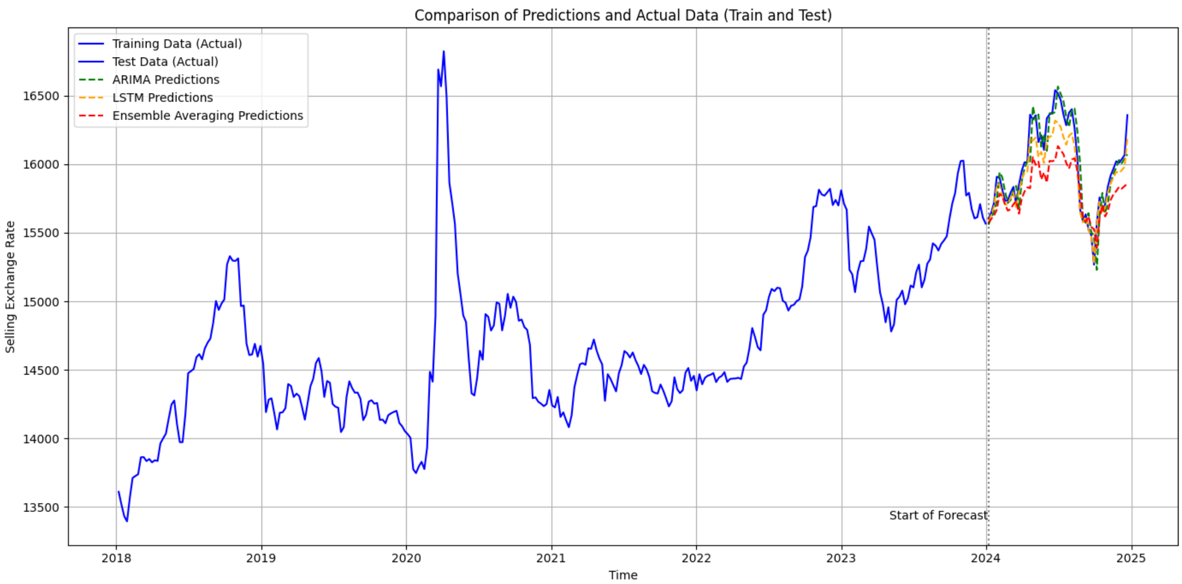

Figure 7 comparison of model predictions

9) **Forecast Results for Future Periods of All Models**

After identifying the best model based on forecasting accuracy and noting that the accuracy of all models is satisfactory, the subsequent step involves forecasting for the next 12 periods. This forecasting aims to evaluate which model performs most effectively for short-term and long-term predictions. The following table displays the forecast outcomes for the upcoming 12 periods across all models.

Table 5 Forecast Results for the Next 12 Periods Using ARIMA, LSTM, and Ensemble Averaging Models

NO	Period	Forecast Results		
		ARIMA	LSTM	Ensemble Averaging
1	2025-01-05	16327.81308	16195.52539	16268.962
2	2025-01-12	16327.05997	16078.14453	16216.32447
3	2025-01-19	16326.93214	15989.02734	16176.60777
4	2025-01-26	16326.91044	15918.83496	16145.3691
5	2025-02-02	16326.90676	15862.0166	16120.09016
6	2025-02-09	16326.90613	15815.04785	16099.19474
7	2025-02-16	16326.90603	15775.56836	16081.63136
8	2025-02-23	16326.90601	15741.93457	16066.66862
9	2025-03-02	16326.90601	15712.96094	16053.77996
10	2025-03-09	16326.906	15687.76953	16042.57211
11	2025-03-16	16326.906	15665.69238	16032.7506
12	2025-03-23	16326.906	15646.21582	16024.08603

Based on Table 5, it can be concluded that each forecasting model has different characteristics in handling short-term and long-term forecasting. The ARIMA model shows relatively consistent results in each period, with a value of around 16,326. This indicates that ARIMA is capable of capturing stable and repetitive patterns, making it suitable for short-term forecasting, where stability and reliability are key factors. However, this model is less flexible in capturing complex patterns or dynamic changes over the long term, as it relies heavily on recurring historical patterns.

In contrast, the LSTM model demonstrates strong capabilities in capturing complex and dynamic patterns, making it highly suitable for long-term forecasting. However, LSTM may become less stable if the historical data used is small or fails to reflect clear patterns for distant future periods. As a balanced alternative, the Ensemble Averaging model combines the strengths of ARIMA in short-term stability and LSTM in long-term adaptability. This model provides more stable predictions compared to LSTM and is more adaptive than ARIMA. By combining both methods, the ensemble model delivers more accurate and reliable forecasting results for various time horizons. This makes it the best choice for applications requiring a balance between short-term accuracy and long-term stability.

V. CONCLUSIONS AND SUGGESTIONS

The evaluation results indicate that all models show good performance, with MAPE values below 10%. The ARIMA model shows the lowest RMSE and MSE, with an RMSE of 133.58 and an MSE of 17,844.65, signifying the lowest prediction error. However, ARIMA is more stable in capturing short-term patterns, with an accuracy of 99.37% based on a MAPE of 0.63%. On the other hand, the ensemble approach (Weighted Average), which integrates ARIMA and LSTM, achieves the lowest MAPE of 0.013%, resulting in the highest accuracy of 99.87%. Although the RMSE and MSE values of the ensemble model are higher than those of ARIMA, this method remains superior in terms of accuracy, delivering the smallest overall error, particularly in long-term forecasting.

The LSTM model, with an RMSE of 166.691 and an MSE of 27,785.8876, demonstrates that this model is more prone to prediction errors compared to ARIMA and the ensemble model. However, it still maintains an impressive accuracy of 99.99% with a MAPE of 0.0086%, indicating its effectiveness in capturing complex patterns in the data. Despite its high accuracy, LSTM tends to be less stable for long-term forecasting, particularly when dealing with limited datasets.

Overall, although the ARIMA model exhibits the lowest prediction errors in terms of RMSE and MSE, the ensemble averaging model, despite having the lowest accuracy among the models, still provides sufficiently good predictions with acceptable performance. The ensemble model strikes a balance between stability and adaptability, making it a suitable choice for long-term forecasting, especially in scenarios where both short-term stability and long-term trend adaptability are crucial. Therefore, the ensemble averaging model emerges as a reliable forecasting strategy, effective for economic planning, investment decision-making, and business strategies that rely on exchange rate and market trend predictions.

REFERENCES

[1] G, Shmueli and P, Julia, *Practical Time Series Forecasting With R*, Third edit, Axelrod schnall publisher, 2024,

[2] V, E, M, Spyros Makridakis, Steven C,Wheelwright, *Metode dan aplikasi peramalan*, 1999,

[3] M, Abror Gustiansyah *et al*,, "Aplikasi Model ARIMA dalam Peramalan Data Harga Emas Dunia Tahun 2010-2022," *Jurnal Statistika dan Aplikasinya*, vol, 7, no, 1, 2023,

[4] Y, X, Wu, Q, B, Wu, and J, Q, Zhu, "Improved EEMD-based crude oil price forecasting using LSTM networks," *Physica A: Statistical Mechanics and its Applications*, vol, 516, pp, 114–124, Feb, 2019, doi: 10,1016/j,physa,2018,09,120,

[5] Md, R, Karim *et al*,, "Interpreting Black-box Machine Learning Models for High Dimensional Datasets," pp, 1–16, Aug, 2022, [Online], Available: http://arxiv,org/abs/2208,13405

[6] G, V, Datla, H, Jiang, and J, T, L, Wang, "An Interpretable LSTM Network for Solar Flare Prediction," in *International Conference on Tools with Artificial Intelligence, ICTAI*, IEEE Computer Society, 2023, pp, 526–531, doi: 10,1109/ICTAI59109,2023,00084,

[7] S, Sen, D, Sugiarto, and A, Rochman, "Komparasi Metode Multilayer Perceptron (MLP) dan Long Short Term Memory (LSTM) dalam Peramalan Harga Beras," *ULTIMATICS*, vol, XII, no, 1, p, 35, 2020,

[8] Y, S, Swarinoto, Y, Koesmaryono, E, Aldrian, and A, H, Wigena, "Model Sistem Prediksi Ensemble Total Hujan Bulanan Dengan Nilai Pembobot (Kasus Wilayah Kabupaten Indramayu)," *Jurnal Meteorologi dan Geofisika*, vol, 13, no, 3, Dec, 2012, doi: 10,31172/jmg,v13i3,134,

[9] S, Rahayu, S, Martha, S, Wira, and R, Intisari, "Prediksi Curah Hujan Dengan Metode Ensemble Averaging," *Buletin Ilmiah Math, Stat, dan Terapannya (Bimaster)*, vol, 11, no, 4, pp, 633–640, 2022,

[10] S, Chen and S, A, Billlngs, "Modelling and analysis of non-linear time series," *Int J Control*, vol, 50, no, 6, pp, 2151–2171, 1989, doi: 10,1080/00207178908953491,

[11] R, Santoso and S, Sudarno, "Metode Nonlinear Least Square (NLS) untuk Estimasi Parameter Model Wavelet Radial Basis Neural Network (WRBNN)," *MEDIA STATISTIKA*, vol, 10, no, 1, p, 49, Aug, 2017, doi: 10,14710/medstat,10,1,49-59,

[12] Moch Farryz Rizkilloh and Sri Widiyanesti, "Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term Memory (LSTM)," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol, 6, no, 1, pp, 25–31, Feb, 2022, doi: 10,29207/resti,v6i1,3630,

[13] Abdelhadi Azzouni and Guy Pujolle, "NeuTM: A Neural Network-based Framework for Traffic Matrix Prediction in SDN," *IEEE*, pp, 1–5, 2018,

[14] I, Zaier, C, Shu, T, B, M, J, Ouarda, O, Seidou, and F, Chebana, "Estimation of ice thickness on lakes using artificial neural network ensembles," *J Hydrol (Amst)*, vol, 383, no, 3–4, pp, 330–340, Mar, 2010, doi: 10,1016/j,jhydrol,2010,01,006,

[15] D, Yulianti, I, Made Sumertajaya, and I, D, Sulvianti, "Pemodelan Harga Beras di Pulau Sumatera dengan Menggunakan Model Generalized Space Time ARIMA," *Xplore*, vol, 2, no, 2, pp, 49–57, 2018,

[16] S, Rakhmawan, K, A, Notodiputro, and I, M, Sumertajaya, "A Study of Arima and Garch Models To Forecast Crude Palm Oil (Cpo) Export In Indonesia," *Proceeding of International Conference On Research, Implementation And Education Of Mathematics And Sciences*, pp, 251–260, 2015, [Online], Available: www,comtrade,un,org

[17] F, Chollet, *Deep Learning with Python*, Shelter Island: Manning Publications Co,, 2021,

[18] R, Liu, Y, Jiang, and J, Lin, "Forecasting the Volatility of Specific Risk for Stocks with LSTM," *Procedia Comput Sci*, vol, 202, pp, 111–114, 2022, doi: 10,1016/j,procs,2022,04,015,