

**EXTRA TREES METHOD FOR STOCK PRICE FORECASTING  
 WITH ROLLING ORIGIN ACCURACY EVALUATION**

**Dani Al Mahkya<sup>1,2</sup>, Khairil Anwar Notodiputro<sup>1</sup>, Bagus Sartono<sup>1</sup>**

<sup>1</sup>Department of Statistics, IPB University, Indonesia

<sup>2</sup>Actuarial Science Study Program, Institut Teknologi Sumatera, Indonesia

**e-mail:** [dani.almahkya@at.itera.ac.id](mailto:dani.almahkya@at.itera.ac.id)

**DOI: 10.14710/medstat.15.1.36-47**

**Article Info:**

Received: 31 March 2022

Accepted: 13 July 2022

Available Online: 27 July 2022

**Keywords:**

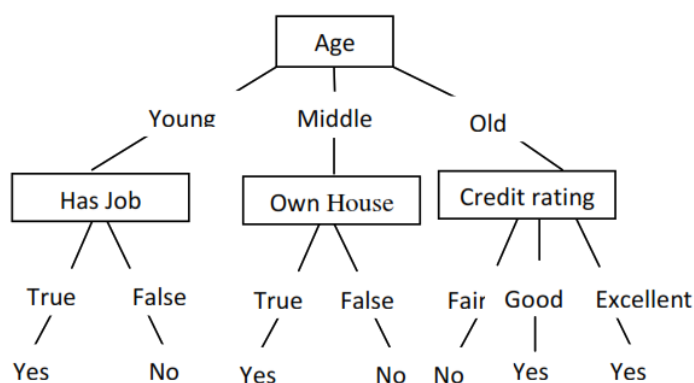
*Randomized Trees; Extra Trees;  
 Regression; Stock; Forecasting*

**Abstract:** Stock is an investment instrument that has risk in its management. One effort to minimize this risk is to model and make further forecasts of stock price movements. Time series data forecasting with autoregressive models is often found in several cases with the most popular approach being the ARIMA model. The tree-based method is one of the algorithms that can be used to forecast both in classification and regression. One ensemble approach to tree-based methods is Extra Trees. This study aims to forecast using the Extra Trees algorithm by evaluating forecasting accuracy with Rolling Forecast Origin on BRMS stock price data. Based on the results obtained, it is known that Extra Trees produces a fairly good accuracy for forecasting up to 6 days after training data with a MAPE of less than 0.1%.

**1. INTRODUCTION**

Stock is proof of ownership of the assets in the company that issued the stock. While the stock price is the value of a stock that reflects the wealth of the company that issued the stock, where changes or fluctuations are largely determined by the forces of supply and demand that occur in the stock market (Sulia, 2017). All companies that issue stock in Indonesia are listed on the Indonesia Stock Exchange (IDX). There are several sectors on the IDX including agriculture, mining, basic and chemical industries, various industries, consumer goods industry, property, infrastructure, finance, and trade services (Saham OK, 2011). One of the important aspects in the capital market that is the focus of investors in making investment decisions is the movement of stock prices, the movement of stock prices illustrates how the performance of the issuer is concerned (Agustina & Sumartio, 2014). Analyzing and predicting stock prices can be an instrument in making investment decisions. In general, time series data forecasting with autoregressive models is found in several cases with the most popular approach being the parametric Autoregressive Integrated Moving Average (ARIMA) model with various developments. Several related studies such as forecasting inflation in Demak using the ARIMA method (Astutik, et al., 2018), forecasting the ARIMA method for stock data of PT. Telekomunikasi Indonesia (Rezaldi & Sugiman, 2021), forecasting the stock price index of PT Verena Multi Finance Tbk using the ARIMA and ARCH-GARCH approaches (Fitriyani, et al., 2021), forecasting the number of passengers at Bakauheni port in the Sunda Strait Tsunami period with ARIMA approach

Intervention and outliers (Mahkya & Anggraini, 2019), as well as Wavelet and ARIMA analysis used for forecasting the gold price of PT. Antam Tbk Indonesia (Kurnia, et al., 2014). However, in addition to parametric models, forecasting in some cases can use a tree-based model approach. The tree-based method or decision tree is one of the learning methods that can be used to analyze complex data. Tree-based methods produce predictions with high accuracy, stability, and ease of interpretation (Banarjee, et al., 2019). Just like the characteristics of a tree, a decision tree has several parts including root nodes, branches, and leaf nodes. Each attribute at each node will be tested to determine other tree branches (Patel & Prajapati, 2018). An illustrative example of a decision tree is described in figure 1 (Saini, et al., 2014). The decision tree method can be developed into various approaches, such as research related to poverty classification using the Classification and Regression Tree (CART) approach (Ispriyanti, et al., 2019) and research on the comparison of CART, Bagging and Random Forest on the classification of simulation objects (Jatmiko, et al., 2019).



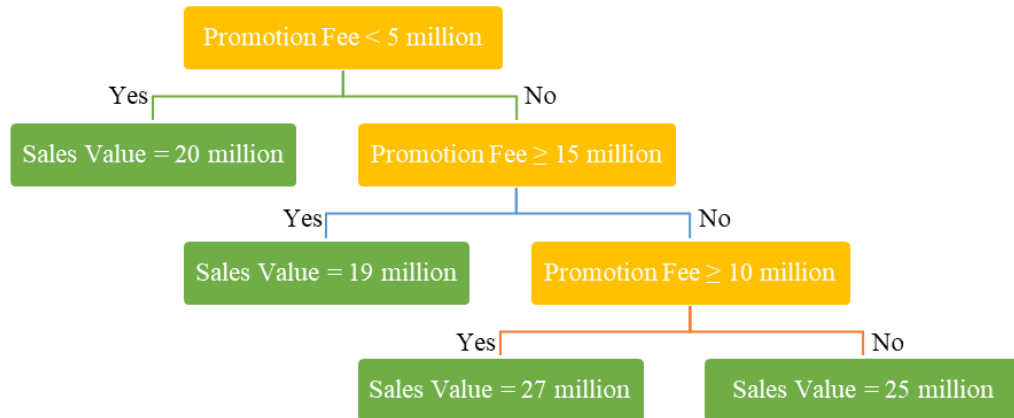
**Figure 1.** Example of a Decision Tree

In this study, the extra-trees method will be used to forecast time-series data, namely the daily closing stock price with the basic autoregressive model. Several studies related to extra-tree decision trees such as the highly random decision tree method (Geurts, et al., 2006), extra-trees regression model for prediction of discharge coefficient in the hydraulic sector (Hameed, et al., 2021), as well as an ensemble cascading extremely randomized tree approach for short-term traffic flow prediction (Zhang, et al., 2019). Several other studies used extra trees and several other tree-based methods as a comparison in modeling phenomena such as prediction of daily precipitation and temperature (Jose, et al., 2022) and prediction of blood cancer (Rupapara, et al., 2022). In other studies, other tree-based methods have been applied to forecast time series data (Lazar & Lazar, 2015) (Rady, et al., 2021) and the application of decision trees as a method for weather forecasting (Kumar, 2013). In several previous studies on decision trees (classification and regression trees), explanatory variables were used as input attributes. In particular, this research utilizes the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) information from the response variables as well as information from the Cross-Correlation Function (CCF) between the response and explanatory variables as input attribute determination. It aims to produce a better output in predicting time series data. Based on these explanations, this study aims to apply a decision tree-based method with an ensemble extra-trees approach to time series data.

## 2. LITERATURE REVIEW

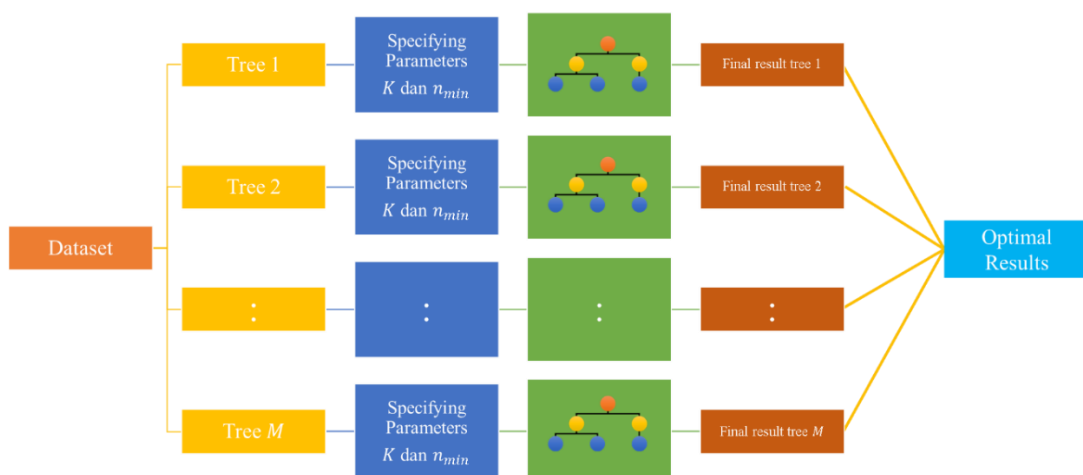
### 2.1. Extra Trees Algorithm

Tree-based methods (regression tree and classification tree) have been widely applied in several studies. The level of accuracy of the allegations and the small error is one of the driving factors for the popularity of this method (Sartono & Syafitri, 2010). In general, the algorithm for the regression tree-based method can be described in Figure 2:



**Figure 2.** Simple Regression Tree

Figure 2 explains how a regression tree can be formed. As an illustration, we want to form a regression tree with the response variable as Sales Value and the predictor variable as Promotional Cost. This study proposes an ensemble method in forming trees, namely the Extra Trees Algorithm. The “Extra Trees” algorithm builds an ensemble method based on a top-down procedure. The fundamental difference from other ensemble methods lies in the process of splitting nodes (Geurts, et al., 2006). Ensemble Tree is one of the developments of a tree-based method with the concept of making a combined tree by utilizing several trees to make assumptions (Sartono & Syafitri, 2010). In general, the Extra Trees algorithm can be described in Figure 3:



**Figure 3.** Extra Trees Algorithm in General

Based on Figure 3, it can be seen that Extra Trees is an Ensemble tree approach with the concept of combining many trees. Merging is done by majority vote for classification cases and the arithmetic average for regression cases (Geurts, et al., 2006). In a simple

procedure, the Extra Trees algorithm can be explained as follows and some important parameters in the algorithm are explained:

- a. Determine the number of trees to be formed based on the parameter  $M$  (the number of ensemble trees to be formed) using the entire learning sample. In principle, the larger value of  $M$ , the better prediction accuracy, but it will affect the computational performance (Geurts, et al., 2006).
- b. In each ensemble tree that is formed, the parameter  $K$  is determined, namely the number of random attributes used in each node. In the case of classification,  $K = n$  is used and for the case of regression,  $K = n$  is used where  $n$  is the number of attributes (Geurts, et al., 2006).
- c. In each ensemble tree that is formed, the parameter  $n_{min}$  is determined, which is the minimum sample used in the splitting process. The greater the value of  $n_{min}$ , the ensemble tree will shrink, and conversely the smaller the value of  $n_{min}$ , the larger the tree formed (Geurts, et al., 2006).
- d. Every ensemble tree that is formed is not pruned, therefore the  $n_{min}$  parameter determines how big the tree will be.

After  $M$  trees are formed, the next step is to carry out the final prediction process with a majority vote for classification cases and the arithmetic average for regression cases.

## 2.2. Accuracy Evaluation

Measurement of prediction accuracy in this study uses Mean Absolute Percentage Error (MAPE) with the following equation (Hyndman & Athanasopoulos, 2021):

$$MAPE = \frac{100\%}{N} \sum_{t=1}^N \left| \frac{A_t - F_t}{A_t} \right| \quad (1)$$

Where  $N$  is the number of learning samples,  $A_t$  is the actual value and  $F_t$  is the forecast value. The accuracy evaluation procedure has developed quite rapidly in the process. One approach that can be used is Time Series Cross-Validation (TSCV) or it can be called the Rolling Forecast Origin procedure. This procedure begins by preparing a series of training and testing data. Forecasting accuracy is calculated by determining the average forecasting error in this case is the MAPE value for each  $h$ , namely the forecast horizon (Hyndman & Athanasopoulos, 2021). The TSCV procedure can be described in Figures 4 and 5.

Origin	Training												Testing							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
12																				
13																				
14																				
15																				
16																				
17																				
18																				

**Figure 4.** Illustration of the Distribution of Training and Testing

Origin	Training												Testing							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
12	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Red	Red	White	White	White	White	White	White
13	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	White	Red	White	White	White	White	White
14	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	White	Red	White	White	White	White
15	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	White	Red	White	White	White
16	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	White	Red	White	White
17	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	White	Red	White
18	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	White	Red

**Figure 5.** Evaluation Process when  $h=2$

Figures 4 and 5 explain how the data set is divided into training and testing groups. Blue observations indicate training data, and red observations indicate testing data. In particular, Figure 5 explains how the evaluation is carried out on the forecast horizon  $h = 2$ . This process can be continued for several other forecast horizons.

### 2.3. Residual Diagnostics

The residual generated by the time series model needs to satisfy several assumptions. This assumption includes the absence of autocorrelation between the lags in the residual model that is formed. One of the tests that can be done is the Ljung-Box test. If the residual is denoted as  $e_t = y_t - \hat{y}_t$  then the test statistics for the Ljung-Box test are as follows (Wei, 2006):

$$Q = n(n + 2) \sum_{k=1}^K (n - k)^{-1} \hat{\rho}_k^2 \quad (2)$$

where  $K$  is the maximum lag specified,  $n$  is the number of observations and  $\hat{\rho}_k$  is the autocorrelation at lag  $k$ . The Ljung-Box test has a hypothesis  $H_0$  that no residual autocorrelation between lags.

## 3. MATERIAL AND METHOD

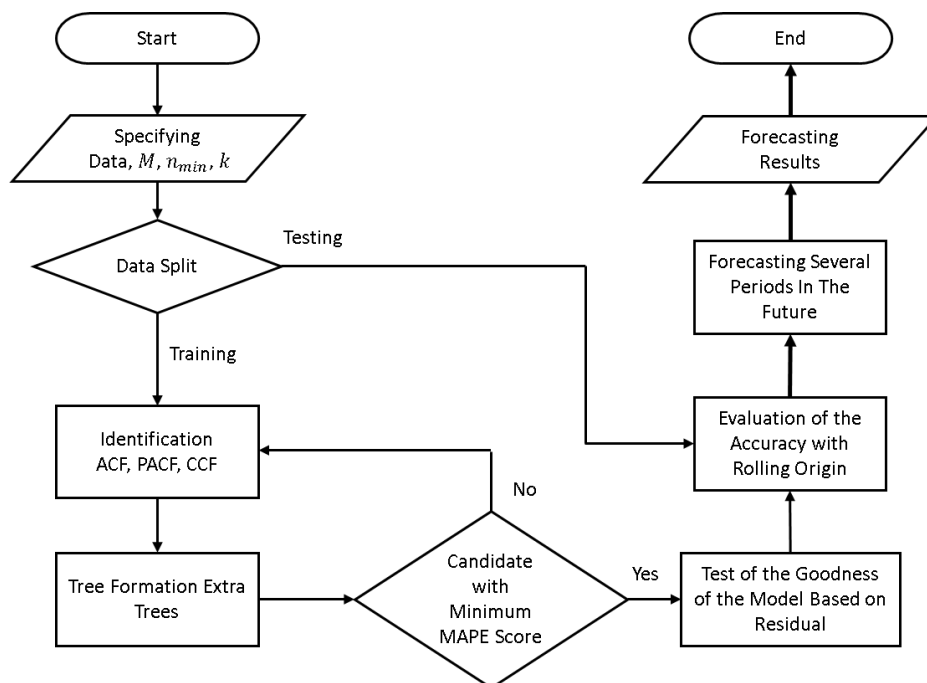
### 3.1. Data

The dataset used in this study is information related to the closing stock price and the number of stock transactions of PT. Bumi Resource Mineral Tbk with BRMS code. The time range used is daily from January 2, 2020, to December 30, 2021. Bumi Resource Mineral is one of the companies with a mining focus. BRMS's Operational Business Unit consists of Dairi Prima Minerals: Zinc and Lead, Citra Palu Minerals: Gold and Molybdenum, and Gorontalo Minerals: Gold and Copper (Bumi Resource, 2003). The main objective of this research is to use the Extra Trees approach for data with Autoregressive cases. In general, several stages used in this research are identifying data, conducting the process of forming Extra Trees, obtaining the best model, evaluating the accuracy, and forecasting BRMS stock prices for some time to come.

### 3.2. Data Analysis Procedure

There are several procedures carried out in the data analysis process to obtain the expected final results. This study uses the R software tool in processing the data and the package used is extraTrees. In general, the procedure is described in Figure 6. Some of the stages shown can be explained as follows:

- a. Prepare the data to be analyzed along with determining the parameters of extra trees to be used, including  $M$ ,  $n_{min}$ , and  $k$ .
- b. Dividing data into training data and testing data.
- c. Identify the form of the relationship as input consideration based on the training data. This identification includes the establishment of Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and Cross-Correlation Function (CCF).
- d. Creating an extra trees ensemble tree using the extraTrees package available in R.
- e. Obtain the best candidate model by comparing MAPE values.
- f. Test the goodness of the model based on the residual model of the training data.
- g. Evaluating the accuracy using Rolling Forecast Origin.
- h. Get the conclusion.

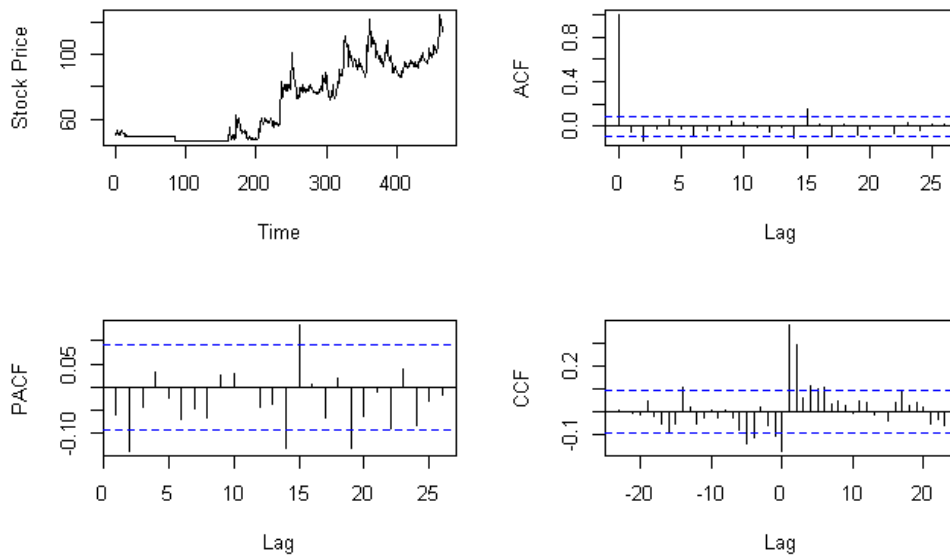


**Figure 6.** Data Analysis Procedure

## 4. RESULTS AND DISCUSSION

### 4.1. Identification

The first step in this research is to divide the data into training and testing with 95% for training and 5% for testing. The main variables used are stock closing price ( $Y$ ) and stock transaction volume ( $X$ ). The next step is to identify to determine the appropriate input for the formation of the model in the training data section. This stage is done by calculating the Autocorrelation Function (ACF) of  $Y$ , Partial Autocorrelation Function (PACF) of  $Y$ , and Cross-Correlation Function (CCF) values between  $X$  and  $Y$ . The identification results are described in Figure 7.



**Figure 7.** Data Identification

Figure 7 explains the time series plot of the stock price (Y), the stationary stock price ACF (Y), the stationary stock price PACF (Y), and the CCF between the transaction volume (X) and the stationary stock price (Y). Based on the identification, it was decided that some of the attributes used in tree formation are described in table 1.

**Table 1.** Attribute Determination

Name	Attribute	Description
Stock Closing Price	$Y_t$	Target
Stock Closing Price (lag $t - 1$ )	$Y_{t-1}$	
Stock Closing Price (lag $t - 2$ )	$Y_{t-2}$	
Stock Closing Price (lag $t - 3$ )	$Y_{t-3}$	
Stock Transaction Volume	$X_t$	Input
Stock Transaction Volume (lag $t - 1$ )	$X_{t-1}$	
Stock Transaction Volume (lag $t - 2$ )	$X_{t-2}$	
Stock Transaction Volume (lag $t - 3$ )	$X_{t-3}$	

#### 4.2. Extra Trees

Based on some of the attributes described in table 1, experiments will be conducted for several possible optimum attributes by comparing the MAPE values from the training data. Because the Extra Trees algorithm is random in tree formation and node determination, each possible combination of attributes (model) will be repeated 50 times. This repetition aims to obtain consistent results. The number of trees used in this study is 500 with the notation  $M = 500$ . In particular, the larger the M value, the more accurate the results obtained (Geurts, et al., 2006). In this study, the number of trees is limited to 500. Parameter M is also closely related to computational performance in a tree formation. The number of random attributes used in each experiment is  $K = n$  (all input attributes). This research basically uses the principle of a regression tree. In the regression tree, by default the value of K is all input attributes (Geurts, et al., 2006). In this study, the minimum sample used in the separation process uses the default minimum sample for the regression tree with a value of  $n_{min} = 5$  (Geurts, et al., 2006).

### 4.3. Best Model Selection

After it was repeated 50 times for each model used, the next step was to determine the average MAPE value to get the best model. The results of the comparison of models are described in table 2.

**Table 2.** Comparison of Average MAPE

No	Target	Input	Average MAPE
1	$Y_t$	$X_t, Y_{t-1}$	1.2187%
2	$Y_t$	$X_t, Y_{t-1}, Y_{t-2}$	1.0002%
3	$Y_t$	$X_t, Y_{t-1}, Y_{t-2}, Y_{t-3}$	0.8706%
4	$Y_t$	$X_t, X_{t-1}$	6.2090%
5	$Y_t$	$X_t, X_{t-1}, X_{t-2}$	4.9446%
6	$Y_t$	$X_t, X_{t-1}, X_{t-2}, X_{t-3}$	4.1273%
7	$Y_t$	$X_t, Y_{t-1}, X_{t-1}$	0.8000%
8	$Y_t$	$X_t, Y_{t-1}, X_{t-1}, X_{t-2}$	0.6925%
9	$Y_t$	$X_t, Y_{t-1}, X_{t-1}, X_{t-2}, X_{t-3}$	0.6469%
10	$Y_t$	$X_t, Y_{t-1}, Y_{t-2}, X_{t-1}$	0.7198%
11	$Y_t$	$X_t, Y_{t-1}, Y_{t-2}, X_{t-1}, X_{t-2}$	0.6254%
12	$Y_t$	$X_t, Y_{t-1}, Y_{t-2}, X_{t-1}, X_{t-2}, X_{t-3}$	0.5930%
13	$Y_t$	$X_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, X_{t-1}$	0.6585%
14	$Y_t$	$X_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, X_{t-1}, X_{t-2}$	0.5889%
15	$Y_t$	$X_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, X_{t-1}, X_{t-2}, X_{t-3}$	0.5563%

Based on table 2, three candidate models will be taken that have the smallest average MAPE value. The three candidate models are 12, 14, and 15 with a MAPE value of less than 0.6%.

### 4.4. Model Diagnostic Test

**Table 3.** Ljung Box Test

Lag	Ljung Box Test		
	Model 12 <i>p-value</i>	Model 14 <i>p-value</i>	Model 15 <i>p-value</i>
1	0.76471	0.45769	0.23275
2	0.00909	0.05652	0.04597
3	0.02437	0.11358	0.09277
4	0.03964	0.12612	0.12484
5	0.05027	0.20660	0.20166
6	0.06529	0.25173	0.22764
7	0.06802	0.25806	0.23578
8	0.08519	0.28256	0.29219
9	0.07788	0.29634	0.32510
10	0.09804	0.37873	0.41195

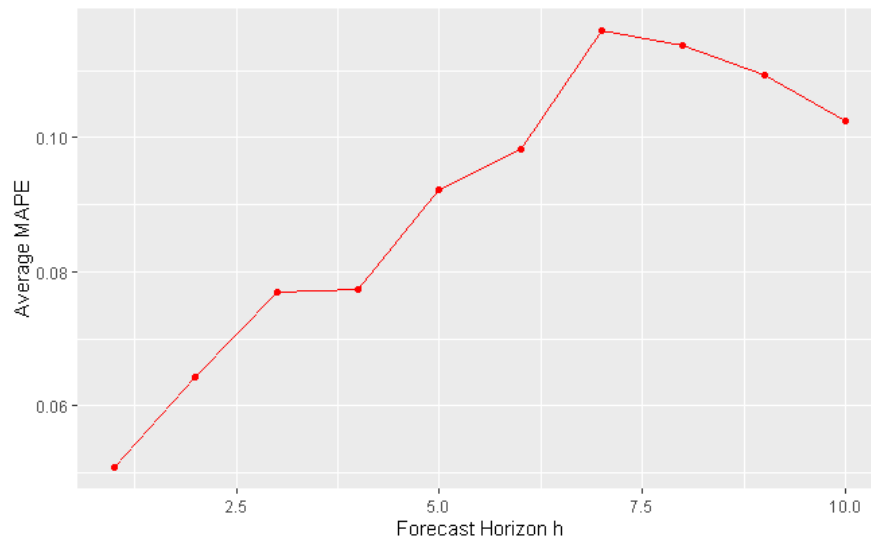
The goodness of the time series model is used to ensure that the residuals from the formed model meet the white noise requirements. The goodness of the model used at this stage is the Ljung-Box test with the hypothesis  $H_0$  that there is no residual autocorrelation between lags. and a significance level of 5%. The results of the Ljung-Box test are shown in table 3.



Based on table 3, it is concluded that model 14 is the best model obtained. This is because the  $p - value > \alpha$  for each lag and causes  $H_0$  to fail to be rejected or there is no residual autocorrelation between lags.

#### 4.5. Prediction Accuracy Evaluation

In the process of evaluating prediction accuracy, the TSCV Rolling Origin approach is used with a forecast horizon of  $h = 10$ . The results obtained are shown in Figure 8.

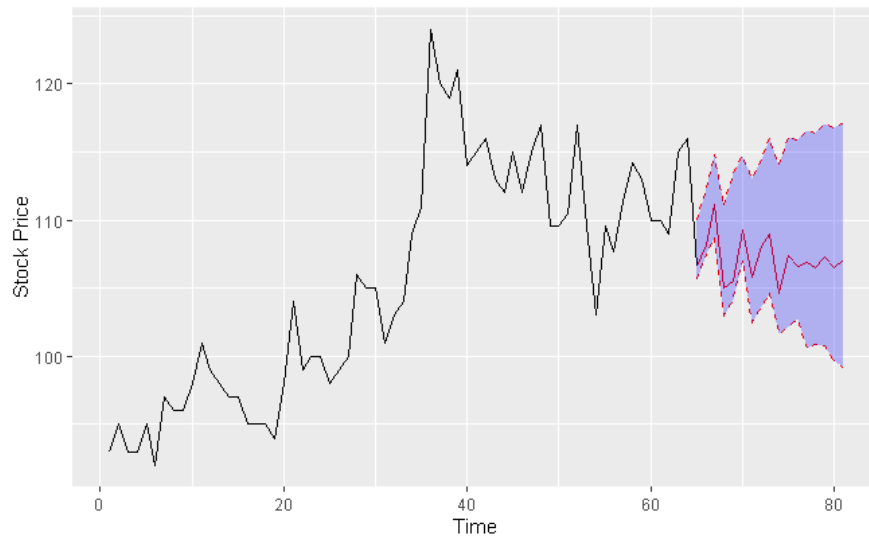


**Figure 8.** TSCV results for  $h = 10$

Figure 8 shows how accurate the predictions are based on MAPE for each forecast horizon. These results indicate that the Extra Trees approach for time series data on the closing stock price of BRMS is accurate enough to predict up to 6 days after training data with a MAPE of less than 0.1%. Meanwhile, the following period tends to produce a large MAPE.

#### 4.6. Forecasting Several Future Periods

The next step is to forecast for the next several periods with the condition that the actual input data is not yet known. This forecast will be carried out for the closing stock price of BRMS in January 2022 with the illustration that the actual closing price data is not yet known. The results of this forecast are shown in Figure 9.



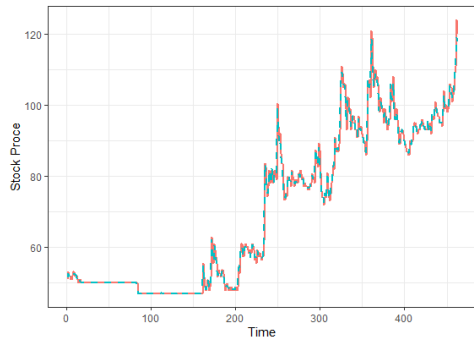
**Figure 9.** Forecasting Results for the Next Several Periods

Figure 9 shows how the forecasting results will be in January 2022 if the input is not known and fully uses the input forecasting results in the previous period. The solid black line is the actual value from October 2021 to December 2021. While the solid red line is the forecast value, the red dashed line and the blue area are the 95% prediction interval. Figure 10 shows the 95% prediction interval enlarges as the forecasting period increases. The point forecast results for the next 18 periods tend to form a constant pattern for a long forecasting period.

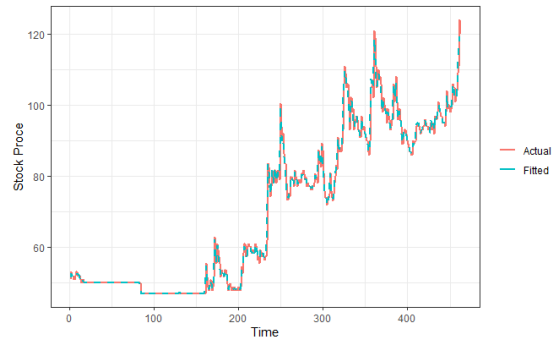
#### 4.7. Discussion of Results

In particular, the idea proposed in this research is how to apply one of the ensemble tree approaches, namely extra tree in making predictions on univariate data and by considering other variables. Based on the results obtained, the prediction accuracy is very good, this can be seen from the small MAPE value. For example, the two best candidate models (which have the smallest MAPE value) from table 2 are model 14 with input attributes  $X_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, X_{t-1}, X_{t-2}$  and model 15 with input attributes  $X_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, X_{t-1}, X_{t-2}, X_{t-3}$  has MAPE 0.5889% and 0.5563% respectively. The MAPE value is very small, especially supported by the comparison plot between the actual and fitted data in Figures 10 and 11. At the model diagnosis stage, from the three best candidate models, namely models 12, 14 and 15, model 14 was chosen as the best model based on the Ljung Box test.

These results indicate that in this study, the smallest MAPE value (model 15) does not necessarily meet the goodness of the model based on the diagnostic test stage. Based on the evaluation of prediction accuracy, these results show that this approach can produce a fairly good prediction (MAPE less than 0.1%) in 6 periods after the actual data. In the final results, prediction intervals are added to support decision making and interpretation of results. The prediction interval is based on the bootstrapping process with 1000 iterations with the results in Figure 9.



**Figure 10.** Comparison of Actual and Fitted on Model 14



**Figure 11.** Comparison of Actual and Fitted on Model 15

## 5. CONCLUSION

Based on the results and discussion that have been described, it can be concluded that for forecasting the closing stock price of BRMS, the Extra Trees algorithm approach shows good results for models with a combination of inputs  $X_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, X_{t-1}, X_{t-2}$  with the average MAPE value for 50 trials is 0.5889% with white noise error. It can be explicitly said that the closing stock price of BRMS will be influenced by the previous price up to 3 days ago and the previous transaction volume up to 2 days ago. In evaluating the prediction results using TSCV, it is concluded that the Extra Trees algorithm approach can produce predictions that are quite accurate for predictions for the next 6 days after training data with a MAPE of less than 0.1%. Meanwhile, the following period tends to produce a large MAPE. The last conclusion is to forecast several future periods with unknown inputs and fully use the input forecasting results in the previous period. The results of point forecasting values tend to approach a constant pattern for a long forecasting period. Meanwhile, the 95% prediction interval enlarges as the forecasting period increases.

## REFERENCES

- Agustina & Sumartio, F., 2014. Analisis Faktor-faktor yang Mempengaruhi Pergerakan Harga Saham pada Perusahaan Pertambangan. *Jurnal Wira Ekonomi Mikroskil*, 4(1).
- Astutik, S. R. P., Sukestiyarno & Hendikawati, P., 2018. *Peramalan Inflasi di Demak Menggunakan Metode ARIMA Berbantuan Software R dan MINITAB*. Semarang, s.n.
- Banarjee, M., Reynolds, E., Andersson, H. B. & Nallamouthu, B., 2019. Tree-Based Analysis: A Practical Approach to Create Clinical Decision Making Tools. *Circ Cardiovasc Qual Outcomes*, 12(5).
- Bumi Resource, 2003. *PT BUMI Resources Tbk..* [Online] Available at: <http://www.bumiresources.com/en/about-us/subsidiaries/detail/test> [Accessed 1 Maret 2022].
- Fitriyani, F., A, S. F., R, M. I. & T, T. A., 2021. Peramalan Indeks Harga Saham PT Verena Multi Finance Tbk Dengan Metode Pemodelan ARIMA Dan ARCH-GARCH. *J Statistika*, 14(1), pp. 11-23.
- Geurts, P., Ernst, D. & Wehenkel, L., 2006. Extremely randomized trees. *Mach Learn*, Volume 63, pp. 3-42.
- Hameed, M. M., AlOmar, M. K., Khaleel, F. & Al-Ansari, N., 2021. An Extra Tree Regression Model for Discharge Coefficient Prediction: Novel, Practical

- Applications in the Hydraulic Sector and Future Research Directions. *Hindawi Mathematical Problems in Engineering*, Volume 2021.
- Hyndman, R. J. & Athanasopoulos, G., 2021. *Forecasting: Principles and Practice (3rd ed)*. Melbourne: OTexts.
- Ispriyanti, D., Prahutama, A. & M., 2019. Klasifikasi Kemiskinan Di Kota Semarang Menggunakan Algoritma Chisquare Automatic Interaction Detection (Chaid) Dan Classification And Regression Tree (Cart). *Media Statistika*, 12(1), pp. 63-72.
- Jatmiko, Y. A., Padmadisastra, S. & Chadidjah, A., 2019. Analisis Perbandingan Kinerja Cart Konvensional, Bagging Dan Random Forest Pada Klasifikasi Objek: Hasil Dari Dua Simulasi. *Media Statistika*, 12(1), pp. 1-12.
- Jose, D. M., Vincent, A. M. & Dwarakish, G. S., 2022. Improving multiple model ensemble predictions of daily precipitation and temperature through machine learning techniques. *Sci Rep*, 12(4678).
- Kumar, R., 2013. Decision Tree for the Weather Forecasting. *International Journal of Computer Applications*, 76(2).
- Kurnia, M. T., Nugrahani, E. H. & Sumarno, H., 2014. Analisis Wavelet dan ARIMA untuk Peramalan Harga Emas PT. ANTAM Tbk. Indonesia. *Journal of Mathematics and Its Applications*, 13(2).
- Lazar, C. & Lazar, M., 2015. Using the Method of Decision Trees in the Forecasting Activity. *Economic Insights – Trends and Challenges*, 4(1), pp. 41-48.
- Mahkya, D. A. & Anggraini, D., 2019. *Forecasting the Number of Passengers from Bakauheni Port during the Sunda Strait Tsunami Period Using Intervention Analysis Approach and Outlier Detection*. Lampung Selatan, s.n.
- Patel, H. H. & Prajapati, P., 2018. Study and Analysis of Decision Tree Based Classification Algorithms. *International Journal of Computer Sciences and Engineering*, 6(10).
- Rady, E. H. A., Fawzy, H. & Fattah, M. A., 2021. Time Series Forecasting Using Tree Based Methods. *Journal of Statistics Applications & Probability*, 10(1), pp. 229-244.
- Rezaldi, D. A. & Sugiman, 2021. *Peramalan Metode ARIMA Data Saham PT. Telekomunikasi Indonesia*. Semarang, s.n.
- Rupapara, V. et al., 2022. Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model. *Sci Rep*, 12(1000).
- Saham OK, 2011. *Saham OK*. [Online] Available at: <https://www.sahamok.net/emiten/sektor-bei/> [Accessed 22 Maret 2022].
- Saini, P., Rai, S. & Jain, A. K., 2014. Decision Tree Algorithm Implementation Using Educational Data. *International Journal of Computer-Aided technologies (IJCAx)*, 1(1).
- Sartono, B. & Syafitri, U. D., 2010. Metode Pohon Gabungan: Metode Gabungan untuk Mengatasi Kelemahan Pohon Regresi dan Klasifikasi Tunggal. *Forum Statistika dan Komputasi*, 15(1), pp. 1-7.
- Sulia, 2017. Analisis Faktor-faktor yang Mempengaruhi Harga Saham pada Perusahaan LQ45 yang Terdaftar di Bursa Efek Indonesia. *Jurnal Wira Ekonomi Mikroskil*, 7(2).
- Wei, W. W. S., 2006. *Time Series Analysis Univariate and Multivariate Method*. New York: Pearson Education.
- Zhang, F. et al., 2019. An Ensemble Cascading Extremely Randomized Trees Framework for Short-Term Traffic Flow Prediction. *KSII Transactions on Internet and Information Systems*, 13(4).