Implementation of X-Gradient Boosting in Banking Stock Price Predictions

Amelia Maharani Fatihah[1], Komang Dharmawan[2,*] Putu Veri Swastika[3]

[123] Udayana University, Bali, Indonesia
*k.dharmawan@unud.ac.id

Abstract. The abstract should summarize the contents of the paper in short terms, i.e. 150-250 words. Accurate and computationally efficient prediction of stock price or stock return or any financial problem is required to gives ad- ditional point of view in investment. Machine learning methods have recently become a part of financial model prediction due to their robustness in dealing with large and complex data. In this paper, an extreme gradient boosting (XGBoost)-based machine learning method is introduced for predicting the stock return of five financial banking institution in Indonesia. The data were collected from January 2021 to November 2023. The technical indicator are investigated for each cases. The hyperparameter tuning through the grid search approach was performed to obtain an optimized XGBoost model. The performance of the XGBoost method is compared to the actual data. The validation evaluation results demonstrate that the XGBoost algorithm is highly accurate with the error less than one percent. This demonstrates the machine learning performance in the financial realm as a useful tools to help investors implement better prediction strategies

Keywords: XGBoost, Machine learning, Stock predictions, Banking

1 Introduction

Investment activities are gaining popularity in Indonesia. Investors, both local and foreign, are increasingly drawn to the capital market. Investment refers to committing resources for future profit. Investors are individuals who make investments. In general, there are two main types of investments: real assets and financial assets. Real asset investment involves allocating capital to physical assets like land, gold, and buildings. Investing in financial assets involves allocating funds to securities like shares, land certificates, and deposits.

In stock investment, every investor hopes to achieve a high rate of return. Stock return refers to the profit level obtained by investors. In investment management, returns are divided into two, namely expected returns and realized returns. Investors expect a return, while realized return is what is obtained. Investors must not only focus on returns, but also on the risks involved in investing. Risk refers to the variance between the realized return and the expected return. It can be classified into two types: systematic risk and market risk. Investing in individual assets comes with a certain level of risk, which is essentially the market risk of a well-diversified portfolio. Therefore, to minimize the risks associated with investing in shares, it is advisable to spread the funds across multiple shares instead of just one

According to Saputra and Saepudin [1], a portfolio is a collection of different assets in the form of shares that are strategically combined to achieve the best possible results. The objective of the portfolio is to minimize risk and maximize returns. In order to create an optimal portfolio, an efficient portfolio must be formed first. An efficient portfolio is a combination of assets that provides low risk at a certain return or high return at a certain risk. The optimal portfolio is one that balances expected return with risk while also considering the characteristics of the assets included. The sachets in question have certain characteristics that determine their expected returns. The selection of shares made by investors can impact the performance of their portfolio. To group shares with similar characteristics, we will be utilizing the XGBoost algorithm.

XGBoost is a machine learning method that uses a decision tree-based en- semble learning model. It is an effective algorithm for optimizing weak trees (classifiers) to improve the system's learning performance. The eXtreme Gradi- ent Boosting (XGBoost) method is a development of gradient boosting proposed by Dr. Tianqi Chen from the University of Washington in 2014. Gradient Boost- ing is an algorithm used to find optimal solution for various problems, especially in regression, classification and ranking [2].

Several studies have examined utilizing stock return predictions to choose shares in creating an optimal portfolio [1] optimized portfolios for shares listed on LQ45, based on stock return predictions using

K. Dharmawan and N. A. Sanjaya ER (eds.), *Proceedings of the First International Conference on Applied Mathematics, Statistics, and Computing (ICAMSAC 2023)*, Advances in Computer Science Research 110,
https://doi.org/10.2991/978-94-6463-413-6_17

Hybrid XGBoost and the Imported Firefly Algorithm. The performance of a portfolio of 7 shares was evaluated with the help of predictions. The results showed that the portfolio performed better than a portfolio without considering predictions. The portfolio had a mean return of 0.0029, standard deviation of 0.0158, and Sharpe ratio of 0.1837, which are all good indicators of its performance.

Jange [3] conducted a study on technical indicators used to predict BCA bank's share prices with XGBoost. The study examined four technical indi- cators: Simple Moving Average (SMA), Exponential Moving Average (EMA), Moving Average Convergence/Divergence (MACD), and Relative Strength Index (RSI). According to the results, Exponential Moving Average (EMA) is the most effective technical indicator with a MAPE value of 4.01%.

Stock price prediction is a crucial aspect of financial decision-making, and the XGBoost algorithm has been shown to be effective in this regard. Yun et al. [4] and Ansary [5] both emphasize the importance of feature engineering in stock price prediction, with the latter highlighting the significance of volume as a predictive feature. [6] and Kumar et al. [7] both demonstrate the effectiveness of the XGBoost algorithm in stock price prediction, with the former using a hybrid LSTM-BO-XGBoost model and the latter using a SARIMA-XGBoost hybrid model. [8] and Gumelar et al. [9] both show that the XGBoost algorithm, when combined with other models, can improve prediction performance. Vuong et al. [10] further enhances the XGBoost algorithm's performance by using it as a feature-selection technique in combination with a deep LSTM network. Basak et al. [11] also highlights the effectiveness of tree-based classifiers, including XGBoost, in predicting the direction of stock market prices.

Thus in this study, we investigate the peformance of XGBoost to predict several banking stock returns. In Section 2, the xgboost literature are reviewed while the description of its algorithm are considered in Section 3. Case studies and simulations on various banks are given Section 4. Conclusions are discussed in the last Section 5.

2 Literature Review

A range of machine learning techniques have been applied to stock price pre- diction, with promising results. Wang and Bai [12] found that Logistic Regres- sion and Support Vector Machines were effective in predicting stock movements, with the latter yielding the best results. Similarly, Vijh et al. [13] used Artificial Neural Network and Random Forest to predict closing prices, achieving high efficiency. Pawaskar [14] also explored the use of machine learning algorithms, in- cluding Multiple Linear Regression and Polynomial Regression, to predict stock prices. Chen [15] further improved prediction accuracy by using Long Short-Term Memory, Convolutional Neural Networks, and Support Vector Regression. Nikou et al. [16] found that Long Short-Term Memory and Deep Learning algo- rithms outperformed other methods in predicting stock values. Jeevan et al. [17] also demonstrated the effectiveness of Linear Regression, Random Forest, and Multilayer Perceptron in stock prediction.

A range of studies have explored the use of XGBoost in predicting stock price returns, with promising results. Yang et al. [8] and Basak et al. [11] both found that XGBoost outperformed other models, while Ansary [5] highlighted the im- portance of technical indicators and internet search query volumes in improving prediction accuracy. Vuong et al. [10] and Liwei et al. [6] further enhanced XG- Boost's performance through feature engineering and Bayesian optimization, respectively. Wang and Bai [12] introduced a boosting-ANN model for stock price forecasting, which also showed promising results. These studies collectively suggest that XGBoost, particularly when combined with other techniques, can be a valuable tool in predicting stock price returns.

3 Methods

This section describes the XGboost-based machine learning method, including its framework to predict stock return

3.1 Return

Stock return is the level of profit obtained by investors. The main components in the source of return are yield and capital gain (loss). Yield is a reflection of the path of income generated periodically in investing. Capital gain (loss) is a component of adding or depreciating the quality of securities, either long-term debt securities or stocks, which results in profit or loss by investors. The return value can be calculated with the following equation

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{1}$$

with R_t is the return value at the time t, P_t is the stock price at the time t, and P_{t-1} is the stock price at the time $t - 1$.

3.2 XGBoost

XGboosting is a machine learning method with a decision tree-based ensemble learning model, which is an algorithm to effectively optimize weak trees (classi- fiers) so as to improve system learning performance.These classifiers are formed with the aim that each new classifier focuses on learning the weaknesses (mis- classified data) of the previous classifier [1]. Based on ensemble learning, the new classifier or tree experience a gradient increase based on the following equation

$$\widehat{yl} = \sum_{t=1}^{K} f_t(x_i) \ , \qquad f_t \in F \tag{2}$$

where f_t is a $t - th$ tree in the regression tree space F and K is the number of trees. The objective function $\mathcal{L}(j)$ to be minimized in XGBoost method is given by the following equation

$$\mathcal{L}(j) = \sum_{i=1} \iota(y_i, \hat{y}_i^t) + f_t(x_i) + \sum_{t=1}^{K} \Omega(f_t) \tag{3}$$

where ι is the loss function and $\Omega(f_t)$ denotes the regularization. The loss function can be logistic loss or square loss, which represents the degree of similarity between the training set and the model. Regularization is a term to control the complexity of the model and usually controls overfitting. The regularization is the complexity equation given by the following

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \left|\left|\omega\right|\right|^2 \tag{4}$$

where γ, λ denotes the L1 regularization coefficient, and L2 regularization coefficient and T, ω denotes the number of nodes, and node weight. The main goals of the XGBoost algorithm is to minimize the objective function (3), which consists of the loss function l and regularization terms (4). The hyperparameters such as T , γ, and ω are given firsthand but, the parameter ft in Eqn (4) must be calculated during the training phase. Several researchers discovered that hyper- parameter variables significantly affect model accuracy [18]. Thus, ideal hyperparameters, which may be discovered by hyperparameter tuning, are necessary. The three most often used approaches for optimizing hyperparameters are grid search, random search, and Bayesian optimization. Because of its simplicity, we used the grid search strategy in this investigation.

4 Result and Discussion

In this section, we discuss the results of the XGBoost method for predicting stock returns as well as the in formation of market through technical indicator analysis.

4.1 Technical Indicator

To begin with, First, in Figure 1 we provide a plot of the stock price of five big financial banks from January 02 2021 until November 07 2023. The five large banks are Bank Rakyat Indonesia (code: BBRI), Bank Central Asia (code: BBCA), Bank Negara Indonesia (code: BBNI), Bank Mandiri (code: BMRI) and Bank Syariah Indonesia (code: BRIS). From the plot in Figure 1 , it can be seen that Bank Central Asia, has the largest price among the four other banks. Bank Rakyat Indonesia occupies the second posi

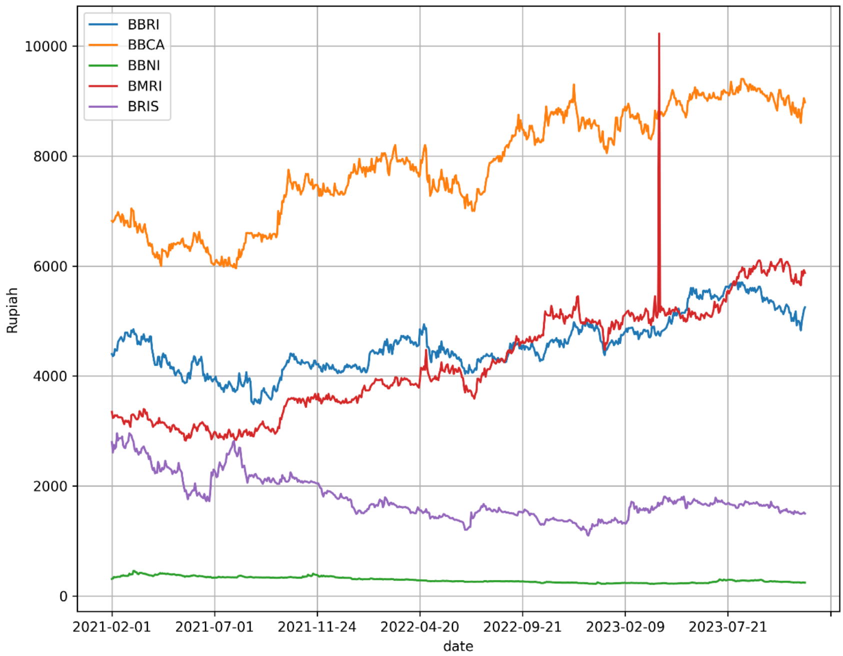

Fig. 1. Plot of stock price of five major banks from January 02 2021 until November 07 2023

tion with the largest prices. However, in the 2022-2023 period, the stock price from Bank Mandiri increased rapidly and shifted BRI's position. Meanwhile, the fourth and the fifth position respectively is occupied by Bank Syariah Indonesia and Bank Negara Indonesia whose stock prices have visibly decreased. For clarity, the chart trend consist of: trend, seasonality and residuals for BRIS and BBNI are given in Figure 2.

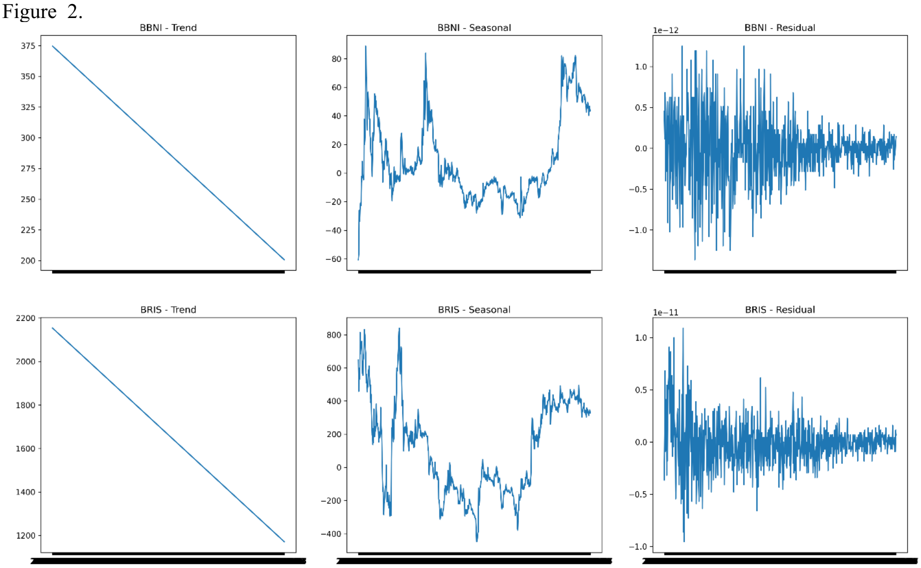

Fig. 2. Decomposition Chart (Trend, Seasonal, Residual) of BNI and BRIS stock prices respectively

Figure 2 illustrates the decomposition of BBNI and BRIS stock prices respec- tively taken based on the decomposition function at daily closing. The function produces a decomposition chart consisting of trend, seasonal and residual plots. Clearly, it can be seen that the BBNI and BRIS stock prices trend tends to de- crease. Meanwhile, the stock price of BBCA and BBRI are tends to increase as plotted in Figure 3. The seasonal plot shows that stock prices fluctuate throughout the year, while the residual stock price plot is very small with a precision close to zero.

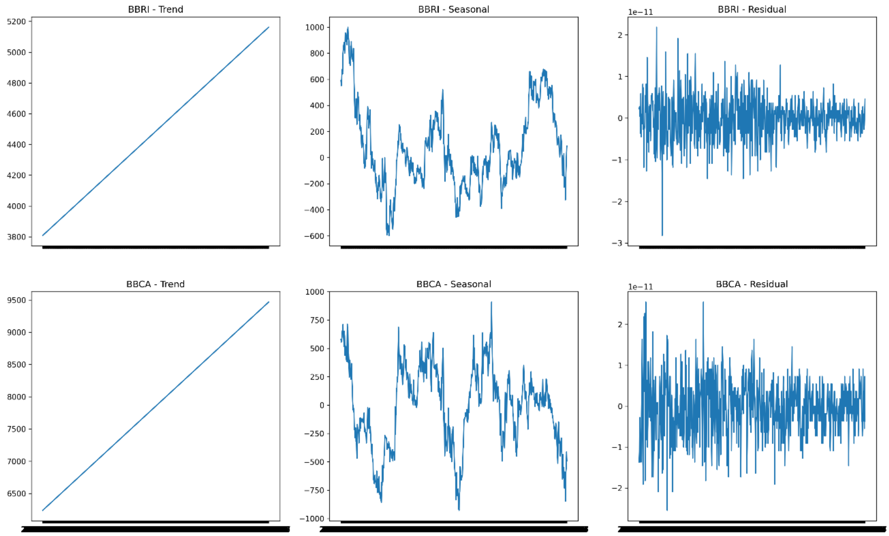

Fig. 3. Decomposition Chart (Trend, Seasonal, Residual) of BBRI and BMRI stock prices respectively

Before we determined the indicator, first we calculate the return of each stock and plotted as in Figure 4 (right) for each stock. The blue curve denoted Bank Rakyat Indonesia’s return or R1, the orange curve denoted Bank Central Asia’s return or R2, the green curve denoted Bank Negara Indonesia’s return or R3, the fourth position with red curve denoted Bank Mandiri’s return or R4 and the last with purple curve denoted Bank Syariah Indonesia’s return or R5.

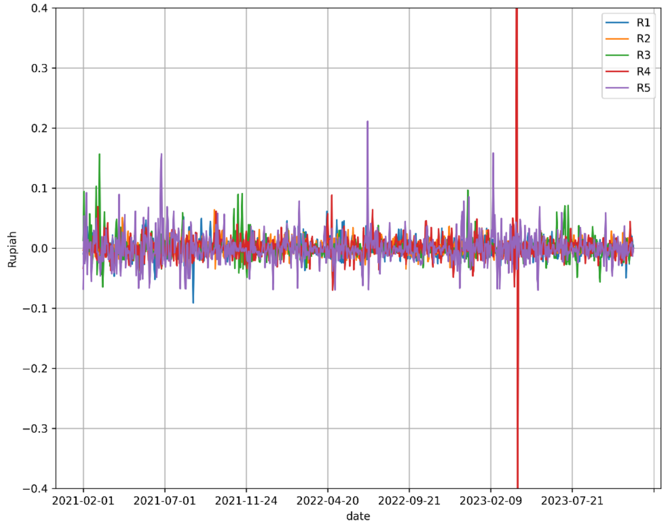

Fig. 4. Plot of stock return of five major banks from January 02 2021 until November 07 2023

Now, to calculate the stock return prediction model, it is necessary to carry out feature engineering using technical indicators as shown in Table 4.1. The indicators used here such as EMA (Exponential Moving Average), SMA (Simple Moving Average), RSI (Relative Strength Index) and MACD (Moving Average Convergence Divergence).

Technical Indicator	Description
EMA 9	The nine days of stock return data moving weighted
SMA 5	Data on 5, 15, and 30 days of short-term average stock prices movements
SMA 15	
SMA 30	
RSI	The momentum indicator used to evaluates the speed and size of recent price fluctuations
MACD	EMA 12 - EMA 26
MACD SIGNAL	EWM index weighted moving average for MACD with a time span of 9 days

Table 1. Technical Indicator for feature enginerring with its description

Moving averages indicate a stock's trend direction or support and resistance levels. In this study, we investigate using both simple moving averages and exponential moving averages. A simple moving average (SMA) is calculated by taking the arithmetic mean of a set of values over a specific time period. We utilized the five, fifteen, and thirty-day periods, abbreviated SMA 5, 15, and 30, respectively. Meanwhile, the Exponential Moving Average (EMA) performs precisely what the name implies: it is exponential, with the more recent values weighted higher than the less recent data. The EMA used here is EMA 9, which is the average of the last nine closing prices. It provides short-term continuation and reversal trading signals. The EMA can be compared and contrasted with the simple moving average. Another important characteristic is the relative strength index (RSI). The RSI is a momentum indicator used in technical analysis. It evaluates the speed and size of recent price fluctuations in an investment to determine if it is overpriced or undervalued. The RSI is shown as an oscillator (a line graph) with a range of zero to 100

Fig. 5. Technical Indicator Chart (EMA and SMA) of BCA Bank Share Returns from January 02 2021 until November 07 2023

As an illustration, we will only calculate technical indicators for BBCA stock returns. For the other stock returns, calculations can be done in the same way. The following is a visualization of feature engineering using technical indicators on BCA stock returns from January 02 2021 until November 07 2023 as shown in the Figure 5. The technical indicators displayed are the EMA 9, SMA (5,15 and 30) together with BBCA stock returns. The blue line represent EMA 9 indicator whereas the red, green and purple lines represents SMA 15, SMA 15 and SMA 30 resprectively as well as the light blue line represent stock return for BBCA.

Once calculated, the RSI indicator of BBCA is plotted beneath the asset price chart, as seen in Figure 6. In general, when the RSI indicator crosses 30, it is a positive indication, while crossing 70 is a negative sign. In other words, an RSI reading of 70 or higher suggests that an investment is becoming overbought or overpriced. An RSI reading of 30 or lower implies an oversold or undervalued state. It can be infereed from RSI index in Figure 6, that BBCA asset rarely hit overpriced and underpriced.

Fig. 6. Relative Strength Index (RSI) of BCA share Returns from January 02 2021 until November 07 2023

Moving average convergence divergence (MACD) is another technical indica- tor that may help analysts and investors get a more complete view of a market. These trend-following momentum indicators display the connection between two moving averages of a security's price. The MACD is computed by subtracting the 26-period EMA from the 12-period EMA. The outcome of this computation is the MACD line. Figure 7 depicts a nine-day MACD EMA, known as the signal line, placed on top of the BBCA return's MACD line. It may be used to trigger buy and sell signals. When the MACD crosses above the signal line, an investor may purchase the asset; when the MACD crosses below the signal line, the in- vestor may sell or short the investment. The MACD measures the connection between two EMAs, whereas the RSI monitors price momentum in respect to previous highs and lows.

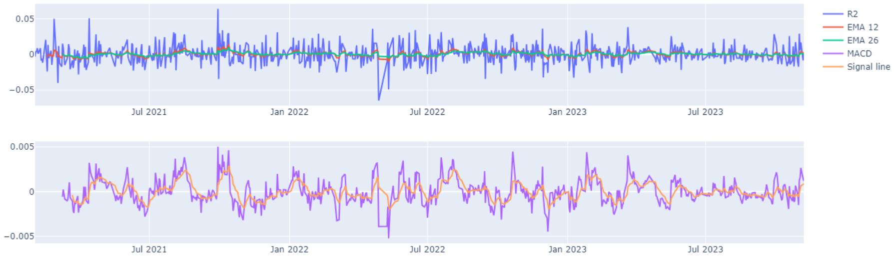

Fig. 7. MACD technical indicator of BBCA stock return

4.2 XGBoost Implementation

To begin with, After the discussion of technical indicators, next we will use the XGBoost Algorithm to predict the stock returns. The initial stage in using XGBoost is to divide the data into two categories: training data and test data.

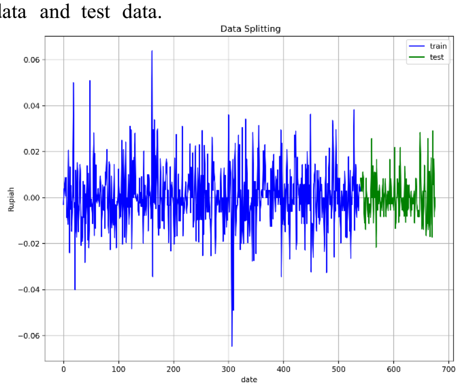

Fig. 8. Splitting Data of BBCA return

Typically, the appropriate ratio for training to test data is 70:30 or 80:20. Approximately 70% or 80% of stock return data is used for training, while the rest is used for testing. The ratio varies with dataset size. However, if training data is less than 70%, the XGBoost model may be undertrained. This suggests that the generated model may lack sufficient data to learn from, rendering it unable to effectively predict the phenomenon [18]. Here, in this study, we used 80% of the allocated data as training data to train the system to produce all prediction models and 20% of the data is used as test data to test the final model obtained after all training has been carried out. The splitting plot are given in Figure 8.

For the next step, the XGBoost method needs hyperparameters to be defined before the training process begins. These hyperparameters have a substantial influence on the performance of XGBoost. As a result, we use hyperparameter tuning using grid search to determine the best values for the hyperparameters, as indicated in Table 4.2. Grids are created by listing all potential hyperparameter combinations in the search interval. Using the training dataset, the objective function in Eq (3) evaluates the error in each grid. The hyperparameters with the lowest error are considered optimal.

Parameter	Original					After Tuning				
	BBRI	BBCA	BMRI	BRIS	BBNI	BBRI	BBCA	BMRI	BRIS	BBNI
n estimators	100	100	100	100	100	300	190	300	300	290
max depth	3	3	3	3	3	3	4	2	5	3
learning rate	0.1	0.1	0.1	0.1	0.1	0.1	0.05	0.1	0.1	0.1
min child weight	1	1	1	1	1	5	6	5	5	5
subsample	1	1	1	1	1	0.7	1	0.9	0.9	0.7
colsample bytree	1	1	1	1	1	0.5	0.5	0.5	0.5	0.5
colsample bylevel	1	1	1	1	1	0.5	0.5	0.5	0.5	0.5
gamma	0	0	0	0	0	0.1	0.4	0.1	0.1	0.1
RMSE	0.2	0.002	0.065	0.003	0.003	0.03	0.03	0.07	0.007	0.003

Table 2. Hyperparameters used in the XGBoost method.

Next, the actual data and the prediction curve of share return of Bank Central Asia are shown in Figure 9. It can be seen that the prediction curve agrees well with the actual data, indicating the accuration of the model. Meanwhile, the prediction results for the other four banks are presented simultaneously in the Figure 10. The stock return predictions for each bank have quite good accuracy with eror less than 1%.

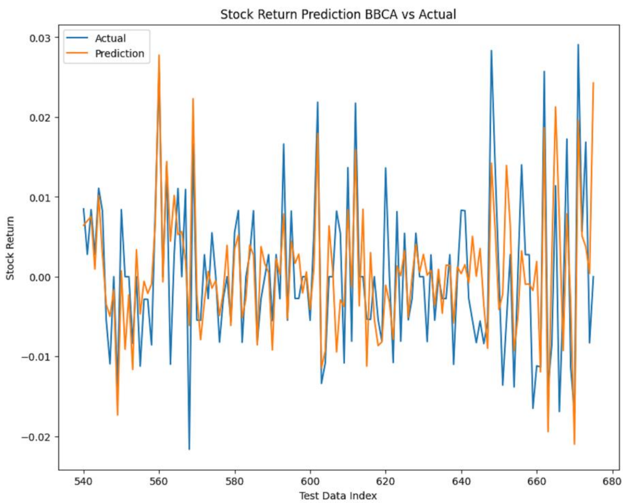

Fig. 9. BBCA Stock return prediction graph

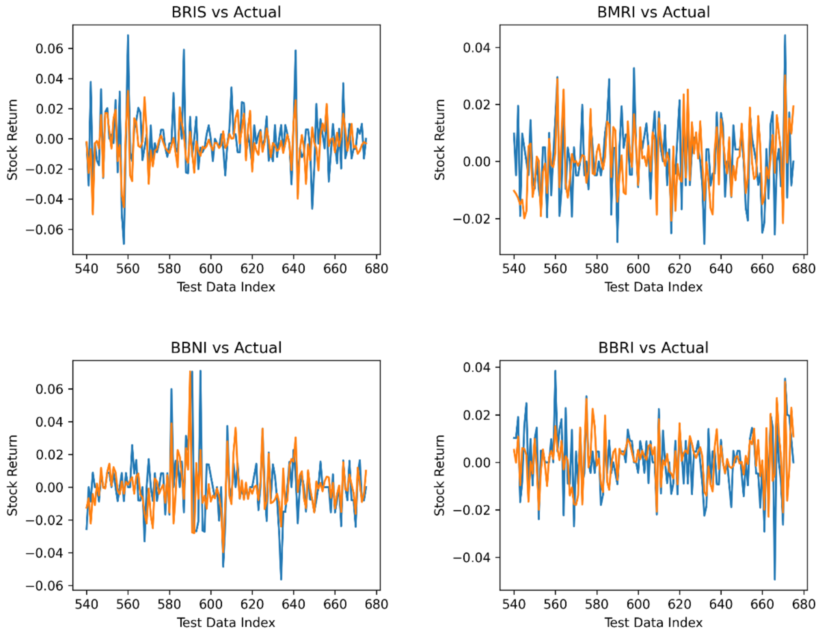

Fig. 10. Stock return prediction vs actual for BBRI, BMRI, BRIS and BBNI

5 Conclusion

The eXtreme Gradient Boosting (XGBoost) machine learning technology has been successfully used to estimate the stock returns of five major Indonesian financial institutions from 2021 to 2023. The grid search method proved success- ful in refining hyperparameters and producing a highly accurate XGBoost model with less than 1% inaccuracy. Apart from that, we were successful in discover- ing technical indicators for each case. This demonstrates the XGBoost model's high performance in the financial realm. Further studies may be conducted to determine the optimal portfolio that can be built as an alternate point of view utilizing machine learning methods.

Acknowledgement: This research was supported by Program Unggulan Udayana (PUU) 2023 with contract number [No. B/1.612/UN14.4.A/PT.01.03/2023].

References

[1] M. R. Saputra, D. Saepudin, Optimasi Portofolio Berbasis Prediksi Return Saham Menggunakan Hybrid XGBoost dan Improved Firefly Algorithm un- tuk Saham – Saham dalam Indeks LQ45, Fakultas Informatika Universitas Telkom (2023).

[2] S. E. Herni Yulianti, Oni Soesanto, Yuana Sukmawaty, Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit, Jomta (2022) 21–26.

[3] B. Jange, Prediksi Harga Saham Bank BCA Menggunakan XGBoost 3 (2022).

[4] K. K. Yun, S. W. Yoon, D. Won, Prediction of stock price direction using a hybrid ga-xgboost algorithm with a three-stage feature engineering process, Expert Syst. Appl. 186 (2021) 115716.

[5] M. S. Ansary, Machine learning for predicting the stock price direction with trading indicators, ENP Engineering Science Journal (2023).

[6] T. Liwei, F. Li, S. Yu, Y. Guo, Forecast of lstm-xgboost in stock price based on bayesian optimization, Intelligent Automation & Soft Computing (2021).

[7] D. Kumar, T. B. C, S. N. R. C, V. A, A. S. Devi, D. Kavitha, Analysis and prediction of stock price using hybridization of sarima and xgboost, 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT) (2022) 1–4.

[8] Y. Yang, Y. Wu, P. Wang, X. Jiali, Stock price prediction based on xgboost and lightgbm, E3S Web of Conferences (2021).

[9] A. B. Gumelar, H. Setyorini, D. P. Adi, S. Nilowardono, Latipah, A. Widodo, A. T. Wibowo, M. T. Sulistyono, E. Christine, Boosting the ac- curacy of stock market prediction using xgboost and long short-term mem- ory, 2020 International Seminar on Application for Technology of Informa- tion and Communication (iSemantic) (2020) 609–613.

[10] P. H. Vuong, T. T. Dat, T. K. Mai, P. H. Uyen, P. T. Bao, Stock-price forecasting based on xgboost and lstm, Comput. Syst. Sci. Eng. 40 (2022) 237–246.

[11] S. Basak, S. Kar, S. Saha, L. Khaidem, S. R. Dey, Predicting the direction of stock market prices using tree-based classifiers, The North American Journal of Economics and Finance (2019).

[12] C. Wang, X. Bai, Boosting learning algorithm for stock price forecasting, IOP Conference Series: Materials Science and Engineering 322 (2018).

[13] M. Vijh, D. Chandola, V. A. Tikkiwal, A. Kumar, Stock closing price prediction using machine learning techniques, Procedia Computer Science 167 (2020) 599–606.

[14] S. Pawaskar, Stock price prediction using machine learning algorithms, In- ternational Journal for Research in Applied Science and Engineering Tech- nology (2022)

[15] L. Chen, Using machine learning algorithms on prediction of stock price, 2020. doi:10.32732/jmo.2020.12.2.84.

[16] M. Nikou, G. Mansourfar, J. Bagherzadeh, Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms, Intell. Syst. Account. Finance Manag. 26 (2019) 164–174.

[17] B. Jeevan, E. Naresh, B. P. V. Kumar, P. Kambli, Share price prediction using machine learning technique, 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C) (2018) 1–4.

[18] D. Tarwidi, S. R. Pudjaprasetya, D. Adytia, M. Apri, An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach, MethodsX 10 (2023) 102119