Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2405−2415
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9414

A Comparative Study of LSTM and BiLSTM Performance in Predicting XAU/USD Prices

I Ketut Agung Enriko[1,*], Fikri Nizar Gustiyana[2]

[1] Faculty of Electrical Engineering, Telecommunication Engineering Study Program, Telkom University, Bandung, Indonesia
[2] Faculty of Electrical Engineering, Electrical Engineering Study Program, Telkom University, Bandung, Indonesia
Email: [1,*]iketutagungenriko@telkomuniversity.ac.id, [2]fikrinizargustiana7899@gmail.com
Correspondence Author Email: iketutagungenriko@telkomuniversity.ac.id
Submitted: **16/02/2026**; Accepted: **05/03/2026**; Published**: 06/03/2026**

Abstract−Gold price forecasting in the XAU/USD market is challenging due to nonlinear dynamics, high volatility, and sensitivity to global macroeconomic factors. This study compares the performance of Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) architectures in forecasting XAU/USD closing prices using historical data from 2023–2026. Data preprocessing includes cleaning, chronological ordering, normalization, and transformation using a sliding window approach. A window size of 60 time steps is selected to represent approximately three months of daily trading activity, enabling the models to capture short- to medium-term temporal dependencies while limiting excessive noise and computational burden. The dataset is divided chronologically into training and out-of-sample testing sets to ensure proper generalization assessment. Both models employ identical architectures with two recurrent layers (50 hidden units each) and are trained using the Adam optimizer with epoch variations (20–100). Evaluation on unseen test data uses MAE, MSE, RMSE, MAPE, and R^2 metrics. LSTM achieves its lowest MAE of 21.26 at 40 epochs, while BiLSTM attains its best performance at 80 epochs with an MAE of 20.86 and R^2 of 0.9981. However, extending training to 100 epochs leads to performance degradation in BiLSTM, indicating sensitivity to overtraining. Overall, optimal performance is achieved through balanced training duration rather than increased architectural complexity.

Keywords: XAU/USD; LSTM; BiLSTM; Value Forecasting; Time Series

1. INTRODUCTION

The global financial market plays a fundamental role in facilitating capital flows, investment activities, and risk management across countries. Within this system, the foreign exchange and commodity markets serve as key pillars influencing macroeconomic stability and global capital dynamics[1]. One of the most actively traded instruments in these markets is gold, commonly represented by the XAU/USD currency pair, which reflects the price of gold quoted in United States dollars[2]. Gold functions not only as a commodity but also as a hedge instrument and safe-haven asset during periods of economic instability, inflationary pressure, and geopolitical uncertainty. Consequently, XAU/USD price movements attract significant attention from investors, traders, and policymakers due to their strong sensitivity to monetary policy decisions, interest rate fluctuations, inflation expectations, global risk sentiment, and U.S. dollar strength[3].

Despite its strategic role, XAU/USD prices exhibit highly volatile and dynamic behavior. Price fluctuations occur continuously due to the interaction of multiple internal and external factors, including global macroeconomic indicators, central bank policies, geopolitical tensions, supply-demand imbalances, and investor sentiment. Unlike many traditional financial instruments, gold prices are particularly responsive to uncertainty shocks, resulting in complex, non-linear, and stochastic time-series patterns. This high level of volatility poses significant challenges for market participants in designing optimal trading and hedging strategies, as price movements are difficult to forecast accurately using conventional analytical approaches[4].

Traditional forecasting techniques, such as technical and fundamental analysis[5], primarily rely on historical price patterns[6] and macroeconomic indicators[7]. Although these methods provide useful insights, they often struggle to capture deep non-linear dependencies and long-term temporal relationships embedded in financial time-series data. The dynamic and evolving structure of financial markets limits the effectiveness of classical statistical models, especially under rapidly changing market regimes. In response to these limitations, artificial intelligence (AI)[8] and deep learning approaches have gained increasing attention in financial forecasting due to their ability to process large-scale data[9], automatically extract features, and identify hidden patterns that are not easily detectable using traditional methods[10].

Among deep learning architectures, Long Short-Term Memory (LSTM), an extension of Recurrent Neural Networks (RNN)[11], has been widely applied in time-series modeling because of its capability to learn long-term dependencies through gating mechanisms that mitigate the vanishing gradient problem[12]. LSTM has demonstrated promising performance in various financial prediction tasks[13]. However, financial time-series data such as XAU/USD prices often contain complex contextual relationships that may benefit from more advanced sequential modeling techniques. Bidirectional Long Short-Term Memory (BiLSTM) extends the conventional LSTM framework by processing sequential data in both forward and backward directions, enabling the model to capture information from past and future contexts simultaneously[14]. This bidirectional mechanism may enhance the model's ability to represent intricate temporal structures[15], particularly in highly volatile instruments such as gold.

Although numerous studies have implemented LSTM-based models for financial forecasting, several research gaps remain evident[16][17][18]. Although LSTM and its variants are widely used in financial time-series forecasting,

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2405−2415
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9414

several methodological and empirical gaps remain. Most studies focus on stock indices or cryptocurrencies, with limited systematic investigation of XAU/USD, whose safe-haven nature and sensitivity to macroeconomic uncertainty may influence recurrent learning dynamics differently. Moreover, prior comparisons between LSTM and BiLSTM are often conducted under heterogeneous experimental settings, reducing internal validity and obscuring the true architectural effect. In addition, limited research examines training stability and epoch sensitivity in bidirectional models, particularly regarding the trade-off between richer contextual representation and overfitting risk under controlled architectural conditions.

Therefore, a controlled benchmarking study that isolates architectural directionality while maintaining identical preprocessing procedures, network depth, hidden units, optimizer configuration, and evaluation metrics is required to provide clearer empirical evidence regarding the comparative effectiveness and stability of LSTM and BiLSTM in modeling XAU/USD price dynamics.Based on this context, the present study aims to develop and implement both LSTM and BiLSTM models for forecasting XAU/USD prices using historical time-series data and to conduct a comprehensive comparative analysis under consistent experimental conditions. Model performance is evaluated using standardized quantitative metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), the coefficient of determination (R^2 Score), and duration of computation , in order to assess predictive accuracy, robustness, and goodness-of-fit[19][20]. Through this evaluation, the study seeks to identify the more effective deep learning architecture for capturing non-linear temporal dependencies in gold price dynamics.

The findings of this study contribute by clarifying the impact of bidirectional recurrent processing on temporal representation in highly volatile financial time series, specifically XAU/USD prices, through a controlled empirical comparison of LSTM and BiLSTM under identical architectural and training conditions. Rather than introducing a new model, the study isolates the effect of sequence directionality to reveal the relative strengths, limitations, and stability characteristics of unidirectional and bidirectional recurrent structures across varying epoch configurations. The results highlight how bidirectional processing affects generalization performance, training stability, and sensitivity to extended epochs. Methodologically, the study establishes a controlled benchmarking framework by maintaining consistent hyperparameters, window size, and optimization settings across both models, enabling a transparent assessment of architectural effectiveness. This structured comparison helps isolate the impact of directional sequence learning on forecasting accuracy and overfitting behavior. From a practical perspective, the findings provide evidence-based insights into when the additional computational complexity of BiLSTM is justified compared to standard LSTM. By quantifying performance differences under out-of-sample evaluation, the study offers actionable guidance for selecting forecasting architectures in volatile market environments, particularly where model stability and training efficiency are critical considerations.

2. RESEARCH METHODOLOGY

2.1 Research Flow

The overall research workflow follows a structured experimental benchmarking framework, as illustrated in Figure 1. The process begins with a comprehensive literature review aimed at identifying established methodologies in financial time-series forecasting, particularly deep learning approaches based on recurrent neural networks. This stage focuses on understanding prior comparative studies involving Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM), identifying research gaps, and formulating a clear experimental objective centered on architectural effectiveness under controlled conditions.

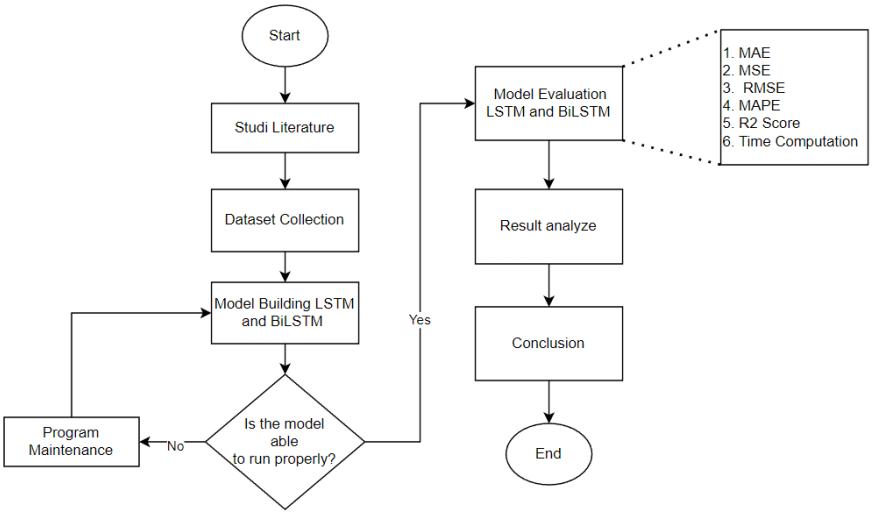


Figure 1. Research Flow

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2405−2415
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9414

2.2 Model Design

Historical XAU/USD closing price data are collected from reliable financial sources and arranged chronologically to preserve temporal order, which is critical in sequential modeling. Since time-series models depend strictly on past-to-future information flow, any disruption in chronological structure could introduce data leakage and invalidate the forecasting framework. Data cleaning procedures are performed to handle missing values, remove inconsistencies, and ensure numerical continuity. This step reduces noise that could otherwise distort gradient updates during training.

A normalization procedure is then applied using MinMaxScaler, transforming the data into a bounded interval of [0,1]. The selection of MinMaxScaler is technically grounded in the characteristics of LSTM and BiLSTM architectures. These models utilize sigmoid and tanh activation functions, both of which are sensitive to input magnitude. When input values are excessively large or unscaled, the activation functions may enter saturation regions, causing vanishing gradients and slowing convergence during backpropagation through time (BPTT). By constraining inputs to a fixed range, MinMax scaling enhances numerical stability and facilitates smoother gradient flow.

Compared to Z-score standardization, which centers data around zero with unit variance, MinMax scaling preserves the original distribution shape and proportional relationships between price movements. In financial time-series forecasting, relative magnitude and trend continuity are often more critical than variance normalization, particularly when modeling directional price dynamics. Additionally, while robust scaling techniques are effective in handling extreme outliers, they may distort sequential amplitude patterns when applied to already-cleaned financial datasets. Given that extreme anomalies are addressed during preprocessing, MinMaxScaler provides a computationally efficient and theoretically compatible normalization strategy.

After normalization, the dataset is transformed using a sliding window mechanism to convert the univariate time series into supervised learning sequences. This step constructs input-output pairs by mapping a fixed number of previous observations to the next predicted value, enabling the models to learn temporal dependencies explicitly. The processed sequences are then used for model development, training validation, and comparative performance evaluation, following the structured experimental pipeline presented in the figure.

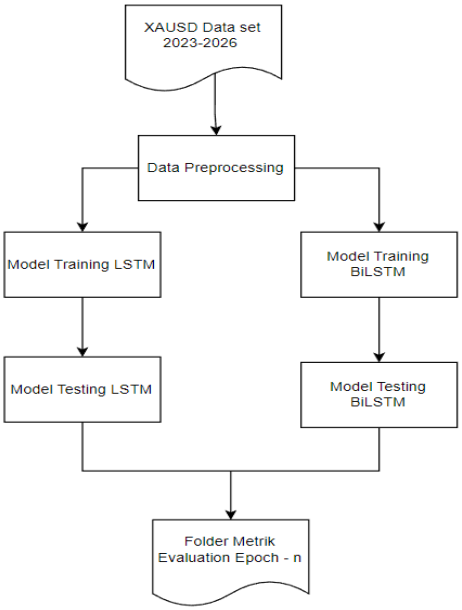


Figure 2. Program Flow

Based on Figure 2, the model development process begins with the utilization of the XAU/USD dataset covering the period 2023–2026 as the primary data source. The dataset undergoes a preprocessing stage that includes data cleaning, handling of missing values, chronological sorting, and normalization to ensure numerical stability during model training. This stage transforms the raw financial time-series data into a structured format suitable for deep learning modeling.

After preprocessing, the dataset is prepared using a sequential time-series approach without random shuffling in order to preserve its temporal structure. The data are organized chronologically so that the model learns patterns from historical observations and generates predictions based on subsequent time steps. This approach maintains the integrity of the financial time-series characteristics and avoids temporal distortion in sequence formation.

In this study, model evaluation is conducted within an in-sample framework, where the sliding window mechanism is applied sequentially across the entire normalized dataset to generate supervised input–output pairs. The same chronologically ordered data are used for model fitting and prediction, meaning the evaluation reflects the models' capacity to learn and reconstruct temporal patterns within the observed series.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2405−2415
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9414

Although in-sample evaluation may raise concerns regarding look-ahead bias, the primary objective of this research is architectural benchmarking rather than deployment-oriented forecasting. By exposing both LSTM and BiLSTM to identical data and experimental conditions, the study isolates structural learning behavior, convergence stability, and sensitivity to epoch variation without the additional variability introduced by train-test splits or walk-forward validation. This controlled setting ensures internal validity for comparative analysis, while broader generalization testing remains a direction for future research.

Following data preparation, two separate model development paths are implemented: the Long Short-Term Memory (LSTM) model and the Bidirectional Long Short-Term Memory (BiLSTM) model. Each architecture undergoes a training phase using identical structural configurations and hyperparameters to ensure a fair comparison, as illustrated in Figure 3. After training, both models proceed to the testing phase, where predictions are generated based on the constructed time-series sequences derived from the same dataset.

The final stage consists of performance evaluation and metric storage for each epoch configuration. Evaluation results, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), R^2 Score, and training time are recorded and organized systematically in dedicated metric folders for each model and epoch variation. This structured workflow enables a controlled and transparent comparison between unidirectional and bidirectional recurrent neural network architectures in modeling XAU/USD price dynamics within an in-sample experimental framework.

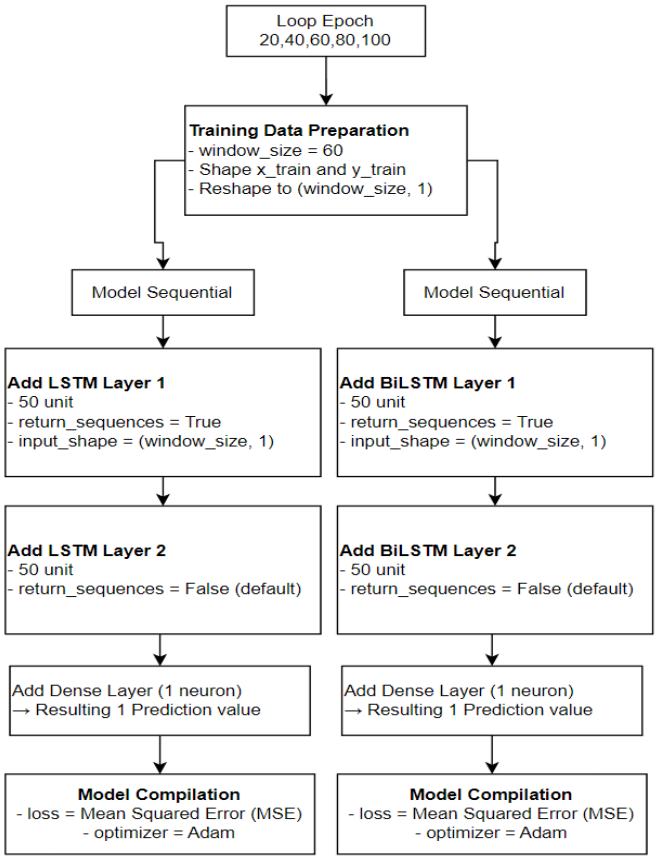


Figure 3. Model Configuration

Based on Figure 3, a sliding window technique with a window size of 60 time steps is applied to transform the time-series data into a supervised learning format, where sequences of past observations are used to predict the next closing price. The study develops two deep learning architectures, namely LSTM and BiLSTM, with identical structural configurations to ensure experimental fairness. Each model consists of two recurrent layers with 50 hidden units per layer, followed by a dense output layer that produces a single predicted value. Both models are trained using the Adam optimizer and mean squared error (MSE) as the loss function. To analyze the effect of training duration on model performance, multiple epoch configurations are tested, specifically 20, 40, 60, 80, and 100 epochs, while other architectural parameters remain constant to maintain controlled experimental conditions.

Model performance evaluation is conducted by generating predictions on the processed dataset and transforming the predicted values back to the original scale using inverse normalization. The predictive accuracy and goodness-of-fit of each configuration are assessed using standard regression metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2 Score). In addition, training time is recorded to compare computational efficiency between LSTM and BiLSTM architectures.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2405−2415
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9414

2.3 Data Set Sample

Meanwhile, representative samples of the processed dataset are provided in Table 1 to illustrate the format, structure, and characteristics of the data used in the experiment. These tables collectively offer a clearer understanding of how the time-series data are organized prior to model training and evaluation.

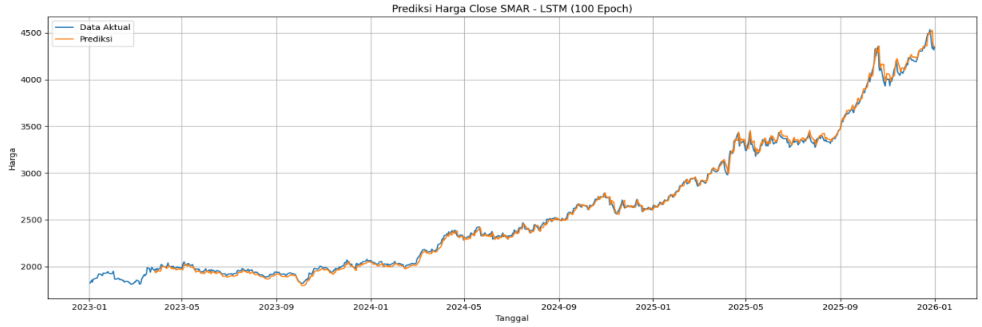
Table 1. Example Data Historical XAU/USD

Date	Close	Open	High	Low
01/01/2026	4.348,35	4.329,42	4.351,65	4.324,42
31/12/2025	4.315,09	4.346,47	4.373,89	4.274,51
…….	…….	…….	…….	…….
…….	…….	…….	…….	…….
04/01/2023	1.854,09	1.839,51	1.865,49	1.835,92
03/01/2023	1.839,49	1.826,14	1.850,69	1.824,42
02/01/2023	1.823,69	1.823,85	1.823,94	1.823,85

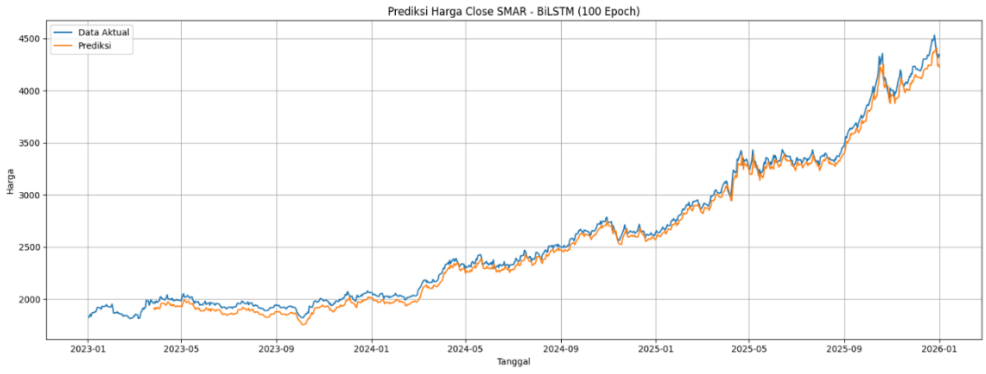
3. RESULT AND DISCUSSION

3.1 Model Prediction Result

The graphical results of the in-sample prediction illustrate the comparative performance of the LSTM and BiLSTM models in capturing the historical movement of XAU/USD closing prices over the 2023–2026 period. As observed in the figure 4, the predicted values generated by both architectures closely follow the actual price trajectory, indicating that the models are capable of learning underlying temporal patterns within the dataset. The prediction curves demonstrate a strong alignment with the overall upward trend, short-term fluctuations, and intermediate corrections present in the historical data.



(a)



(b)

Figure 4. Result of Prediction Graph using LSTM (a) and BiLSTM (b) (sample using Epoch 100)

3.2 Mean Absolute Error (MAE) Result

To further evaluate the predictive performance of the LSTM and BiLSTM models, the Mean Absolute Error (MAE) is employed to quantify the average magnitude of prediction errors without considering their direction. MAE provides an intuitive measure of how closely the predicted values align with the actual data, where lower values indicate higher predictive accuracy and better model fit. The MAE results obtained from different epoch configurations for both architectures are summarized in Table 2, while the comparative performance patterns between LSTM and BiLSTM are visually presented in Figure 5.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2405−2415
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9414

Table 2. Mean Absolute Error Result

Epoch	LSTM	BiLSTM
20	36.12	22.20
40	21.26	28.63
60	24.61	21.62
80	31.04	20.86
100	25.74	56.71

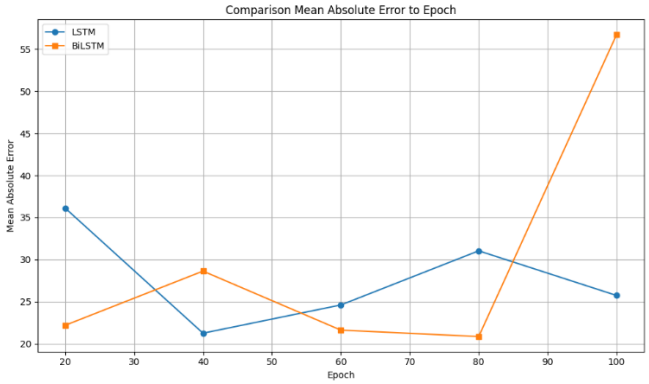


Figure 5. Mean Absolute Error Result

3.3 Mean Squared Error (MSE) Result

Mean Squared Error (MSE) is employed as an evaluation metric to measure the predictive performance of the LSTM and BiLSTM models by calculating the average of the squared differences between the actual values and the predicted outputs. By squaring the error terms, MSE amplifies the impact of larger deviations, making it particularly sensitive to extreme prediction errors that may occur during periods of heightened market volatility. This characteristic is especially relevant in financial time-series modeling, where sudden price movements can significantly influence model reliability and forecasting stability.

In the context of XAU/USD price prediction, the use of MSE enables a more rigorous assessment of how well each architecture handles sharp fluctuations and non-linear dynamics embedded in the dataset. While MAE provides an average absolute deviation measure, MSE emphasizes the consistency of predictions by penalizing large residuals more heavily. Therefore, lower MSE values indicate not only higher predictive accuracy but also greater robustness in managing substantial forecasting errors. The detailed MSE results across different epoch configurations are presented in Table 3, and the comparative performance trends between LSTM and BiLSTM are illustrated in Figure 6, allowing for a clearer interpretation of error distribution patterns across training durations.

Table 3. Mean Squared Error Result

Epoch	LSTM	BiLSTM
20	2850.19	1021.72
40	987.75	1424.87
60	1275.70	1006.12
80	1524.17	950.40
100	1201.67	4030.32

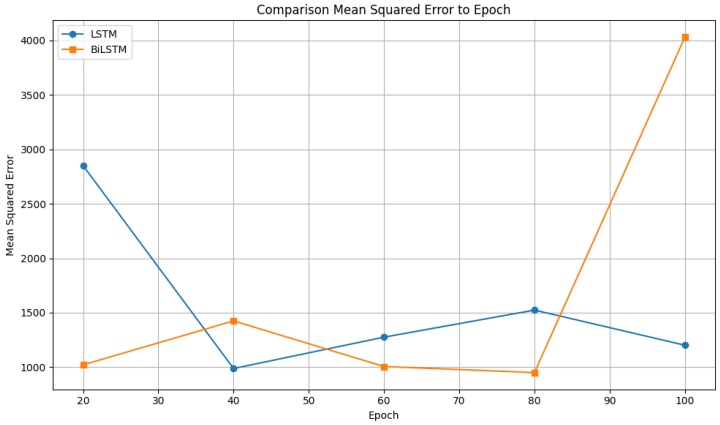


Figure 6. Mean Squared Error Result

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2405−2415
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9414

Based on the MSE values presented in Table 3, clear differences in error magnitude can be observed between the LSTM and BiLSTM models across various epoch configurations. At 20 epochs, LSTM records a relatively high MSE of 2850.19, whereas BiLSTM achieves a substantially lower value of 1021.72, indicating that the bidirectional architecture produces smaller squared prediction errors during early training stages.

At 40 epochs, the pattern shifts, with LSTM achieving a lower MSE of 987.75 compared to BiLSTM's 1424.87, suggesting that LSTM adapts more effectively at this training duration. When the number of epochs increases to 60, BiLSTM again demonstrates better performance, obtaining an MSE of 1006.12, slightly lower than LSTM's 1275.70. A similar trend is observed at 80 epochs, where BiLSTM reaches its lowest MSE value of 950.40, while LSTM records 1524.17, indicating stronger stability of the bidirectional model under this configuration.

However, at 100 epochs, a significant increase in BiLSTM's MSE is observed, reaching 4030.32, whereas LSTM maintains a comparatively moderate value of 1201.67. This sharp rise suggests that prolonged training may introduce instability or overfitting behavior in the BiLSTM model under the in-sample setting. Overall, the fluctuation of MSE values across epochs highlights that model performance does not improve monotonically with additional training iterations, and optimal results are achieved at specific epoch configurations for each architecture.

3.4 Root Mean Squared Error (RMSE) Result

Root Mean Squared Error (RMSE) is utilized as an additional evaluation metric to measure the average magnitude of prediction errors while maintaining the original unit scale of the data. RMSE is derived from the square root of the Mean Squared Error (MSE), which allows it to reflect the dispersion of prediction errors in a more interpretable form. Because it penalizes larger errors through the squaring process before taking the root, RMSE remains sensitive to significant deviations while still being expressed in the same unit as the XAU/USD price.

In financial time-series forecasting, RMSE is particularly useful for assessing how closely the predicted values follow actual price movements, especially during periods of high volatility. Lower RMSE values indicate that the model produces predictions with smaller overall deviation from the observed data, reflecting better predictive accuracy and stability. Because RMSE penalizes larger errors more heavily due to the squaring process, it provides additional insight into model robustness during extreme price fluctuations. Therefore, this metric is effective for evaluating the consistency of predictive performance across varying training durations. The detailed RMSE results across different epoch configurations are summarized in Table 4, while the comparative performance trends between LSTM and BiLSTM are illustrated in Figure 7.

Table 4. Root Mean Squared Error Result

Epoch	LSTM	BiLSTM
20	53.39	31.96
40	31.43	37.75
60	35.72	31.72
80	39.04	30.83
100	34.67	63.48

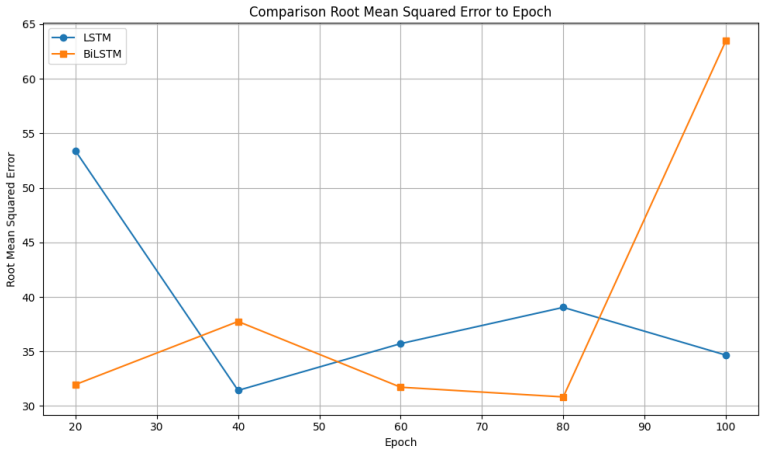


Figure 7. Root Mean Squared Error Result

The RMSE results in Table 4 reveal noticeable differences in error dispersion between the LSTM and BiLSTM architectures across epoch variations. At 20 epochs, the LSTM model produces an RMSE of 53.39, which is substantially higher than BiLSTM's 31.96, indicating that the bidirectional network is able to approximate the actual price values more closely during shorter training duration. This suggests that BiLSTM converges faster in capturing temporal dependencies at the initial stage.

When the number of epochs increases to 40, the pattern reverses. LSTM reduces its RMSE significantly to 31.43, outperforming BiLSTM, which records 37.75. This indicates that the unidirectional architecture benefits more

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2405−2415
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9414

from moderate training iterations under this configuration. At 60 epochs, the RMSE values become relatively close, with LSTM at 35.72 and BiLSTM at 31.72, where the bidirectional model again demonstrates slightly better stability. A similar tendency appears at 80 epochs, as BiLSTM achieves its lowest RMSE value of 30.83, while LSTM shows a higher deviation of 39.04, reflecting stronger predictive consistency from the bidirectional structure at this training duration.

A contrasting behavior emerges at 100 epochs. BiLSTM experiences a sharp increase in RMSE to 63.48, whereas LSTM maintains a comparatively stable value of 34.67. This substantial rise suggests that prolonged training may introduce instability or overfitting tendencies in the BiLSTM configuration within the in-sample experiment. Overall, the RMSE findings indicate that predictive performance fluctuates depending on epoch selection, and the relationship between training duration and error reduction is not strictly linear for either architecture.\

3.5 Mean Absolute Percentage Error (MAPE) Result

The MAPE results for each optimizer under varying epoch and hidden unit settings are presented in Table 8, with the corresponding percentage error patterns illustrated in Figure 8.Mean Absolute Percentage Error (MAPE) is employed to evaluate the relative prediction accuracy of the LSTM and BiLSTM models by measuring the average percentage difference between the actual and predicted values. Unlike MAE and RMSE, which express errors in absolute price units, MAPE presents the error in percentage form, making it easier to interpret the magnitude of forecasting deviation in proportional terms. This characteristic is particularly useful in financial time-series analysis, as it allows performance comparison across different price levels and time periods.

By expressing prediction errors as percentages, MAPE provides insight into how large the forecasting deviation is relative to the actual XAU/USD price. Lower MAPE values indicate higher predictive accuracy and better model reliability in tracking price movements. However, since MAPE involves division by actual values, it is sensitive when the observed data approach very small magnitudes, although this limitation is generally minimal in gold price data due to its relatively high price scale. The detailed MAPE results across epoch configurations are presented in Table 5, and the comparative performance trends between LSTM and BiLSTM are illustrated in Figure 8.

Table 5. Mean Absolute Percentage Error Result

Epoch	LSTM(%)	BiLSTM(%)
20	1.21	0.81
40	0.76	1.08
60	0.87	0.77
80	1.23	0.75
100	0.97	2.24

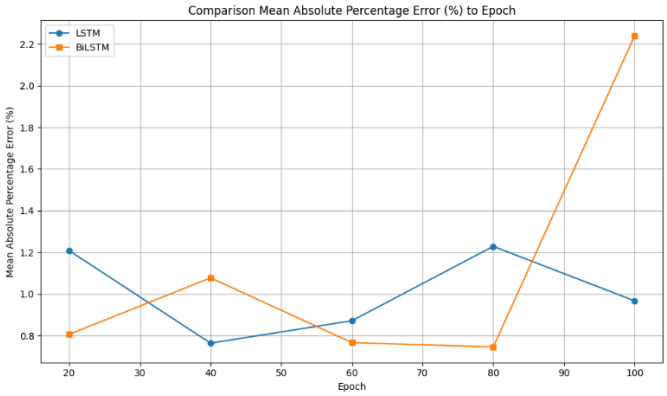


Figure 8. Mean Absolute Percentage Error Result

Based on the MAPE values presented in Table 5, both LSTM and BiLSTM exhibit low percentage errors across most epoch configurations, indicating that the predicted values closely follow the actual XAU/USD price movements in proportional terms. At 20 epochs, LSTM records a MAPE of 1.21%, whereas BiLSTM achieves a lower value of 0.81%, suggesting that the bidirectional architecture produces more accurate relative predictions during the initial training stage.

When the number of epochs increases to 40, LSTM improves significantly to 0.76%, outperforming BiLSTM, which records 1.08%. This indicates that at this configuration, the unidirectional model provides more precise percentage-based forecasts. At 60 epochs, both models demonstrate comparable performance, with LSTM at 0.87% and BiLSTM slightly lower at 0.77%, reflecting stable proportional prediction accuracy.

At 80 epochs, BiLSTM achieves its lowest MAPE value of 0.75%, while LSTM shows 1.23%, indicating stronger relative prediction capability of the bidirectional structure under this training duration. However, at 100 epochs, BiLSTM experiences a marked increase in MAPE to 2.24%, whereas LSTM maintains a relatively moderate value of 0.97%. This rise suggests that extended training may reduce proportional stability in the BiLSTM

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2405−2415
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9414

configuration within the in-sample setting. These findings demonstrate that percentage-based prediction accuracy is influenced by epoch selection, and performance improvements do not occur consistently as training iterations increase

3.6 R² Score Result

The coefficient of determination (R² Score) is employed to measure the goodness-of-fit of the LSTM and BiLSTM models in explaining the variability of the actual XAU/USD closing prices. R² represents the proportion of variance in the dependent variable that can be explained by the predictive model. Its value ranges from 0 to 1, where values closer to 1 indicate a stronger explanatory capability and a better alignment between predicted and observed data.

In financial time-series forecasting, the coefficient of determination (R²) reflects how well a model explains the variability of price movements beyond merely minimizing prediction error. Rather than emphasizing the magnitude of residuals, R² evaluates the extent to which the predicted values align with the overall trend and dynamic structure of the observed series. Higher R² values indicate stronger explanatory power and a closer fit to the underlying fluctuations of the dataset, while lower values suggest that substantial variance remains unaccounted for by the model. In volatile markets such as XAU/USD, maintaining a consistently high R² is particularly important to ensure that both trend direction and short-term corrections are adequately represented. The R² results for each epoch configuration are presented in Table 6, and the comparative trends between LSTM and BiLSTM are illustrated in Figure 9.

Table 6. R2 Score Result

Epoch	LSTM	BiLSTM
20	0.9942	0.9979
40	0.9980	0.9971
60	0.9974	0.9980
80	0.9969	0.9981
100	0.9976	0.9919

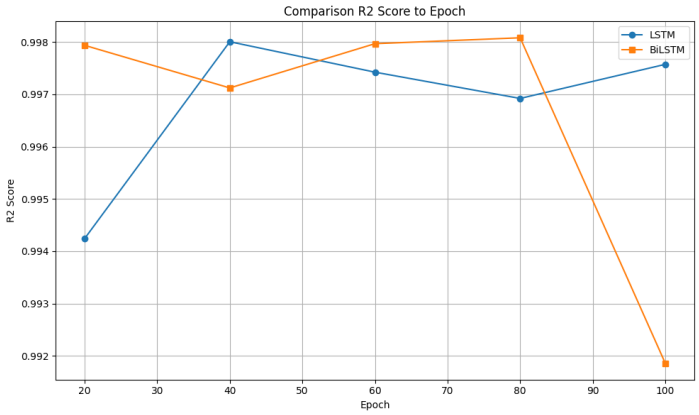


Figure 9. R2 Score Result

Based on the R² values presented in Table 6, both LSTM and BiLSTM demonstrate a strong ability to explain the variance of the XAU/USD closing prices within the in-sample evaluation framework. At 20 epochs, LSTM achieves an R² of 0.9942, while BiLSTM records a higher value of 0.9979, indicating that the bidirectional architecture captures a greater proportion of the data variability at the early training stage.

At 40 epochs, LSTM improves to 0.9980, slightly surpassing BiLSTM, which obtains 0.9971. This suggests that the unidirectional model provides stronger explanatory capability under this configuration. When the epoch increases to 60, BiLSTM reaches 0.9980, marginally higher than LSTM's 0.9974, reflecting effective modeling of temporal dependencies at this training duration. A similar trend is observed at 80 epochs, where BiLSTM achieves its highest R² value of 0.9981, compared to LSTM's 0.9969, indicating superior goodness-of-fit for the bidirectional structure under this setting.

However, at 100 epochs, BiLSTM shows a noticeable decline to 0.9919, whereas LSTM maintains a relatively stable R² of 0.9976. This reduction suggests that extended training may reduce the explanatory consistency of the BiLSTM model in the in-sample context. The R² results indicate that both architectures are capable of modeling the majority of price variability, although their explanatory strength fluctuates depending on the selected epoch configuration.

3.7 Computation Time Result

Computation Time is analyzed to evaluate the training efficiency of the LSTM and BiLSTM models across different epoch configurations. This metric measures the total duration required for each model to complete the training process, providing insight into computational complexity and resource requirements. In deep learning-based financial

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2405−2415
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9414

forecasting, training time is an important consideration, particularly when models are intended for practical implementation or real-time analysis.

Variations in computation time are influenced by architectural design, parameter magnitude, and the number of recurrent operations executed during backpropagation through time. Because BiLSTM processes sequences in both forward and backward directions, it inherently requires more computational resources than standard unidirectional LSTM. This bidirectional mechanism effectively increases the number of parameter updates and gradient calculations within each training iteration, contributing to longer training duration. Evaluating computation time together with accuracy metrics therefore enables a more balanced assessment of model performance, taking into account both predictive effectiveness and computational cost. The detailed training time results for each epoch configuration are presented in Table 7 and illustrated in Figure 10.

Table 7. Computation Time Result

Epoch	LSTM (s)	BiLSTM (s)
20	104.84	173.36
40	222.85	345.25
60	339.14	519.16
80	417.23	691.42
100	527.44	877.85

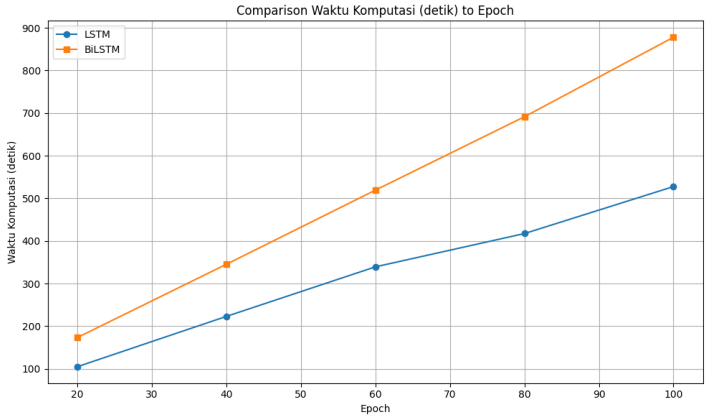


Figure 10. Computation Time Result

Based on the computation time results presented in Table 7, a consistent increase in training duration is observed as the number of epochs increases for both architectures. For the LSTM model, the training time rises from 104.84 seconds at 20 epochs to 222.85 seconds at 40 epochs, and further to 339.14 seconds at 60 epochs. This upward trend continues at 80 epochs with 417.23 seconds and reaches 527.44 seconds at 100 epochs. The gradual increase reflects the direct relationship between the number of training iterations and computational workload.

A similar pattern is observed for the BiLSTM model, although with noticeably longer training durations at every epoch configuration. BiLSTM requires 173.36 seconds at 20 epochs, increasing to 345.25 seconds at 40 epochs and 519.16 seconds at 60 epochs. At 80 epochs, the computation time reaches 691.42 seconds, and at 100 epochs it extends to 877.85 seconds. The consistently higher training time compared to LSTM is attributed to the bidirectional architecture, which processes sequential data in both forward and backward directions, effectively doubling recurrent computations within each layer.

The comparison indicates that while BiLSTM may achieve competitive predictive accuracy under certain epoch configurations, it requires substantially greater computational resources. This difference highlights the trade-off between model complexity and training efficiency, which becomes an important consideration in practical financial forecasting applications.

4. CONCLUSION

The experimental findings indicate that the predictive performance of the LSTM and BiLSTM models is strongly influenced by epoch configuration and architectural characteristics. Both models achieve high explanatory capability, with R^2 values exceeding 0.99 across most configurations, demonstrating their ability to model the dominant variance of XAU/USD price movements within the observed dataset. The LSTM model exhibits stable and consistent behavior across epoch variations, achieving its best performance at 40 epochs with an MAE of 21.26 while maintaining strong explanatory power. Even as epochs increase, performance remains relatively stable. This robustness can be attributed to its unidirectional structure, which aligns naturally with forward-moving financial time-series forecasting and limits parameter complexity, thereby promoting smoother convergence and reduced sensitivity to prolonged training. In

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2405−2415
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9414

contrast, the BiLSTM model reaches its peak performance at 80 epochs, achieving a lower MAE of 20.86 and the highest R^2 of 0.9981. The bidirectional mechanism enhances contextual representation within each training window, enabling improved short-term pattern extraction. However, at 100 epochs, performance declines, with R^2 decreasing to 0.9919 and error values increasing, indicating higher susceptibility to overtraining due to its more complex parameter structure. From a computational perspective, BiLSTM requires longer training time, reflecting the cost of its enhanced representational capacity. Therefore, architectural superiority depends on the trade-off between marginal accuracy gains and training stability rather than metric magnitude alone.

REFERENCES

[1] M. R. S, M. L. Siraj, A. Caezar, and T. Tadampali, "Understanding Financial Risk Dynamics : Systematic Literature Review inquiry into Credit , Market , and Operational Risks," *J. Ilm. Akunt.*, vol. 7, no. 2, pp. 1186–1213, 2024, doi: https://doi.org/10.57178/atestasi.v7i2.927.

[2] J. T. Pakpaharan, "Analysis in Decision Making for XAUUSD (Gold / USD) Pair Transactions in Futures Trading," *J. Indones. Soc. Sci.*, vol. 5, no. 10, pp. 2587–2595, 2024, doi: https://doi.org/10.59141/jiss.v5i10.1235.

[3] H. Mangiring and P. Simarmata, "Gold as a Safe Haven During the 2025 Global Tax War : A Qualitative Literature Review," *J. Semesta Ilmu Manaj. dan Ekon.*, vol. 1, no. 4, 2025, doi: https://doi.org/10.71417/j-sime.v1i4.352.

[4] O. Chuang, R. Gupta, C. Pierdzioch, and B. Shu, "Financial Uncertainty and Gold Market Volatility : Evidence from a Generalized Autoregressive Conditional Heteroskedasticity Variant of the Mixed-Data Sampling (GARCH-MIDAS) Approach with Variable Selection," *Econmetrics*, vol. 12, no. 38, pp. 1–17, 2024, doi: https://doi.org/10.3390/econometrics12040038.

[5] I. A. F. Putri and Zawawi, "Pengaruh Penggunaan Analisis Fundamental Terhadap Pengambilan Posisi Transaksi Dalam Trading Emas," *J. EKBIS (Ekon. Bisnis) Politek. Piksi Ganesha*, vol. 12, no. 2, pp. 228–234, 2024, doi: https://doi.org/10.56689/ekbis.v12i2.1496.

[6] M. Isnin, T. Informatika, and U. G. Putih, "Predicted Increase in Gold Price Every Year with Impact on Economic Factors," *Int. J. Econ. Manag. Sci.*, vol. 1, no. 4, pp. 366–375, 2024, doi: https://doi.org/10.61132/ijems.v1i4.359.

[7] H. Xiong and W. Zhang, "How the US Macroeconomic Factors Affect the Gold Price ?," in *Proceedings of the 8th International Conference on Economic Management and Green Development*, 2024, pp. 32–37. doi: 10.54254/2754-1169/124/2024.MUR17825.

[8] E. Golafshani, A. A. Chiniforush, P. Zandifaez, and T. Ngo, "An artificial intelligence framework for predicting operational energy consumption in office buildings," *Energy Build.*, vol. 317, no. January, p. 114409, 2024, doi: http://doi.org/10.1016/j.enbuild.2024.114409.

[9] H. Handoko, A. Asrofiq, J. Junadhi, and A. S. Negara, "Sentiment Analysis of Sirekap Tweets Using CNN Algorithm," *INTENSIF J. Ilm. Penelit. dan Penerapan Teknol. Sist. Inf.*, vol. 8, no. 2, pp. 312–329, 2024, doi: https://doi.org/10.29407/intensif.v8i2.23046.

[10] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, and S. Shahab, "Deep Learning for Stock Market Prediction," *Entropy*, vol. 22, no. 840, pp. 1–23, 2020, doi: https://doi.org/10.3390/e22080840.

[11] S. J. Pipin, R. Purba, and H. Kurniawan, "Prediksi Saham Menggunakan Recurrent Neural Network (RNN-LSTM) dengan Optimasi Adaptive Moment Estimation," *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 806–815, 2023, doi: 10.47065/josyc.v4i4.4014.

[12] I. K. A. Enriko, F. N. Gustiyana, and R. H. Putra, "Komparasi Hasil Optimasi Pada Prediksi Harga Saham PT . Telkom Indonesia Menggunakan Algoritma Long Short Term Memory," *J. Media Inform. Budidarma*, vol. 7, no. April, pp. 659–667, 2023, doi: 10.30865/mib.v7i2.5822.

[13] I. K. A. Enriko, F. N. Gustiyana, and H. Krishna, "Forecasting JPFA Share Price using Long Short Term Memory Neural Network," *Jaict*, vol. 8, no. 1, p. 157, 2023, doi: 10.32497/jaict.v8i1.4285.

[14] K. Kwanda, D. E. Herwindiati, and M. D. Lauro, "Perbandingan LSTM dan Bidirectional LSTM pada Sistem Prediksi Harga Saham Berbasis Website," *J. Multidisciplinary Res. Dev.*, vol. 7, no. 1, pp. 26–35, 2024, doi: https://doi.org/10.38035/rrj.v7i1.

[15] D. I. Puteri *et al.*, "Prediksi Harga Saham Syariah menggunakan Bidirectional Long Short Term Memory (BiLSTM) dan Algoritma Grid Search," *Jambura*, vol. 6, no. 1, pp. 39–45, 2024, doi: https://doi.org/10.37905/jjom.v6i1.23297.

[16] A. Khumaidi, R. Raafi'udin, and I. P. Solihin, "Pengujian Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung," *J. Telemat.*, vol. 15, no. 1, pp. 13–18, 2020, doi: https://doi.org/10.61769/telematika.v15i1.340.

[17] A. A. Ningrum, I. Syarif, A. I. Gunawan, E. Satriyanto, and R. Muchtar, "Algoritma Deep Learning-LSTM untuk Memprediksi Umur Transformator," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 3, p. 539, 2021, doi: 10.25126/jtiik.2021834587.

[18] G. Budiprasetyo, M. Hani, and D. Zahira, "Prediksi Harga Saham Syariah Menggunakan Algoritma Long Short-Term Memory (LSTM)," *J. Nas. Teknol. dan Sist. Inf.*, vol. 08, no. 3, pp. 164–172, 2022, doi: https://doi.org/10.25077/TEKNOSI.v8i3.2022.164-172.

[19] Mushliha, "Implementasi CNN-BiLSTM untuk Prediksi Harga Saham Bank Syariah di Indonesia," *Jambura*, vol. 6, no. 2, pp. 195–203, 2024, doi: ht- tps://doi.org/10.37905/jjom.v6i2.26509 ©.

[20] D. J. Andriansyach, Sarwido, and H. Mulyo, "Short-Term Cryptocurrency Price Prediction Using Bi-LSTM Method with Interactive Web," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 4, pp. 845–850, 2024, doi: https://doi.org/10.47233/jteksis.v6i4.1645 Abstract.