JIKO (Jurnal Informatika dan Komputer) Accredited KEMDIKTISAINTEK, No.0173/C3/DT.05.00/2025
Vol. 9, No. 1, April 2026, pp. 68-73 p-ISSN: 2614-8897
DOI: 10.33387/jiko.v9i1.11102 e-ISSN: 2656-1948

A HYBRID DEEP LEARNING AND TREE BOOSTING APPROACH FOR BBCA STOCK PRICE FORECASTING WITH SHAP EXPLAINABILITY

Muhamad Sabri Ahmad[1*], H. Hadiyanto,[2] Ridwan Sanjaya[3]

[1*] Doctoral Program in Information Systems, Universitas Diponegoro, Semarang, Indonesia
[1]Faculty of Engineering, Universitas Khairun, Ternate, Indonesia
[2] Chemical Engineering Department, Faculty of Engineering, Diponegoro University,Semarang, Indonesia
[3] Department of Information Systems, Faculty of Computer Science, Soegijapranata Catholic University, Semarang, Indonesia

Email: [1]msabri@unkhair.ac.id, [2]hadiyanto@lecturer.undip.ac.id,[3]ridwan@unika.ac.id

(Received: 11 December 2025, Revised: 24 February 2026, Accepted: 23 March 2026)

Abstract

Forecasting stock price movement is a complex task due to nonlinear patterns, market volatility, and the influence of various technical and fundamental factors. This study proposes a hybrid forecasting framework that integrates the sequential learning capability of the Gated Recurrent Unit (GRU) with the nonlinear regression strength of Extreme Gradient Boosting (XGBoost) to predict the daily closing price of Bank Central Asia Tbk (BBCA). The dataset consists of historical BBCA prices from 2017 to 2025 and includes technical indicators such as moving averages, RSI, MACD, and Bollinger Bands. An 80:20 chronological split was used to evaluate model generalization through MAE, RMSE, MAPE, and R^2 metrics. Experimental results show that the hybrid GRU–XGBoost model outperforms both standalone GRU and XGBoost models, achieving the best performance with MAE of 229.09, RMSE of 312.26, and R^2 of 0.874 MAPE of 2.37%. Furthermore, SHAP-based explainability analysis highlights that price-based features and trend–momentum indicators contribute most significantly to the prediction output, while the GRU-derived sequential feature enhances temporal pattern recognition. These findings demonstrate that combining deep learning and boosting techniques produces a more accurate and interpretable forecasting model suitable for financial decision-making and risk analysis.

Keywords: *stock forecasting, GRU, XGBoost, hybrid model, SHAP, BBCA*

**Corresponding Author: Muhamad Sabri Ahmad*

1. INTRODUCTION

The stock market plays a crucial role in supporting national economic growth, where price movements of blue-chip stocks often become indicators of market stability. One of the most influential stocks in the Indonesian capital market is Bank Central Asia Tbk (BBCA), BBCA represents a blue-chip stock with high market capitalization and strong investor interest, making it an ideal case study for evaluating predictive models in a relatively stable yet dynamic financial environment.,which is known for its high liquidity and relatively stable long-term performance [1], [2], [3]. Despite its stability, BBCA's price remains highly dynamic due to macroeconomic conditions, investor sentiment, and global market volatility. These dynamics make accurate stock price forecasting an important aspect for investors, analysts, and financial institutions.

Traditional time series forecasting approaches such as the Autoregressive Integrated Moving Average (ARIMA) model have been widely used in financial analysis due to their mathematical simplicity and interpretability. However, these linear models struggle to capture nonlinear and complex patterns commonly found in stock market movements [4], [5] In recent years, machine learning and deep learning techniques have gained significant attention for financial forecasting tasks, with models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) demonstrating strong capabilities in learning temporal dependencies [5], [6] Despite these strengths, deep learning models often behave as black boxes, making it difficult to interpret their internal mechanisms and limiting their practical use in decision-critical domains such as finance [7], [8]

Tree-based ensemble methods, particularly Extreme Gradient Boosting (XGBoost), have shown strong predictive performance on structured financial datasets due to their ability to model nonlinear relationships and handle complex feature interactions efficiently [9],[10]. However, GRU alone is limited in capturing complex nonlinear relationships, while XGBoost lacks temporal sequence understanding. Therefore, combining both models is expected to leverage their complementary strengths. Another advantage of XGBoost is its compatibility with Explainable Artificial Intelligence (XAI) tools such as SHapley Additive exPlanations (SHAP), enabling transparent interpretation of model predictions [11], [12], [13]. Prior studies have explored hybrid approaches combining sequence-based models with gradient boosting techniques, reporting improved accuracy compared to single-model approaches [14], [15], [16] However, research focusing on hybrid GRU-XGBoost architectures for Indonesian stock data, especially BBCA, remains limited.

To address this gap, this study proposes a hybrid deep learning and tree boosting framework for predicting BBCA stock prices. The GRU model is utilized to extract sequential temporal information, while XGBoost serves as a meta-regressor to enhance predictive performance. Furthermore, SHAP-based XAI is integrated to provide interpretability by identifying the most influential technical indicators contributing to the model's predictions. This approach aims to combine predictive accuracy with interpretability, ensuring applicability in real-world financial decision support systems.

The objectives of this study are threefold: (1) to build GRU, XGBoost, and hybrid GRU-XGBoost models for stock price prediction; (2) to compare their predictive performance based on MAE, RMSE, and R^2 metrics; and (3) to analyze feature importance using SHAP to provide model transparency. The main contribution of this study lies in the development of a hybrid GRU–XGBoost model with integrated SHAP-based explainability for financial forecasting. The findings of this study are expected to contribute to the development of interpretable machine learning solutions for the Indonesian stock market, particularly in modeling the price movement of BBCA stock..

2. RESEARCH METHOD

This study implements a multi-stage workflow consisting of Data Collection, Data Preprocessing, Feature Engineering, Modeling, and Evaluation. The overall research workflow is illustrated in Fig. 1 and each stage is described in detail in the following subsections.

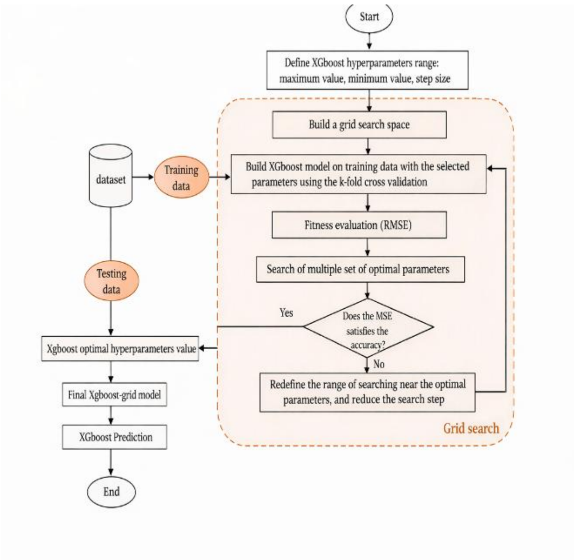

Figure 1. Hybrid GRU-XGBoost Research Workflow

2.1 Data Collection

The dataset used in this research consists of historical daily stock prices of Bank Central Asia Tbk (BBCA) from January 2017 to January 2025. The data includes Open, High, Low, Close, and Volume attributes, obtained from publicly accessible financial data sources such as Yahoo Finance and MarketWatch. Daily closing prices were selected as the target variable because of their relevance in stock forecasting tasks. All records were inspected to ensure consistency and completeness before further processing.

2.2 Data Preprocessing

Preprocessing was performed to ensure that the dataset is clean, consistent, and suitable for the modeling stage. The following steps were applied:

1. Date Formatting - Converting the date attribute into a uniform datetime format and sorting the records chronologically.
2. Numeric Cleaning - Removing comma delimiters and converting all price and volume fields into numeric values.
3. Target Generation - Creating the prediction target by shifting the closing price by one day.
4. Missing Value Handling - Removing rows affected by undefined technical indicators and the final row with an undefined target.
5. Scaling for GRU - Applying MinMaxScaler to normalize features for GRU input, using only the training portion of the dataset to avoid data leakage.
6. Sequence Construction - Transforming the normalized dataset into sequential windows (20-60 timesteps) required for GRU-based time series modeling.

2.3 Feature Engineering

Technical indicators commonly used in financial analysis were generated to enhance predictive capability. The indicators computed in this study include:

- Moving Average (MA5, MA10, MA20)
- Exponential Moving Average (EMA10)
- Relative Strength Index (RSI14)
- MACD and Signal Line
- Bollinger Bands (Upper, Middle, Lower)
- Daily Return

These features were selected to provide trend, momentum, and volatility information, offering complementary perspectives to the raw price data.

2.4 Modeling Framework

Three predictive models were developed and compared: a GRU model, an XGBoost model, and a hybrid GRU-XGBoost model.

2.4.1 GRU Model

The GRU model was built to capture temporal dependencies in the sequential stock data. The architecture consists of:

- A GRU layer with 64 units
- A second GRU layer with 32 units
- Dropout regularization
- Dense layers for regression output

The model was trained using the Adam optimizer and mean squared error (MSE) loss function. Early stopping was applied to prevent overfitting by monitoring validation loss, The validation set was obtained by splitting the training data into 90% training and 10% validation.

2.4.2 XGBoost Model

XGBoost was implemented as a tabular-based baseline model. It was trained using all engineered technical features and lagged values. The key hyperparameters used include, These hyperparameters were determined through empirical tuning and iterative experimentation.:

- n_estimators = 500
- max_depth = 6
- learning_rate = 0.05
- subsample = 0.8
- colsample_bytree = 0.8

The model was optimized to capture nonlinear interactions and complex feature relationships.

2.4.3 Hybrid GRU-XGBoost Model

The hybrid model integrates sequential learning from the GRU model with the nonlinear tabular modeling capability of XGBoost. The GRU predictions are appended to the original feature set and used as additional inputs for an XGBoost meta-regressor. The hybrid model integrates, This stacked ensemble approach enables the hybrid model to utilize both temporal patterns and engineered feature interactions.

2.5 Evaluation Metrics

Model performance was measured using several regression metrics:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- Coefficient of Determination (R^2)

The dataset was split chronologically into 80% for training and 20% for testing to simulate real-world forecasting conditions. The evaluation compares the GRU, XGBoost, and hybrid models using identical train-test partitions. In addition, SHAP-based Explainable AI analysis was applied to the hybrid model to identify dominant features contributing to stock price predictions.

3. RESULT AND DISCUSSION

This section presents the experimental results and discusses the performance of the GRU, XGBoost, and Hybrid GRU-XGBoost models in predicting the BBCA stock closing price. The evaluation focuses on accuracy, generalization ability, and interpretability, followed by feature importance analysis using SHAP.

3. 1 Data Overview

The final dataset consisted of daily BBCA stock prices from 2017 to 2025 after preprocessing and feature construction. Several technical indicators-such as MA, EMA, RSI, MACD, and Bollinger Bands-were successfully computed and aligned with the raw OHLCV features. The data distribution showed stable long-term growth with occasional volatility spikes, reflecting macroeconomic fluctuations affecting the Indonesian banking sector.

3. 2 Model Performance Evaluation

Three models were evaluated: GRU, XGBoost, and the Hybrid GRU-XGBoost model. Performance was measured using MAE, RMSE, MAPE, and R^2 on the test set. Table 1 summarizes the results.

Table 1. Model Performance Comparison on Test Data

Model	Test MAE	Test RMSE	Test MAPE	Test R^2
GRU	1243.01	1469.28	12.69%	-1.78
XGBoost	238.18	325.87	2.46%	0.862
Hybrid GRU+XGBoost	229.09	312.26	2.37%	0.874

The GRU model produced the lowest performance, indicating difficulty in generalizing temporal patterns over long horizons. This is consistent with the behavior of recurrent neural networks when trained on volatile financial data with nonlinear fluctuations.

In contrast, XGBoost achieved substantially lower MAE and RMSE, making it the strongest standalone baseline. Its ability to model nonlinear interactions among engineered technical indicators contributed to its predictive stability.

The Hybrid GRU-XGBoost model outperformed both individual models across all evaluation metrics on the test set. Although the improvement over XGBoost is modest (MAE reduced by 3.81%, RMSE reduced by 4.18%), the hybrid strategy shows consistent enhancement and maintains the highest R^2 value (0.874). This indicates that incorporating sequential information from GRU predictions helped XGBoost refine the regression output.

These results are similar to patterns observed in ensemble-based predictive studies where hybrid models outperform single architectures by combining complementary strengths.

3. 3 Actual vs Predicted Visualization

Figure 2 illustrates a comparison between actual BBCA closing prices and model predictions on the test set. XGBoost and the hybrid model show close alignment with real price movements, while the GRU model exhibits larger deviations.

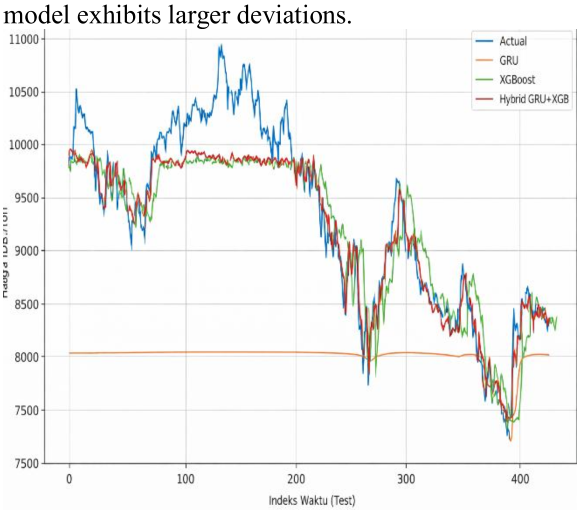

Figure 2. Actual vs Predicted Stock Prices on Test Data

This improvement indicates that the hybrid model effectively integrates temporal and nonlinear feature representations.

The hybrid model not only tracks the directional movement more closely but also reduces error during volatility peaks, demonstrating improved robustness.

3. 4 SHAP-Based Explainability Analysis

Explainability is critical in financial forecasting, where model transparency aids investor trust and ensures that predictions align with logical market behavior. SHAP analysis was applied to the hybrid model to identify key contributing features.

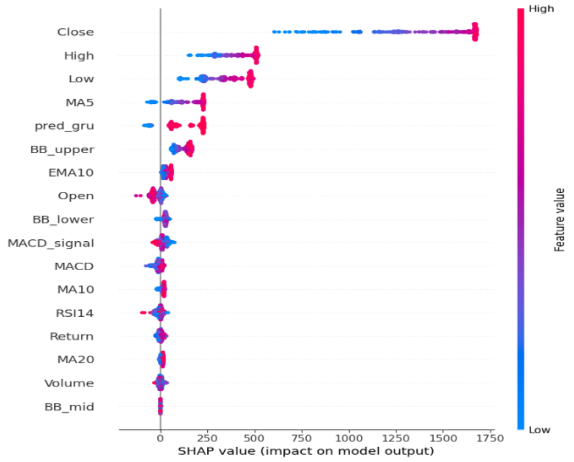

Figure 3. Presents the SHAP summary plot showing the most influential features.

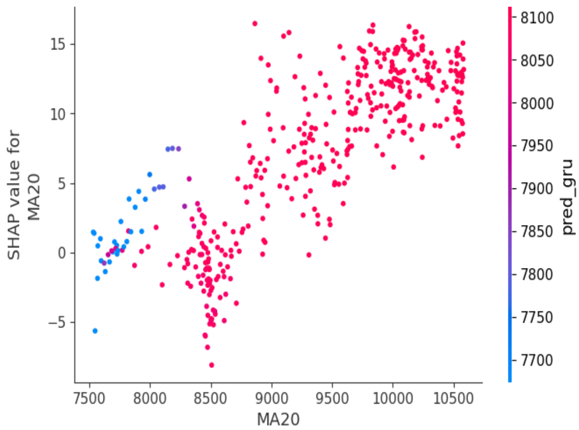

Figure 4. SHAP Summary Plot of Feature Contributions

Figure 3 presents the SHAP dependence analysis for the MA20 feature, illustrating its marginal effect on the hybrid model's output. The plot reveals a strong positive monotonic relationship, where higher MA20 values correspond to increasingly positive SHAP contributions, indicating upward pressure on the predicted closing price. The color gradient representing *pred_gru* highlights the interaction between medium-term trend signals and sequence-derived information. Observations with higher GRU-based predictions consistently exhibit larger SHAP values, confirming that the hybrid model leverages temporal dynamics to refine trend-based forecasting behavior

Figure 4 displays the global SHAP summary plot, ranking all input features by their overall contribution to the hybrid GRU-XGBoost model. Price-based variables (Close, High, Low) dominate the importance spectrum, reflecting their central role in capturing short-term market movements. Technical indicators such as MA5, RSI14, MACD, and Bollinger Bands provide additional nonlinear signals related to momentum and volatility. The pred_gru feature exhibits meaningful positive impact, demonstrating that sequential representations extracted by the GRU module enhance the model's regression capability. The SHAP value dispersion across features indicates stable and interpretable model behavior, with feature contributions aligning with established financial reasoning.

The top features influencing predictions include:

1. MA20 and MA10 - Medium-term trend indicators significantly shaped the stock price estimation.
2. RSI14 - Momentum changes and overbought/oversold signals impacted predicted movements.
3. MACD and MACD Signal - Captured trend reversals and momentum divergence.
4. GRU Prediction Feature - Contributed positively by providing sequence-aware information.
5. Bollinger Bands - Reflected volatility levels and relative price positioning.

The presence of both GRU-based and technical indicators in the top contributors supports the effectiveness of hybrid modeling. SHAP dependence plots further show how increases or decreases in specific indicators correlate with predicted price movement direction.

3. 5 Discussion

The results demonstrate that the hybrid GRU-XGBoost architecture offers meaningful advantages over individual models. The GRU model alone was insufficient due to difficulty in capturing nonlinear temporal patterns, while XGBoost provided strong tabular modeling capabilities but lacked sequential context. Combining both produced improved accuracy while maintaining interpretability.

The SHAP analysis confirmed that the hybrid model relies on economically meaningful indicators, aligning with real-world financial reasoning. This enhances the model's credibility for practical forecasting applications.

4. CONCLUSION

This study proposed a hybrid forecasting framework integrating GRU-based sequential learning with XGBoost regression to predict the daily closing price of BBCA stock. Three models—GRU, XGBoost, and the hybrid GRU–XGBoost—were evaluated using MAE, RMSE, MAPE, and R^2 metrics on an 80:20 chronological split. The results demonstrate that the standalone GRU model struggled to generalize on volatile financial time series, while XGBoost delivered substantially stronger predictive performance on engineered technical indicators. The hybrid model achieved the best overall results, yielding improvements over XGBoost across all evaluation metrics and exhibiting enhanced robustness in capturing trend and momentum fluctuations.

The SHAP-based explainability analysis further revealed that both price-based features (Close, High, Low) and technical indicators (MA5, RSI14, MACD, and Bollinger Bands) play significant roles in shaping model predictions. The GRU-derived sequential feature also contributed meaningfully to the hybrid model's performance, indicating the effectiveness of combining temporal and nonlinear tabular representations. These findings confirm that the proposed hybrid approach provides a balanced integration of accuracy and interpretability, making it suitable for practical financial forecasting applications.

Future work may extend this study by incorporating additional data sources such as macroeconomic indicators, multi-stock relationships, or sentiment features derived from financial news and social media. Exploring alternative deep learning architectures or transformer-based sequence models may also enhance predictive stability under high-volatility market conditions.

5. REFERENCE

[1] M. A. Rad, L. M. Lefsrud, M. T. Hendry, A. C. Cen, and S. Soltaninejad, "Analysis of freight train passing a stop signal using machine learning: Application of XGBoost and SHAP," *J. Rail Transp. Plan. Manag.*, vol. 35, no. April, p. 100532, 2025, doi: 10.1016/j.jrtpm.2025.100532.

[2] B. Gülmez, "Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm," *Expert Syst. Appl.*, vol. 227, no. May, p. 120346, 2023, doi: 10.1016/j.eswa.2023.120346.

[3] J. Che, "A Short-Term Hybrid Prediction Model for Offshore Wind Power Based on EEMD and LSTM," *2023 6th Int. Conf. Energy, Electr. Power Eng.*, pp. 1225–1229, 2023, doi: 10.1109/CEEPE58418.2023.10167250.

[4] Q. Liu, Y. Hu, and H. Liu, "Enhanced stock price prediction with optimized ensemble modeling using multi-source heterogeneous data: Integrating LSTM attention mechanism and multidimensional gray model," *J. Ind. Inf. Integr.*, vol. 42, no. November, p. 100711, 2024, doi: 10.1016/j.jii.2024.100711.

[5] C. Ravikumar, J. Manoranjini, S. Pallavi, U. Naresh, S. Telang, and S. Kiran, "Advancing Malware Detection Using Memory Analysis and Explainable AI Approach," *2nd Int. Conf. Intell. Cyber Phys. Syst. Internet Things, ICoICI 2024 - Proc.*, no. ICoICI, pp. 518–523, 2024, doi: 10.1109/ICoICI62503.2024.10696406.

[6] S. Hossain and G. Kaur, "Stock Market Prediction: XGBoost and LSTM Comparative Analysis," *2024 3rd Int. Conf. Artif. Intell. Internet Things, AIIoT 2024*, no. AIIoT, pp. 1–6, 2024, doi: 10.1109/AIIoT58432.2024.10574794.

[7] H. E. Yu, L. I. J. Ming, R. Sumei, and Z. Shuping, "A Hybrid Model for Financial Time Series Forecasting — Integration of EWT , ARIMA With The Improved ABC Optimized ELM," pp. 84501–84518, 2020, doi: 10.1109/ACCESS.2020.2987547.

[8] J. Timko, R. El Shawi, and S. Tomasiello, "Optimizing stock price forecasting: a hybrid approach using fuzziness and automated machine

learning," *Expert Syst. Appl.*, vol. 295, no. December 2024, 2024, doi: 10.1016/j.eswa.2025.128844.

[9] Z. Liu, Z. Su, L. Shang, H. Sun, and B. Zhao, "An approach to stock price prediction based on improved EEMD and attention-enhanced BiLSTM," *Expert Syst. Appl.*, vol. 284, p. 127802, 2025, doi: 10.1016/j.eswa.2025.127802.

[10] H. Notaria, S. Shah, D. Thopte, H. Soneji, P. Bari, and K. Deulkar, "Comparative Analysis of Stock Price Prediction using Time Series Models," *2024 8th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2024*, pp. 1–6, 2024, doi: 10.1109/ICCUBEA61740.2024.10775112.

[11] Z. Xiao, S. Li, and Z. Lin, "A Hybrid Modeling Approach for Forecasting the Volatility of REITs Index in US Market," *2012 Int. Conf. Manag. Sci. Eng. 19th Annu. Conf. Proc.*, pp. 1861–1867, 2012, doi: 10.1109/ICMSE.2012.6414425.

[12] M. R. Hossain, M. T. Ismail, and J. Hossain, "Stock Price Prediction Using Empirical Mode Decomposition Based Theta Method and Forecast Combination," *2021 Int. Conf. Decis. Aid Sci. Appl.*, pp. 1115–1119, 2021, doi: 10.1109/DASA53625.2021.9682235.

[13] L. Heistrene *et al.*, "An Improved and Explainable Electricity Price Forecasting Model via SHAP-Based Error Compensation Approach," *IEEE Trans. Artif. Intell.*, vol. 6, no. 1, pp. 159–168, 2025, doi: 10.1109/TAI.2024.3455313.

[14] T. D. Hastuti, R. Sanjaya, and F. Koeswoyo, "The Investment Opportunity, Information Technology and Financial Performance of SMEs," *Proc. - Int. Conf. Comput. Inf. Sci. Sustain. Tomorrow with Digit. Innov. ICCOINS 2021*, no. July 2021, pp. 247–251, 2021, doi: 10.1109/ICCOINS49721.2021.9497182.

[15] P. Thanathamathee, S. Sawangarreerak, and D. N. M. Nizam, "Enhancing Going Concern Prediction With Anchor Explainable AI and Attention-Weighted XGBoost," *IEEE Access*, vol. 12, no. March, pp. 68345–68363, 2024, doi: 10.1109/ACCESS.2024.3401007.

[16] M. Aravindhan and D. Manikavelan, "Decomposition-Based Transformer-LSTM Framework for Stock Index Forecasting," *2025 3rd World Conf. Commun. & Comput.*, pp. 1–6, 2025, doi: 10.1109/WCONF64849.2025.11233358.