BAREKENG: Journal of Mathematics and Its Applications
September 2025 Volume 19 Issue 3 Page 2155-2168
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss3pp2155-2168

PERFORMANCE EVALUATION OF THE INDF.JK STOCK PRICE MOVEMENT PREDICTION MODEL USING RANDOM FOREST METHOD WITH GRID SEARCH CROSS VALIDATION OPTIMIZATION

Della Zaria[1], Evy Sulistianingsih[2], Shantika Martha[3], Wirda Andani[4*]

[1,2,3,4]Statistics Department, Faculty of Mathematics and Natural Sciences, Universitas Tanjungpura Jln. Prof. Dr. H. Hadari Nawawi, Pontianak, 78124, Indonesia

*Corresponding author's e-mail: * **wirda.andani@math.untan.ac.id***

Article History:

Received: 1st January 2025
Revised: 8th March 2025
Accepted: 15th April 2025
Published: 1st July 2025

Keywords:

Cross Validation;
Grid Search;
Random Forest;
Stock.

ABSTRACT

Investment in financial instruments in Indonesia has shown significant growth over time, with stocks often being the first choice for investors to invest money. Unfortunately, deciding to buy and sell stocks is not easy. When determining the right time to buy or sell stocks, volatile stock price movements and losses caused by wrong decisions are investors' problems. Thus, it is essential to analyze stock price movement predictions. This study aims to evaluate the prediction model's performance for PT Indofood Sukses Makmur Tbk (INDF.JK) stock price movement in the next 30 days to reduce the risk of possible losses and help the decision-making process. We used the Random Forest method and Grid Search Cross Validation (CV) optimization to form the model. The data used is the closing price of INDF.JK stock for the period January 2, 2014, to December 29, 2023, from Yahoo Finance, which is processed into eight types of stock technical indicators, namely SMA_5, SMA_10, SMA_15, SMA_30, EMA_9, MACD, MACD_Signal, and RSI. The research pipeline includes descriptive statistics, preprocessing, feature and target variables determination, data split, model formation without and with optimization, testing accompanied by performance evaluation, and comparison of the formed model. The results show that the prediction model of INDF. JK's stock price movement in the next 30 days has excellent performance, proven accurate by 90.8% with the application of Random Forest and Grid Search CV. The Random Forest prediction model with Grid Search CV optimization has better performance indicators than the Random Forest model without Grid Search CV optimization, which is shown by the increase of all indicator values. The relative Strength Index is the variable with the best performance for the prediction model. It can be used as the primary consideration for investors when deciding on the buying and selling process of INDF.JK stock in the next 30 days.

How to cite this article:

D. Zaria, E. Sulistianingsih, S. Martha and W. Andani., "PERFORMANCE EVALUATION OF THE INDF.JK STOCK PRICE MOVEMENT PREDICTION MODEL USING RANDOM FOREST METHOD WITH GRID SEARCH CROSS VALIDATION OPTIMIZATION," *BAREKENG: J. Math. & App.*, vol. 19, no. 3, pp. 2155-2168, September, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article · Open Access

1. INTRODUCTION

Investment in financial instruments in Indonesia has shown significant growth over time. Data records that as of 2023, the number of capital market investors (stocks, bonds, and mutual funds) has reached 12,16 million, with a significant increase of 1,85 million from the previous year [1]. One type of investment in financial instruments that is popular among the public is stocks. This investment tool is often the primary choice of public investment because the results are believed to be more profitable [2]. A total of 5,7 million Single Investor Identification (SID) have been recorded as the number of stock investors in Indonesia until mid-2024 [3].

Unfortunately, deciding to buy and sell stocks is not easy, especially when determining the right time. Volatile stock price movements and losses incurred due to wrong decisions are special problems often faced by investors [4]. This problem is what raises the importance of analyzing stock price predictions in the future. One of the learning methods that can support this prediction process is Data Mining, with Random Forest as a Machine Learning algorithm that can be applied. As a development method of the Decision Tree, Random Forest is efficient in providing better results because this method is a performance improvement of the Decision Tree method that is prone to overfitting [5].

In its implementation, Random Forest requires a variety of hyperparameters, such as the number of estimators, depth tree, and other parameters that need to be tested to obtain the most optimal results [6]. Hyperparameter tuning can assist in finding the most optimal combination of parameters. One of them is Grid Search Cross Validation (CV) optimization. This optimization can find the best combination of parameters by testing every possible combination of parameters and a validation process by Cross Validation [7]. Therefore, this study applies Random Forest and Grid Search CV in predicting the PT Indofood Sukses Makmur Tbk (INDF.JK) stock price movement, which is a leading industry in Indonesia that operates in the primary consumer goods sector where there is an excellent opportunity always to be relied on. We selected this method based on the good model performance results of a combination of Random Forest and Grid Search CVs from previous studies.

Previous studies underpinned this study, including the combination of Random Forest with Grid Search CV, which resulted in the most optimal performance in the early detection study of survival of heart failure patients compared to Random Forest without optimization [8]. The performance of Random Forest and Grid Search CV is also shown as the most optimal combination of methods in predicting the closing price of Microsoft Corporation (MSFT) stock based on historical data, compared to Linear Regression, Neural Network, K-Nearest Neighbor, and Decision Tree as well as Evolutionary comparison optimization [9]. The variables involved in this study are based on Bank BCA's stock price prediction research using the XGBoost method based on technical indicators [10], which results in Exponential Moving Average (9) as the essential variable for the model, followed by Relative Strength Index, Moving Average Convergence/Divergence, Moving Average Convergence/Divergence Signal, Simple Moving Average (5), Simple Moving Average (10), Simple Moving Average (15), and Simple Moving Average (30) in a row.

Based on that, this study aims to evaluate the prediction model's performance for INDF.JK stock price movement in the next 30 days using the Random Forest method and Grid Search Cross Validation (CV) optimization to reduce the risk of possible losses and help the decision-making process based on performance results. This research will revalidate the performance optimization produced by Grid Search CV on Random Forest and provide a new view of INDF.JK stock, especially the role of each technical indicator in predicting the movement of INDF.JK stock price in the next 30 days. This research brings a novel target variable category system that is based on upward or downward movements in the following 30 days.

2. RESEARCH METHOD

2.1 Materials and Data

The raw data used in this study are historical data on the closing price of PT Indofood Sukses Makmur Tbk (INDF.JK) stock from January 2, 2014, to December 29, 2023. This data is secondary data collected from www.yahoofinace.com. It is processed into stock technical indicators used as independent variables to make predictions later.

Stock technical indicators are indices used to calculate stock technical analysis. This analysis is carried out to predict the direction of stock price movements and other market components based on historical stock data, so it is often used to determine the timing of buying and selling short-term stocks [11]. The technical indicators of the stock are as follows.

2.1.1 Simple Moving Average (SMA), Exponential Moving Average (EMA), Moving Average Convergence/Divergence (MACD), and Relative Strength Index (RSI)

Simple Moving Average (SMA) is an indicator that can calculate the average movement of stock prices over n days. The formula for calculating SMA is as follows [12].

$$SMA_i = \frac{C_i + C_{i-1} + C_{i-2} + \cdots + C_{i-(n-1)}}{n} \tag{1}$$

where SMA_i is the average price movement on the i-th day ($i = 1,2, \ldots, N$) with N as the amount of data, C_i is the closing price of stock on the i-th day, C_{i-1} on the previous 1 day, C_{i-2} on the previous 2 days, $C_{i-(n-1)}$ on $(n-1)$ the previous day, i is the data point, and n is the period ($n = 1,2, \ldots, N$).

Exponential Moving Average (EMA) is a development indicator of the SMA that measures the weighted average stock price over n days. It can be calculated using the following formula [13].

$$EMA_i = (C_i - EMA_{i-1}) * \left(\frac{2}{n+1}\right) + EMA_{i-1} \tag{2}$$

where EMA_i is the EMA value on the i-th day and EMA_{i-1} is the EMA value on the previous 1 day.

Moving Average Convergence/Divergence (MACD) is an indicator that can identify changes in price direction along with buy or sell signals. The MACD calculation formula is as follows [14].

$$MACD = EMA_{12}(C) - EMA_{26}(C) \tag{3}$$

$$SignalLine = EMA_9(MACD) \tag{4}$$

where C is the stock's closing price and EMA_n is the EMA value for n days.

Relative Strength Index (RSI) is an indicator that can provide overbought and oversold signals. It can be calculated using the formula below [15].

$$RSI = 100 - \frac{100}{1 + RS} \tag{5}$$

$$RS = \frac{Average\ Gain}{Average\ Loss} \tag{6}$$

$$\begin{aligned} Average\ Gain &= (u1 + u2 + \cdots + u_n)/n \\ Average\ Loss &= (d1 + d2 + \cdots + d_n)/n \end{aligned} \tag{7}$$

$$u_n = C_i - C_{i-1}\ ;\ d_n = -(C_i - C_{i-1}) \tag{8}$$

where u_n is the last increase in the period n, d_n is the last decrease in the period n, and n is the period (14 days).

An exponential smoothing formula can be used to find the RSI value on day 16 and after, referring to J. Welles Wilder's concept in 1978.

$$New\ Average\ Gain = \frac{(Previous\ Average\ Gain \times 13) + Today's\ Gain}{14} \tag{9}$$

$$New\ Average\ Loss = \frac{(Previous\ Average\ Loss \times 13) + Today's\ Loss}{14} \tag{10}$$

The list of technical indicators used in this study is shown in **Table 1**.

Table 1. List of Technical Indicators

Variable Name	Definition
SMA_5	Average stock price movement in 5 days
SMA_10	Average stock price movement in 10 days
SMA_15	Average stock price movement in 15 days
SMA_30	Average stock price movement in 30 days
EMA_9	Weighted average stock price over 9 days
MACD	EMA12 – EMA26 values
MACD_Signal	Average movement of the EWM index weights for MACD in 9 days
RSI	Overbought and oversold informant

2.2 Applied Methods

2.2.1 Random Forest

Random Forest is one of the Machine Learning methods that can be used for prediction and classification cases [16] by building several trees in a forest. This method's main concept is to form several decision trees from the training data using the bootstrap aggregating (bootstrapping) technique [17] and use the average vote of each formed decision tree to classify or predict classes on the new data. The steps to apply the Random Forest method are as follows [18].

a. Specifies "k" sample of D_a datasets randomly drawn from dataset D with returns, where k is the value of the number of decision trees.

b. Uses D_a data samples in constructing a-decision tree *(a = 1,2, ...,k)* so that the Classification and Regression Trees (CART) method can be applied. The CART method utilizes the information gain value to determine each node in the decision tree. The value of the information gain can be calculated using **Equation (11)**.

$$Gain(A) = Info(D) - Info_A(D) \tag{11}$$

with the values $Info(D)$ and $Info_A(D)$ can be obtained by **Equation (12)** and **Equation (13)**.

$$Info(D) = -\sum_{i=1}^{m} p_i log_2(p_i) \tag{12}$$

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j) \tag{13}$$

where m is the number of target classes and p_b is the probability of the occurrence of class b in partition D where $b = 1,2, \dots, m$, v is the number of partitions, D_j is the total j-th partition where $j = 1,2, \dots, v$ and D is the number of rows on the entire partition.

c. Repeat steps 1 and 2 k times.

An illustration of how Random Forest works is shown in **Figure 1**.

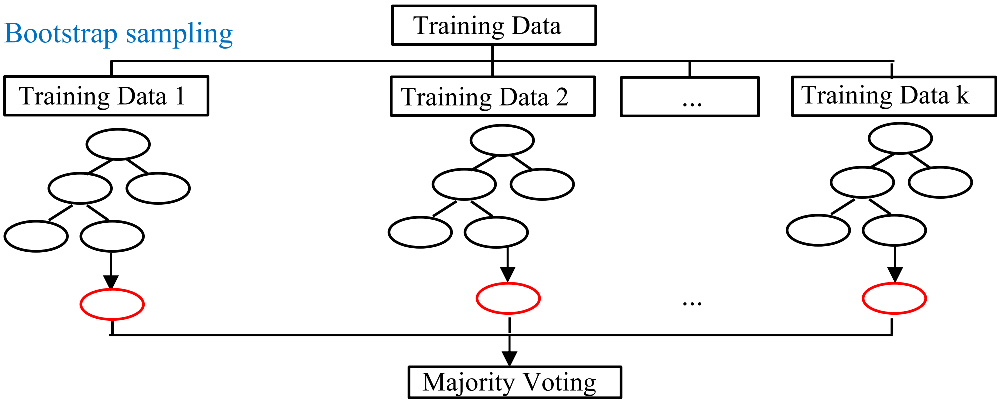


Figure 1. Ilustration of Random Forest Method

In its implementation, Random Forest requires some hyperparameters that will determine the optimization of the model being built. The two core hyperparameters in Random Forest, namely mtry and ntree, with the following description [19].

a. Hyperparameter mtry is a value that sets the number of randomly selected variables and is intended to be a candidate considered on each node in building the tree. Generally, the mtry value can be based on the formula $\sqrt{p}$ for the classification case and $p/3$ for the regression case, where p is the number of independent variables involved in building the model.

b. Hyperparameter ntree is a value that regulates the number of trees built in the model. Some software sets an ntree value of 500 as the default, and for large amounts of data, an ntree value of at least 100 can be set for more optimal model performance.

The level of contribution of variables to the model can be seen through the level of importance of variables, one of which is the Mean Decrease Gini (MDG), which can measure the significance level of variables in separate classes in the model. MDG is a decision tree metric to determine the variable splits by averaging the total decrease of the gini index from all of the trees in the forest [20]. The formula for calculating the MDGs is as follows [21].

$$MDG_{X(h)} = \frac{1}{k}[1 - \sum\nolimits_{k} Gini\ (h)^k] \tag{14}$$

where $MDG_{X(h)}$ is the MDG value that measures the level of importance of independent variables X_h where $h = 1, 2, ..., p$. $Gini\ (h)^k$ is the Gini index for the independent variable X_h in the k-th tree, where k is the number of trees in the Random Forest. The formula for calculating the Gini index is as follows [22].

$$Left\ Node\ (L): imp(t_L) = \sum\nolimits_{d=1}^{2} p_{tL}^{(d)}(1 - p_{tL}^{(d)}) \tag{15}$$

$$Right\ Node\ (R): imp(t_R) = \sum\nolimits_{d=1}^{2} p_{tR}^{(d)}(1 - p_{tR}^{(d)}) \tag{16}$$

$$t\ Node\ (L): imp(t) = \sum\nolimits_{c=1}^{2} p_{t}^{(c)}(1 - p_{t}^{(c)}) \tag{17}$$

$$p_t^{(c)} = \frac{n_t^{(c)}}{n_t} \qquad p_t^{(d)} = \frac{n_t^{(d)}}{n_t} \tag{18}$$

where $p_t^{(c)}, p_t^{(d)}$ are the proportion of objects of the c-th or d-th classification class at node. $n_t^{(c)}, n_t^{(d)}$ are number of observations of the c-th or d-th classification class at node and n_t is number of all observations at t node.

2.2.2 Grid Search and K-Fold Cross Validation

Grid Search is a type of optimization used to find the combination of hyperparameters with the most optimal performance results. It is brought to the prediction model by experimenting with parameter combinations one by one. The way Grid Search works has a special characteristic: each parameter is placed on a grid between 2 parameters that are not the same type, with the most optimal value located in the meeting between the two [23].

K-Fold Cross Validation is a cross validation method used to evaluate the model's performance thoroughly. It divides the data into several parts (folds), trains the model, and evaluates it k times, where k is the number of folds [24].

2.2.3 Model Performance Evaluation

The binary classification model formed from a machine learning method can be measured accurately using the Confusion Matrix, which is a table that contains information comparing the predicted results with the actual results of each class [25]. The visualization of the Confusion Matrix is shown in Table 2.

Table 2. Confusion Matrix

Correct Classification	Classified As	
	Prediction (+)	**Prediction (-)**
Actual (+)	True Positive (TP)	False Negative (FN)
Actual (-)	False Positive (FP)	True Negative (TN)

The formula for the size of the performance evaluation of the classification model is in Table 3.

Table 3. Model Evaluation Size

Size	Definition	Formula
Accuracy	Accuracy of classification models in correctly predicting data classes	$\frac{TP + TN}{TP + FP + FN + TN}$
Precision	The ratio of positive data that the model successfully predicts correctly from the overall positive predictions	$\frac{TP}{TP + FP}$
Sensitivity	The ratio of positive data that the model has successfully predicted correctly, namely as positive	$\frac{TP}{TP + FN}$
Specificity	The ratio of negative data that the model successfully predicts, namely as negative	$\frac{TN}{TN + FP}$

3. RESULT AND DISCUSSION

3.1 Descriptive Statistics

The movement in the closing price of PT Indofood Sukses Makmur Tbk (INDF.JK) stock in the selected period, namely January 2, 2014 to December 29, 2023, is shown in Figure 2. It depicts the fluctuations of INDF.JK stock over the selected period, with a data movement pattern visualized. This stock experienced a drastic decline in the closing price in 2015 but was accompanied by a drastic increase in 2016. In 2020, INDF.JK stock again experienced a drastic decrease in the closing price and an increase above the average line sometime afterward.

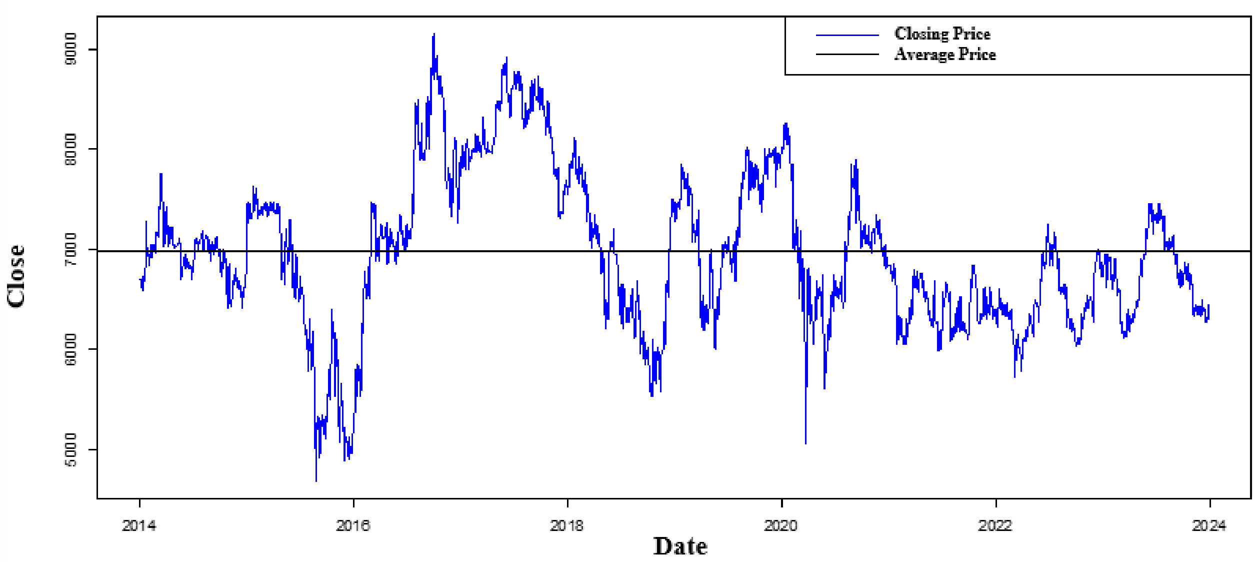


Figure 2. Graph of the Closing Price Movement of indf.jk Stock in 2014-2023

3.2 Data Preprocessing

The check results were obtained that there was no missing value, either in the form of a value in the missing data or an NA value; therefore, the data could be said to be ready to continue the next preprocessing stage. This is followed by an exponential smoothing process to allocate a more significant weight to the current price data and reduce the weight value exponentially on the price data in the past so that the prediction model can later adjust to the latest situation pattern while learning from past historical data. This study uses a simple exponential smoothing method, Single Exponential Smoothing (SES), which is the closing price value of INDF.JK stock is refined exponentially with the calculation formula in Equation (19).

$$F_i = (\alpha \times X_i) + ((1 - \alpha) \times F_{i-1}) \tag{19}$$

where F_i is the forecast value on the i-th day, α is the smoothing parameter ($0 < \alpha < 1$), X_i is the actual value on the i-th day, and F_{i-1} is the forecast value on the previous 1 day.

This study used an optimal alpha of 0.025 (the result of the Excel solver). It means that the model allocates a weight of 2.5% for the most recent value and 97.5% for the historical value of stock in the past.

3.3 Determination of Feature and Target Variables

3.3.1 Feature Variables

The feature variables used are several stock technical indicators formed based on the value of the closing price of INDF.JK stock has been smoothed exponentially. The calculation gives rise to NA values in several rows, so cleaning is carried out, with the final result as follows.

Table 4. Results of Calculations of Technical Indicators after Cleaning

Number	SMA_5	SMA_10	SMA_15	SMA_30	EMA_9	MACD	MACD _Signal	RSI
1	6814.474	6802.340	6789.859	6750.740	6804.520	33.353	32.980	97.247
2	6819.487	6807.281	6794.612	6755.139	6809.521	33.457	33.076	97.491
3	6824.125	6811.912	6799.330	6759.760	6814.248	33.448	33.150	97.636
....	...	...	...	...	...	...	...	...
2448	6566.677	6584.328	6600.190	6648.515	6580.407	-47.428	-48.202	0.079
2449	6561.510	6578.220	6594.268	6641.927	6575.145	-46.797	-47.921	0.076

3.3.2 Target Variables

The target variable used is based on the closing price value of INDF.JK stock without exponential smoothing, which contains two class categories, namely "up" movement notated with "1" and "down" movement notated with "-1".

This study is aimed at evaluating the prediction models of INDF.JK stock price movement in the next 30 days, the class of each row of data is determined based on the result of subtraction between the data 30 days after and the current data. If the reduction results show a value greater than 0, then the row of data is included in class "1," or there is an upward movement. Conversely, if the result of the subtraction shows a value less than or equal to 0, then the data row is included in the "-1" class, or there is a downward movement. The following is the formula for categorizing target variables in mathematical language.

$$Target_i = Sign(C_{i+30} - C_i) \tag{20}$$

where $Target_i$ means the target variable class from the data of the i-th day, *Sign* refers to the use of class signs, namely 1 or -1, C_{i+30} is the closing price of the stock 30 days after the i-th day, and C_i means the closing price of the stock on the i-th day.

The results of the class categorization on the data are shown in **Table 5**.

Table 5. Target Variable Results

Number	Date	INDF.JK Stock Closing Price	Target Variable
1	20/02/2014	7000	1
2	21/02/2014	7050	1
3	24/02/2014	6975	1
...	...	...	...
31	04/04/2014	7150	-1
32	07/04/2014	7225	-1
...	...	...	...
2419	15/11/2023	6425	1

3.4 Data Split

The split of training data and testing data in this study uses an 80:20 composition, where 80% of the existing data is allocated to train the model through training data, and the remaining 20% of the total data is allocated to test the model formed through data testing. This composition selection is used to keep in mind the balance between the training and testing processes; therefore, the model is obtained that does not just memorize existing data (overfitting) but learns from training data and tests on testing data. The composition of the data split is listed in **Table 6**.

Table 6. Composition of Data Split

Data Type	Class		Total	Composition
	1	-1		
Training Data	884	1055	1939	80%
Testing Data	232	248	480	20%
Total	1116	1303	2419	100%

Based on **Table 6**, as many as 1939 (80%) of the data rows used as training data consist of 884 rows of class 1 data and 1055 rows of class -1 data. The testing data contains 480 (20%) rows of 232 rows of class 1 data and 284 rows of class -1 data. According to this number, training, and testing data can be considered balanced because no majority class reaches the number twice plus one of the minority classes.

3.5 Formation and Performance Evaluation of Model by Random Forest without Grid Search CV Optimization

The model's formation of the Random Forest without Grid Search CV is carried out in the same way as the construction of the model by Random Forest with hyperparameters in the default settings. The parameters in Random Forest used to build a model in this study are mtry and ntree. The value of each parameter in the formation of this prediction model is set automatically by the RStudio program by default, where a value of 2 is applied to the mtry parameter and 500 to the ntree.

The prediction results obtained from the Random Forest without Grid Search CV or default through the Confusion Matrix are included in **Table 7**.

Table 7. Confusion Matrix Prediction Results of Random Forest without Grid Search CV

Actual	Prediction	
	-1 (downward)	**1 (upward)**
-1 (downward)	226 (TP)	22 (FN)
1 (upward)	26 (FP)	206 (TN)

Using the values provided in Table 7, it is possible to calculate the values that will also measure the performance of the model as follows:

$$\text{Accuracy} = \frac{226+206}{226+22+26+206} = \frac{432}{480} = 0.9$$

$$\text{Sensitivity} = \frac{226}{226+22} = \frac{226}{248} = 0.911$$

$$\text{Precision} = \frac{226}{226+26} = \frac{226}{252} = 0.897$$

$$\text{Specificity} = \frac{206}{206+26} = \frac{206}{232} = 0.888$$

Based on the calculation results, the model has excellent performance, with an accuracy level of 0.9 or 90%, accurately predicting the closing price movement of INDF.JK stock in the next 30 days. From the overall prediction of downward movements in the price, as many as 89.7% of the data on the closing price of INDF.JK stock that experienced a downward was correctly predicted by the model. In addition, 91.1% of INDF.JK's closing stock price data, which had a downward movement in the next 30 days, was correctly predicted by the model as a downward—conversely, 88.8% of INDF.JK's closing stock price data, which had an upward movement in the next 30 days, was correctly predicted by the model as upward.

The variable importance of the prediction model by Random Forest without Grid Search CV optimization in this study is shown in Table 8.

Table 8. Variable Importance Level of Model by Random Forest without Grid Search CV

Number	Feature Variable	Overall Importance Value
1.	RSI	133.373
2.	SMA_5	128.755
3.	SMA_15	125.977
4.	EMA_9	121.901
5.	SMA_30	120.789
6.	SMA_10	119.565
7.	MACD_Signal	110.119
8.	MACD	100.990

According to Table 8, RSI is the variable that occupies the top position with the highest variable importance of all existing variables, which is 133.373. Followed by SMA_5 worth 128.755; SMA_15 worth 125.977; EMA_9 worth 121.901; SMA_30 worth 120.789; SMA_10 worth 119.565; MACD_Signal worth 110.119; and ended with the MACD occupying the last position with the least variable importance level for the prediction model of 100.990.

3.6 Formation and Performance Evaluation of Model by Random Forest with Grid Search CV Optimization

The model's formation of the Random Forest with Grid Search CV optimization begins with the hyperparameter tuning step, which finds a combination of parameters that can provide the most optimal accuracy value later. The candidates for the mtry parameter value selected are 2, 3, 4, and 8, while the values of the ntree parameter are 50, 100, 200, 500, and 1000. Cross Validation will validate the process of finding the best parameters. This study uses 10-Fold Cross Validation to evaluate the model's performance, meaning each pair of parameters is repeated 10 times in the Grid Search process. The hyperparameter tuning results of Grid Search CV for the Random Forest method are shown in Table 9.

Table 9. **Results of Hyperparameter Tuning Grid Search CV**

Number	Parameter		Accuracy
	mtry	ntree	
1.	2	50	0.884
2.	2	100	0.887
3.	2	200	0.886
4.	2	500	0.885
5.	2	1000	0.886
6.	3	50	0.889
7.	3	100	0.887
8.	**3**	**200**	**0.894**
9.	3	500	0.890
10.	3	1000	0.889
11.	4	50	0.889
12.	4	100	0.887
13.	4	200	0.889
14.	4	500	0.889
15.	4	1000	0.890
16.	8	50	0.891
17.	8	100	0.888
18.	8	200	0.890
19.	8	500	0.891
20.	8	1000	0.890

Based on **Table 9**, from the 20 types of combinations between mtry and ntree parameters, it was found that the best combination of parameters as the producer of the highest accuracy level of 0.894 was owned by the parameter pairs mtry 3 and ntree 200. This means that the mtry parameter of 3 and ntree worth 200 succeeded in predicting the training data with an accuracy of 89.4%. The accuracy of the predictions produced by this combination of parameters shows the highest value compared to other combinations, so this combination is the most feasible to apply to the model to create the most optimal model.

The prediction results from the Random Forest with Grid Search CV optimization through the Confusion Matrix are included in **Table 10**.

Table 10. **Confusion Matrix of Prediction Results by Random Forest with Grid Search CV**

	Prediction	
Actual	**-1 (downward)**	**1 (upward)**
-1 (downward)	227 (TP)	21 (FN)
1 (upward)	23 (FP)	209 (TN)

Using the values provided in **Table 10**, it is possible to calculate the values that will also measure the performance of the model as follows:

$$\text{Accuracy} = \frac{227+209}{227+21+23+209} = \frac{436}{480} = 0.908 \qquad \text{Sensitivity} = \frac{227}{227+21} = \frac{227}{248} = 0.915$$

$$\text{Precision} = \frac{227}{227+23} = \frac{227}{250} = 0.908 \qquad \text{Specificity} = \frac{209}{209+23} = \frac{209}{232} = 0.901$$

Based on the calculation result, the model has excellent performance, with an accuracy level of 0.908 or 90.8%, to accurately predict the closing price movement of INDF.JK stock in the next 30 days. From the overall prediction of downward movements in the price, as many as 90.8% of the data on the closing price of INDF.JK stock that experienced a downward was correctly predicted by the model. In addition, 91.5% of the data on the closing price of INDF.JK stock, which had a downward movement in the next 30 days, was correctly predicted by the model as a downward—conversely, 90.1% of INDF.JK's closing stock price data, which had an upward movement in the next 30 days, was correctly predicted by the model as upward. The variable importance of the prediction model by Random Forest with Grid Search CV optimization in this study is shown in **Table 11**.

Table 11. Variable Importance Level of Model by Random Forest with Grid Search CV

Number	Feature Variable	Overall Importance Value
1.	RSI	141.432
2.	SMA_5	129.507
3.	SMA_30	128.076
4.	EMA_9	121.138
5.	SMA_15	118.513
6.	SMA_10	116.381
7.	MACD_Signal	107.479
8.	MACD	99.153

According to Table 11, RSI is the variable that occupies the top position with the highest variable importance of all existing variables, which is 141.432—followed by SMA_5 worth 129.507; SMA_30 worth 128.076; EMA_9 worth 121.138; SMA_15 worth 118.513; etc; and ended by the MACD which occupies the last position with the least variable importance level for the prediction model of 99.153.

3.7 Comparison of Prediction Model Performance Evaluation Results

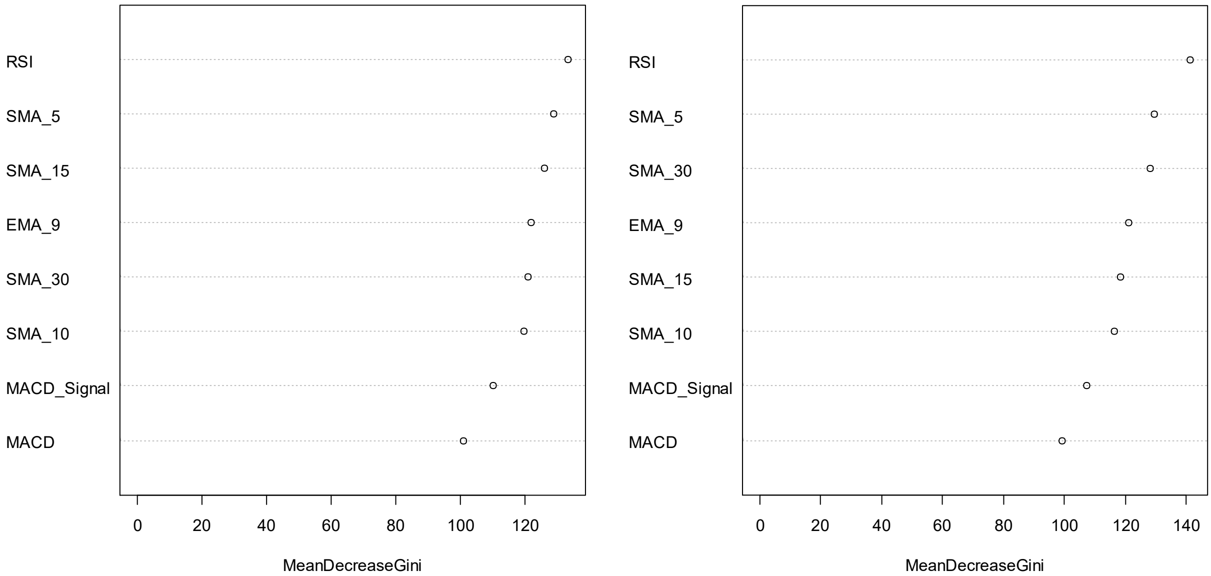
A comparison of the model's performance indicator values of the Random Forest without and with Grid Search CV optimization is attached in Table 12.

Table 12. Comparison of the Performance Indicator Values of the Two Models

Indicator	Without Grid Search CV	With Grid Search CV
Accuracy	0.900	0.908
Precision	0.897	0.908
Sensitivity	0.911	0.915
Specificity	0.888	0.901

It can be seen that all model performance indicators have increased, although it is not too significant. However, this increase shows that using Grid Search CV as a parameter optimization can optimize the model's performance with Random Forest, which is formed by performing better than the model without Grid Search CV optimization.

A comparison of the model's variable importance level of Random Forest without and with CV Grid Search optimization can be seen in Figure 3.



(a) (b)

Figure 3. Comparison of the Importance of the Two Variables of the Model
(a) Random Forest without Grid Search CV, (b) Random Forest with Grid Search CV

Based on Figure 3, both models have almost the same level of relative variable importance, including RSI as the variable with the highest significance level and MACD as the variable with the lowest level of importance. However, there is a difference between the third and fifth positions filled by SMA_15 and SMA_30 in the first model and SMA_30 and SMA_15 in the second model. This level of importance describes the contribution of variables in making predictions with class separation. Because the model of

Random Forest with Grid Search CV is the best, the level of importance of the second model variable can be considered when deciding to buy and sell INDF.JK stock in the next 30 days. In addition, when a contradiction of information occurs between variable A and variable B (another variable with a position of interest below it in this study) in predicting the movement of INDF.JK stock in the next 30 days and prospective investors are recommended to consider more of the signals given by variable A.

This research validated the optimal use of Grid Search CV to improve Random Forest performance, which aligns with previous studies. Research from [10] found RSI as a variable that occupies the second level of importance. This is in line with the results of this research, namely RSI as the variable with the highest level of importance for the prediction model. Despite different application data, RSI shows the most optimal performance for predicting stock price movements. Hence, it should be used as the main consideration when making stock buying and selling decisions.

4. CONCLUSIONS

The evaluation results of the prediction model from PT Indofood Sukses Makmur Tbk (INDF.JK) stock price movement in the next 30 days has excellent performance, proven accurate to be analyzed using the Random Forest method accompanied by Grid Search Cross Validation optimization with an accuracy rate of 90.8%. Based on the evaluation results of the prediction model performance, this study also shows that the implementation of Grid Search CV can optimize the performance of Random Forest in making predictions with an increase in each model performance evaluation indicator. The RSI technical indicator shows the best performance for the prediction result; therefore, it can be used as the primary consideration for potential investors when deciding to buy and sell INDF.JK stock in the next 30 days.

REFERENCES

[1] idx.co.id, "MELALUI BERBAGAI PENCAPAIAN TAHUN 2023, PASAR MODAL INDONESIA TUNJUKKAN OPTIMISME HADAPI TAHUN 2024," BURSA EFEK INDONESIA, Press Release 117/BEI.SPR/12-2023, Desember 2023. [Online]. Available: https://www.idx.co.id/id/berita/siaran-pers/2080

[2] M. N. Syafira and D. Saepudin, "PREDIKSI RETURN SAHAM DENGAN METODE RANDOM FOREST DAN PENERAPANNYA UNTUK SELEKSI PORTOFOLIO," in *e-Proceeding of Engineering*, Oktober 2023, pp. 4996–5001.

[3] idx.co.id, "INVESTOR PASAR MODAL TEMBUS 13 JUTA, KOLABORASI DENGAN KOMUNITAS DAN IDX MOBILE JADI ANDALAN," Bursa Efek Indonesia, Press Release 039/BEI.SPR/06-2024, Jun. 2024. [Online]. Available: https://www.idx.co.id/id/berita/siaran-pers/2166

[4] A. Pratomo, R. F. Umbara, and A. A. Rohmawati, "PREDIKSI PERGERAKAN HARGA SAHAM DENGAN METODE RANDOM FOREST MENGGUNAKAN TREND DETERMINISTIC DATA PREPARATION," in *e-Proceeding of Engineering*, Apr. 2019, pp. 2545–2556.

[5] K. Sidiq and D. R. S. Saputro, "CABLE NEWS NETWORK (CNN) ARTICLES CLASSIFICATION USING RANDOM FOREST ALGORITHM WITH HYPERPARAMETER OPTIMIZATION," *BAREKENG J. Ilmu Mat. Dan Terap.*, vol. 17, no. 2, pp. 0847–0854, Jun. 2023, doi: https://doi.org/10.30598/barekengvol17iss2pp0847-0854.

[6] M. Azhar and H. F. Pardede, "KLASIFIKASI DIALEK PENGUJAR BAHASA INGGRIS MENGGUNAKAN RANDOM FOREST," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 2, pp. 439–446, Apr. 2021, doi: https://doi.org/10.30865/mib.v5i2.2754.

[7] L. N. Aina, V. R. S. Nastiti, and S. K. Aditya, "IMPLEMENTASI EXTRA TREES CLASSIFIER DENGAN OPTIMASI GRID SEARCH CV PADA PREDIKSI TINGKAT ADAPTASI," *MIND Multimed. Artif. Intell. Netw. Database J.*, vol. 9, no. 1, pp. 78–88, 2024, doi: https://doi.org/10.2670/mindjournal.v9i1.78-84.

[8] B. F. Sitanggang and P. Sitompul, "DETEKSI AWAL KELANGSUNGAN HIDUP PASIEN GAGAL JANTUNG MENGGUNAKAN MACHINE LEARNING METODE RANDOM FOREST," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 2, pp. 3347–3357, 2024, doi: 10.31004/innovative.v4i2.8189.

[9] D. Z. H. Iskandar and Y. Ramdhani, "OPTIMASI PARAMETER RANDOM FOREST MENGGUNAKAN GRID SEARCH UNTUK ANALISIS TIME SERIES," *PETIR J. Pengkaj. Dan Penerapan Tek. Inform.*, vol. 16, no. 2, pp. 267–277, Nov. 2023, doi: https://doi.org/10.33322/petir.v16i2.2084.

[10] B. Jange, "PREDIKSI HARGA SAHAM BCA MENGGUNAKAN XGBOOST," *Arbitr. J. Econ. Account.*, vol. 3, no. 2, pp. 231–237, Nov. 2022, doi: https://doi.org/10.47065/arbitrase.v3i2.495.

[11] M. Sadikin and R. Agustina, "ANALISIS FUNDAMENTAL DAN TEKNIKAL SAHAM BCA DAN BRI (TAHUN 2019-2021)," *Semin. Nas. Akunt. Dan Call Pap. SENAPAN*, vol. 3, no. 1, pp. 57–67, Oct. 2023, doi: https://doi.org/10.33005/senapan.v3i1.285.

[12] T. Astutiningsih, D. R. S. Saputro, and Sutanto, "OPTIMASI ALGORITME XTREME GRADIENT BOOSTING (XGBOOST) PADA HARGA SAHAM PT. UNITED TRACTORS TBK.," *SPECTA J. Technol.*, vol. 7, no. 3, pp. 632–641, Dec. 2023, doi: https://doi.org/10.35718/specta.v7i3.1031.

[13] K. Valiant, Y. Lukito, and G. Santosa, "SISTEM PREDIKSI HARGA SAHAM LQ45 DENGAN RANDOM FOREST CLASSIFIER," *JUTEI*, vol. 3, no. 2, pp. 127–136, Oktober 2019, doi: https://doi.org/10.21460/jutei.201932:187.

[14] Arman, A. I. Haanurat, and Nur'aeni, "Analisis Moving Average Convergence Divergence pada Investasi Saham Syariah saat New Era," *Ekombis Sains J. Ekon. Keuang. Dan Bisnis*, vol. 7, no. 1, pp. 1–10, Mar. 2022, doi: 10.24967/ekombis.v7i1.1540.
[15] R. G. Firdaus, "ANALISIS TEKNIKAL SAHAM MENGGUNAKAN INDIKATOR RSI DAN BOLLINGER BANDS PADA SAHAM KONSTRUKSI," *J. Pasar Modal Dan Bisnis*, vol. 3, no. 1, pp. 15–26, Feb. 2021, doi: https://doi.org/10.37194/jpmb.v3i1.60.
[16] I. Maulana, A. M. Siregar, and A. Fauzi, "OPTIMIZATION OF MACHINE LEARNING MODEL ACCURACY FOR BRAIN TUMOR CLASSIFICATION WITH PRINCIPAL COMPONENT ANALYSIS," *J. Tek. Inform. JUTIF*, vol. 5, no. 3, pp. 903–915, 2024, doi: https://doi.org/10.52436/1.jutif.2024.5.3.2058.
[17] L. Muflikhah, Widodo, W. Firdaus, and Solimun, *MACHINE LEARNING DALAM BIOINFORMATIKA*. in 1. Universitas Brawijaya Press, 2021.
[18] R. A. Haristu and P. H. P. Rosa, "Penerapan Metode Random Forest untuk Prediksi Win Ratio Pemain Player Unknown Battleground," *MEANS Media Inf. Anal. Dan Sist.*, vol. 4, no. 2, pp. 120–128, Oct. 2019, doi: https://doi.org/10.54367/means.v4i1.545.
[19] C. Yulianto, "MODEL PENILAIAN TANAH MASSAL BERBASIS BIDANG TANAH MENGGUNAKAN ALGORITMA RANDOMFOREST DI KOTA SURAKARTA," *J. Pertanah.*, vol. 14, no. 1, pp. 26–39, 2023, doi: https://doi.org/10.53686/jp.v4i1.204.
[20] A. S. M. Shafi, M. M. I. Molla, J. J. Jui, and M. M. Rahman, "DETECTION OF COLON CANCER BASED ON MICROARRAY DATASET USING MACHINE LEARNING AS A FEATURE SELECTION AND CLASSIFICATION TECHNIQUES," *SN Appl. Sci.*, vol. 2, no. 7, p. 1243, Jul. 2020, doi: https://doi.org/10.1007/s42452-020-3051-2.
[21] E. Christy and K. Suryowati, "ANALISIS KLASIFIKASI STATUS BEKERJA PENDUDUK DAERAH ISTIMEWA YOGYAKARTA MENGGUNAKAN METODE RANDOM FOREST," *J. Stat. Ind. Dan Komputasi*, vol. 6, no. 1, pp. 69–76, 2021.
[22] F. Ramayanti, D. Vionanda, D. Permana, and Zilrahmi, "APPLICATION OF RANDOM FOREST TO IDENTIFY FOR POOR HOUSEHOLDS IN WEST SUMATERA PROVINCE," *UNP J. Stat. Data Sci.*, vol. 1, no. 2, pp. 97–104, Mar. 2023, doi: https://doi.org/10.24036/ujsds/vol1-iss2/31.
[23] U. Sunarya and T. Haryanti, "PERBANDINGAN KINERJA ALGORITMA OPTIMASI PADA METODE RANDOM FOREST UNTUK DETEKSI KEGAGALAN JANTUNG," *J. Rekayasa Elektr.*, vol. 18, no. 4, pp. 241–247, Dec. 2022, doi: https://doi.org/10.17529/jre.v18i4.26981.
[24] Y. Widyaningsih, G. P. Arum, and K. Prawira, "APLIKASI K-FOLD CROSS VALIDATION DALAM PENENTUAN MODEL REGRESI BINOMIAL NEGATIF TERBAIK," *BAREKENG J. Ilmu Mat. Dan Terap.*, vol. 15, no. 2, pp. 315–322, Jun. 2021, doi: https://doi.org/10.30598/barekengvol15iss2pp315-322.
[25] N. Hadianto, H. B. Novitasari, and A. Rahmawati, "KLASIFIKASI PEMINJAMAN NASABAH BANK MENGGUNAKAN METODE NEURAL NETWORK," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 163–170, Sep. 2019, doi: https://doi.org/10.33480/pilar.v15i2.658.