BAREKENG: Journal of Mathematics and Its Applications
March 2026 Volume 20 Issue 1 Page 0573–0584
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss1pp0573-0584

FORECASTING INDONESIA COMPOSITE INDEX USING HYBRID AUTOREGRESSIVE INTEGRATED MOVING AVERAGE-DOUBLE RANDOM FOREST MODEL

Andika Putri Ratnasari [1*], Luthfia Hanun Yuli Arini [2]

[1,2] *Department of Mathematics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta*
Jln. Colombo No.1, Sleman, Yogyakarta, 55281, Indonesia

*Corresponding author's e-mail: * **andikaputriratnasari@uny.ac.id***

Article Info

Article History:
Received: 21st April 2025
Revised: 29th May 2025
Accepted: 29th June 2025
Available online: 24th November 2025

Keywords:
ARIMA;
Double random forest;
Forecast;
Hybrid model;
Time series.

ABSTRACT

Modeling time series data using autoregressive integrated moving average (ARIMA) has been widely discussed. However, this has limitations in that it can only handle linear data. Machine learning is one of the alternative approaches that can solve this limitation since this method can handle nonlinear cases. Double random forest (DRF) is considered a supervised learning method that can solve regression problems. This research provides a novel hybrid forecasting framework combining ARIMA and DRF, designed to model both linear and nonlinear behaviors, and provide more accurate predictions for volatile financial data like the Indonesia Composite Index (ICI). Previous studies have not examined the performance of the hybrid ARIMA-DRF model. In this study, the performance of ARIMA, DRF, and the hybrid ARIMA-DRF models is compared using ICI data obtained from Bank Indonesia's website. ICI has nonstationary and nonlinear characteristics. This made the ICI data suitable to be modeled using the hybrid ARIMA-DRF model. The comparison results indicate that the hybrid ARIMA-DRF model outperforms the independent ARIMA and DRF models, with a value of its mean absolute percentage error is 4.17%. Therefore, forecasting the future value of ICI data was done by using a hybrid ARIMA-DRF model with forecasting periods from October 2023 to September 2024. The forecasting results show that ICI values fluctuate over the forecasting periods, hence the authority might use the pattern to predict the ICI behaviors and take further decisions. While the forecasting results offer valuable insights for decision-making, this study has limitations as it does not incorporate external macroeconomic variables that may influence ICI behavior.

How to cite this article:

A. P. Ratnasari and L. H. Y. Arini, "FORECASTING INDONESIA COMPOSITE INDEX USING HYBRID AUTOREGRESSIVE INTEGRATED MOVING AVERAGE-DOUBLE RANDOM FOREST MODEL", *BAREKENG: J. Math. & App.*, vol. 20, iss. 1, pp. 0573-0584, Mar, 2026.

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

Time series data have many behaviors. It means that each time series problem cannot be solved with a single technique. Every problem might be solved with different approaches. One of the most popular time series models is Autoregressive Integrated Moving Average (ARIMA). ARIMA belongs to a nonstationary time series model [1]. A previous study compared the ARIMA and Long Short-Term Memory (LSTM) models in forecasting NASDAQ stock exchange rate data. This study obtained that ARIMA performs better than LSTM [2]. Furthermore, the application of ARIMA in modeling Indonesia's export data obtains good performance forecasting accuracy. The results show that the ARIMA model has a smaller than 10% mean absolute percentage error (MAPE) value [3]. Despite its good performance in previous research, ARIMA also has weaknesses. The weakness of ARIMA is its high error value if the normality and linearity in the data are not fulfilled [4].

The use of machine learning to model time series data has been widely studied recently. Machine learning has many advantages in modeling time series data. One of the advantages is that this method can solve nonlinear problems that classical time series cannot handle [5]. Double Random Forest (DRF) is a machine learning model that can be used to solve regression and classification problems [6]. The performance of DRF in modeling time series data has been studied using economic indicator data that have nonlinear behavior. The results of this study show that DRF can give high forecasting accuracy [7].

Most financial time series data have nonlinear behavior which the variance and mean of the series change over time. Indonesia Composite Index (ICI) is one of the factors that is considered by investors to invest in the Indonesia Stock Exchange (IDX). This happens because the changing of ICI data in IDX becomes a general benchmark of Indonesia’s stock condition. It is also a general indicator of all stock prices changing that are listed in IDX [8]. Therefore, this is important to forecast the future value of ICI data.

ICI data have high fluctuations, and they tend to have nonlinear behavior. The hybrid ARIMA and machine learning model can solve linear and nonlinear problems. Previous research developed and compared three hybrid models for forecasting stock index returns: ARIMA-SVM, ARIMA-ANN, and ARIMA-RF. Their study was motivated by the complex, dynamic, and nonlinear nature of financial time series, which often limits the effectiveness of traditional linear models like ARIMA. To address this, they combined ARIMA (to capture linear patterns) with machine learning methods such as Support Vector Machine (SVM), Artificial Neural Network (ANN), and Random Forest (RF) (to capture nonlinear patterns) [9].

Another previous study demonstrated that hybrid models combining linear and nonlinear techniques significantly improve stock market forecasting accuracy compared to individual models. Their results showed that the ANN-ARIMA hybrid, which integrates the strengths of both linear and nonlinear methods, achieved lower forecasting errors and better trend prediction across major Indian stock indices [10]. In addition, a hybrid forecasting approach that combines ARIMA for capturing linear structures and Support Vector Machines (SVM) for modeling nonlinear patterns has also been shown to outperform both individual ARIMA and SVM models, yielding higher forecasting accuracy. These findings highlight that integrating linear and nonlinear modeling techniques can effectively handle the complexity of financial time series and lead to superior prediction performance [11].

Despite the growing interest in hybrid forecasting models, no prior study has explored the integration of ARIMA with the Double Random Forest algorithm. This research aims to fill that gap by developing a novel hybrid ARIMA-DRF model. The novelty of this study lies in combining ARIMA’s strength in capturing linear patterns with DRF’s robust capability in modeling complex nonlinearities, especially in volatile financial data like the ICI. Unlike previous hybrid models, the proposed ARIMA-DRF model is expected to offer improved adaptability and forecasting accuracy due to the ensemble nature and enhanced variance control of DRF. Therefore, the main research question addressed in this study is: Can a hybrid ARIMA-DRF model provide better forecasting performance for the Indonesia Composite Index compared to an individual ARIMA or DRF model.

This study compares the performance of ARIMA, DRF, and the hybrid ARIMA-DRF models in forecasting ICI data. The models are evaluated using three accuracy metrics: Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Deviation (MAD). The best-performing model is then applied for future ICI forecasting, aiming to contribute a novel, reliable forecasting approach for complex financial time series. In addition, the proposed model is expected to provide valuable insights for investors and policymakers by offering more accurate predictions of stock market movements. By addressing both linear and nonlinear components effectively, this study seeks to enhance the robustness of time series forecasting, particularly for volatile financial indicators like the ICI.

2. RESEARCH METHODS

2.1 Data Description and Procedure of Data Analysis

The data used in this study are the monthly Indonesia Composite Index (ICI) data from January 2008 to September 2023. The dataset was obtained from the official website of Bank Indonesia and can be freely downloaded through their online database. ICI data were chosen for this study because they serve as a key benchmark of the overall performance of the Indonesian stock market. The ICI reflects investor sentiment and market dynamics, making it an essential indicator for financial forecasting. Moreover, its availability as an open-source dataset ensures transparency and reproducibility in financial research.

Prior to model implementation, the data were divided into training and testing subsets. The training dataset spans from January 2008 to December 2020, while the testing dataset covers the period from January 2021 to September 2023. This partitioning ensures that the models are trained on past trends while being validated on recent data. This partition was created by considering the data pattern.

To handle the nonstationary nature of the ICI data, differencing was applied to transform the series into a stationary process. The modeling process involved the construction of ARIMA, DRF, and hybrid ARIMA-DRF models. The optimal ARIMA model was selected based on the Box-Jenkins methodology, while the DRF model's hyperparameters were fine-tuned using a grid search cross-validation (CV) method. The hybrid ARIMA-DRF model was developed by first modeling the linear components with ARIMA and then capturing the residual nonlinear patterns using DRF. Model evaluation was performed using three accuracy metrics: Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Deviation (MAD). The model with the lowest error values was selected for forecasting ICI data.

2.2 Autoregressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) is one of the univariate time series models [12]. This model was proposed by Box and Jenkins in 1976 [13]. ARIMA has three components. The first component is AR (p), which is a regression model that takes advantage of correlations between observations and previous observations. p is the order of the AR component. Furthermore, there is an integrated component with d order. This is a stationary process of data using differencing. The last component is MA (q), MA represents the residuals autocorrelation [14]. The model of ARIMA (p, d, q) is written in Eq. (1) [15].

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B) e_t \tag{1}$$

In Eq. (1), B is the backshift operator, which shifts the value of the observation one step backward, i.e., $BY_t = Y_{t-1}$. The symbol $\phi_p(B)$ is the polynomial of the AR component with order p, written as $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, where ϕ_i are the autoregressive parameters. The symbol $(1-B)^d$ denotes the differencing operator applied d times. $\theta_q(B)$ is the polynomial of the MA component with order q, written as $\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$, where θ_i are the moving average parameters. Then, e_t is the residual or white noise at time t, assumed to follow a normal distribution with zero mean and constant variance, i.e., $e_t \sim N(0, \sigma^2)$. The fitting processes of an ARIMA model are identifying the tentative models, estimating the model parameters, carrying out diagnostic checking, and validating the model [16].

2.3 Double Random Forest (DRF)

The proposition of Double Random Forest (DRF) was made in 2020 by S. Han, H. Kim, and Y. Lee with the aim of enhancing the performance of Random Forest (RF). DRF diverges from RF in the sense that it constructs a regression tree based on all of the training data rather than employing bootstrapping as a preliminary step. This deviation has an impact on the size of the tree generated by DRF, with a larger tree size potentially leading to a reduction in prediction bias [6].

The regression case of data modeling using DRF commences by creating a regression tree using the training data, followed by the execution of four stages for the identification of optimal splits. The initial stage involves bootstrapping a random sample when the number of samples in a node is less than 10% of the training data [6].

The selection of independent variables ($m \approx {}^{p}/_{3}$) is performed in the second stage in a random manner. The third stage involves the determination of the most suitable splitting criteria based on the consideration of the mean square error value. The fourth stage entails the repetition of the first, second, and third stages until the fulfillment of the stopping criterion. The process of constructing this regression tree is iterated until b regression trees have been created. The final step involves the determination of the prediction results of the DRF model through the averaging of the predictions made by each regression tree [6].

In summary, the DRF algorithm for regression involves the following main steps [6]:

1. Initial tree construction, where a regression tree is created by using the complete training dataset.
2. Selection of optimal splits, which is performed iteratively until the stopping criteria for tree growth are satisfied.
3. Repetition of steps 1 and 2 until b regression trees are generated by the DRF algorithm.
4. Final prediction, which is obtained by averaging the outputs of all b regression trees.

This DRF model has been applied by Ratnasari et al. (2023). This study used DRF to analyze the economic indicator data. The results showed that DRF had good forecasting ability in analyzing these data [7].

2.4 Hybrid ARIMA-DRF

The pattern of ICI data cannot be easily captured. The hybrid ARIMA-DRF model, which combines linear and nonlinear capabilities, presents itself as a viable solution for effectively modeling data exhibiting both characteristics. This will be a good alternative for modeling data with both characteristics. The hybrid model (Z_t) is written in Eq. (2).

$$Z_t = Y_t + N_t \tag{2}$$

with Y_t represents the linear part and N_t represents the nonlinear part of the hybrid model. The linear part and the nonlinear part are estimated from the training data. $\hat{Y}_t$ is the forecasted value at time t from the ARIMA model. The residuals from the ARIMA model (e_t) can be written in Eq. (3).

$$e_t = Z_t - \hat{Y}_t. \tag{3}$$

The residuals are modeled using the DRF model. The model of residuals can be written in Eq. (4).

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + a_t \tag{4}$$

with f represents the nonlinear function modeled by the DRF and a_t represents the random error. The forecast of hybrid ARIMA-DRF is presented in Eq. (5).

$$\hat{Z}_t = \hat{Y}_t + \hat{N}_t \tag{5}$$

with $\hat{N}_t$ is the forecasted value of the DRF model [11].

Below are the key steps of the ARIMA-DRF modeling framework:

1. Creating the linear model (ARIMA).
2. Calculating residuals from the ARIMA model.
3. Modeling residuals using Double Random Forest (DRF)
4. Forecasting the nonlinear component.
5. Combining linear and nonlinear forecasts.

2.5 Forecast Accuracy Metrics

The model performance is compared using three indices, which are MAPE, RMSE, and MAD. The formula of MAPE is written in Eq. (6) [17].

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_a - Y_f}{Y_a}\right| \times 100\,. \tag{6}$$

The next metric is RMSE and MAD. The formula of RMSE is written in Eq. (7) [18].

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_a - Y_f)^2}. \tag{7}$$

The formula of MAD is written in Eq. (8) [19].

$$MAD = \frac{1}{n}\sum_{t=1}^{n}\left|Y_a - Y_f\right|. \tag{8}$$

Y_a represents the actual value of time series data, n denotes the number of time series observations, and Y_f is the forecasted value.

2.6 Terasvirta Test for Nonlinearity Testing

The Terasvirta test is a statistical procedure used to detect nonlinear structures in time series data. The test can be conducted using the chi-square distribution, with the null hypothesis is the time series exhibits linear behavior. The chi-square-based procedure involves the following steps [20]:

1. Regress Y_i on the predictors $1, X_1, X_2, \dots X_p$ and compute the residuals.
2. Construct an auxiliary regression by regressing the residuals from step 1 on the original predictors and m additional predictors (quadratic and cubic component values of Taylor expansion).
3. Calculate the coefficient of determination (R^2) from step 2.
4. Compute the chi-square test statistics by using $\chi^2 = n.R^2$, where n is the number of observations.
5. Compare χ^2 to the critical value from the chi-square distribution with m degrees of freedom. If $\chi^2 > \chi^2_m$, then reject the null hypothesis. Otherwise, fail to reject H_0.

3. RESULTS AND DISCUSSION

3.1 Data Exploration

The plot of ICI data is shown in Fig. 1. According to the plot, ICI data have a nonstationary pattern, tend to increase, and fluctuate over time. The nonlinearity testing was also carried out. The Terasvirta test was used to check the nonlinearity of ICI data. The results showed that the p-value of the Terasvirta test was less than the significance level 0.05, indicating that the ICI data exhibits nonlinearity.

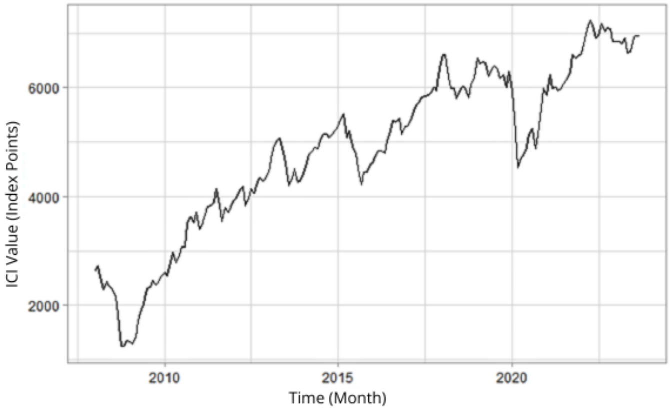

Figure 1. Plot of ICI Data from January 2008 to September 2023.
Source: Bank Indonesia

According to Fig. 1, ICI data are not stationary. Therefore, the data need to be transformed into stationary using differencing before the data are analyzed. The stationarity of data can be detected by using the Augmented Dickey-Fuller (ADF) test and the plot of the data.

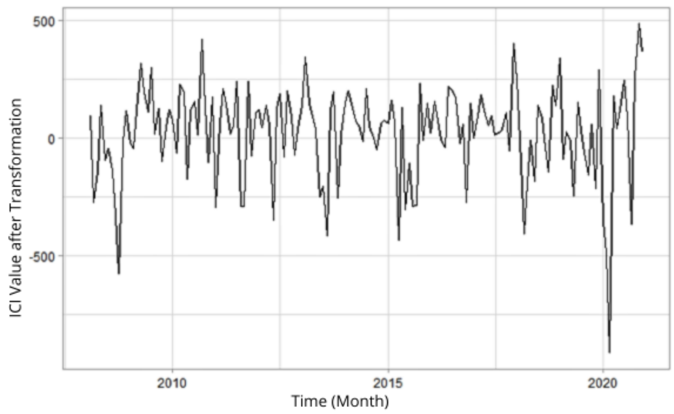

Figure 2. Plot of ICI Data after Differencing from January 2008 to December 2020

Based on the ADF test of transformed data, the p-value result is 0.01. By using a 0.05 significance level, it can be concluded that the data is stationary after transformation. Fig. 2 also shows that after one time differencing, the data show a stationary pattern.

3.2 ARIMA Model for ICI Data

Plots of the autocorrelation function (ACF) and partial autocorrelation function (PACF) are used to identify the order of the tentative ARIMA (p, d, q) models. The ACF plot shown in Fig. 3 is used to specify the order of MA (q), while the PACF plot shown in Fig. 4 is used to specify the order of AR (p). The d symbol represents the order of differencing [1].

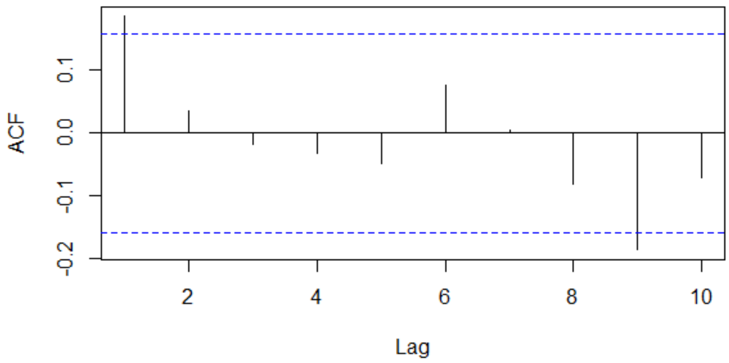

Figure 3. ACF Plot

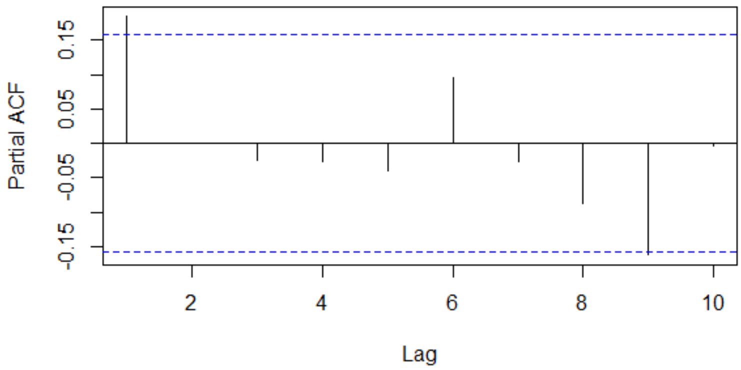

Figure 4. PACF Plot

The ACF plot, as shown in Fig. 3, displays significant spikes outside the confidence bounds (dashed blue lines), indicating the presence of moving average (MA) components. A sharp cutoff after lag 1 suggests that the potential order of the MA term is MA(1). Meanwhile, the PACF plot, shown in Fig. 4, exhibits a significant spike at lag 1 followed by a sharp decline, indicating the presence of autoregressive (AR) components. Thus, the potential order of the AR term is AR(1).

Since Figs. 3 and 4 show that the value of autocorrelation and partial autocorrelation cut off after the first lag. Therefore, the tentative ARIMA models are ARIMA (1, 1, 1), ARIMA (1, 1, 0), and ARIMA (0, 1, 1). The best ARIMA model is determined by considering the smallest Akaike Information Criterion (AIC) score and significance test of parameters, which are shown in Table 1.

Table 1. AIC Value of Each Tentative Model

Model	AIC
ARIMA (1, 1, 1)	2088.259
ARIMA (1, 1, 0)	2086.274
ARIMA (0, 1, 1)	2086.618

According to Table 1, ARIMA (1, 1, 0) yields the lowest AIC value. Additionally, the significance tests indicate that all parameters in the model have p-values below 0.05. Therefore, it can be concluded that ARIMA (1, 1, 0) is the best model for ICI data. The model of ARIMA (1, 1, 0) is written in Eq. (9).

$$\hat{Y}_t = 1.195Y_{t-1} - 0.195Y_{t-2} \tag{9}$$

The diagnostic checking is carried out to check the normality of residuals, the homogeneity of residual variance, and the autocorrelation of residuals. The results show that ARIMA (1, 1, 0) has met all the model assumptions. Furthermore, the model is validated using the testing data. The MAPE, RMSE, and MAD values resulting from the ARIMA (1, 1, 0) model are 9.12%, 722.54, and 627.63, respectively.

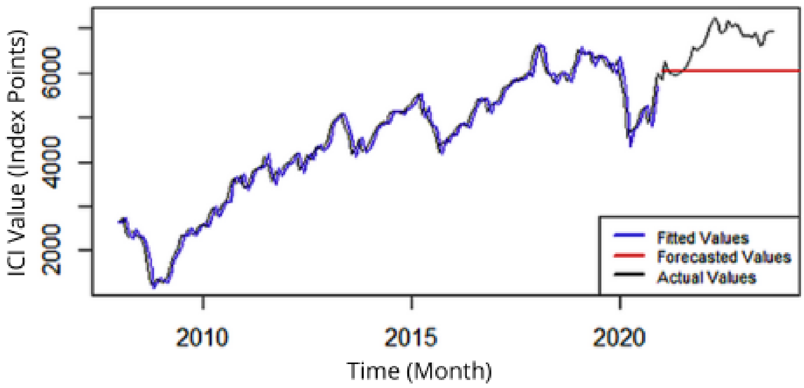

Figure 5. The Visualization of Fitted, Forecasted, and Actual ICI Data using ARIMA (1, 1, 0) from January 2008 to September 2023

The comparison of actual and forecasted values is visualized in Fig. 5. According to Fig. 5, the forecasted values of the ARIMA model cannot fit the actual data and tend to be constant. Therefore, it can be interpreted that the ARIMA model is not suitable for forecasting ICI data.

3.3 DRF Model for ICI Data

DRF is a supervised learning method that uses the independent and dependent variables in its process. The dependent variable used in this study is the data at time t, while the independent variable is chosen by the significance autocorrelation and partial autocorrelation of the ICI data. According to Figs. 3 and 4, the first lag of ACF and PACF is significant. Therefore, the data at the time $t-1$ becomes an independent variable.

Grid search CV is used to identify the optimal hyperparameter for the DRF model. Because the data used in this study are time series data, the cross-validation process is done by using the sliding windows method. The sliding windows method splits the data into K complementary partitions. The value of K used in this study is 5. The partitions of data are shown in Table 2.

Table 2. Data Partitions for Cross-Validation

K	Period of Training Data	Period of Testing Data
1	Jan-2008 to Dec-2018	Jan-2019 to Aug-2020
2	Jan-2008 to Jan-2019	Feb-2019 to Sep-2020
3	Jan-2008 to Feb-2019	Mar-2019 to Oct-2020
4	Jan-2008 to Mar-2019	April-2019 to Nov-2020
5	Jan-2008 to April-2019	May-2019 to Dec-2020

The DRF hyperparameters tuned in this study are ntree and node size. Ntree represents the regression trees built in the DRF model, while node size represents the minimum observations number in a terminal node [7]. There are three values of ntree (500, 600, and 700) and six values of node size (5, 6, 7, 8, 9, 10) tuned in this study. The optimal hyperparameters resulted from the Grid Search CV process are 600 (ntree) and 10 (node size). The validation process is carried out by calculating the forecast accuracy measure of the testing data. The results show that MAPE, RMSE, and MAD values are 5.40%, 435.56, and 371.50, respectively.

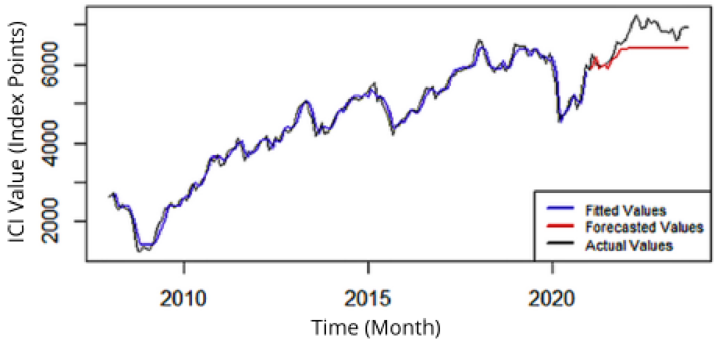

Figure 6. The Visualization of Fitted, Forecasted, and Actual ICI Data Using the DRF Model from January 2008 to September 2023

The visualization of actual and forecasted data resulting from the DRF model is shown in Fig. 6. Based on Fig. 6, the forecasting results of DRF in some early periods are fit with actual data. After that, the forecasting results tend to have a constant pattern. These results were also found in previous research, while DRF was used to model Export data. The forecasting results also remained constant when the data had high fluctuations [9]. This means that the DRF model does not have good performance in modeling ICI data.

3.4 Hybrid ARIMA-DRF Model for ICI Data

The best ARIMA model resulting from the Box-Jenkins procedure is ARIMA (1, 1, 0). The residuals from the ARIMA model are analyzed using the DRF model. The identification of optimal hyperparameters for modeling the residuals is also using Grid Search CV with three values of ntree (500, 600, and 700) and six values of node size (5, 6, 7, 8, 9, 10). The residuals are also split using the sliding windows method with $K = 5$. The partition rules of training and testing data for cross-validation process are also used in Table 2. The DRF optimal hyperparameters for the residuals ARIMA model are 700 (ntree) and 10 (node size). This DRF model is combined with ARIMA (1, 1, 0) to model ICI data.

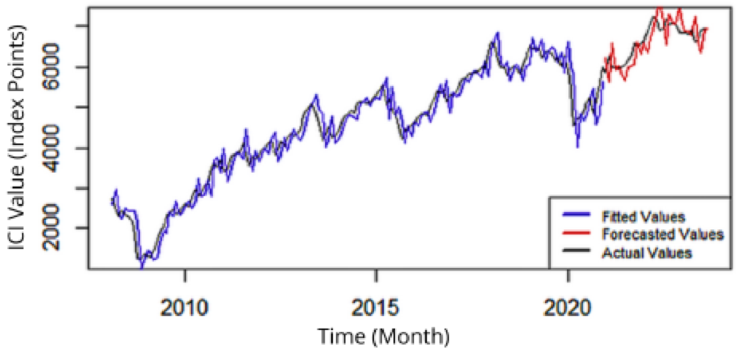

Figure 7. The Visualization of Fitted, Forecasted, and Actual ICI Data Using a Hybrid ARIMA-DRF Model from January 2008 to September 2023

The validation results from testing data show that the MAPE, RMSE, and MAD values for the hybrid ARIMA-DRF model are 4.17%, 355.24, and 276.40. According to these forecast accuracy metrics, the hybrid ARIMA-DRF outperforms ARIMA and DRF in modeling ICI data. The visualization results of forecasted and actual values of ICI data using the hybrid ARIMA-DRF model are shown in Fig.7. Based on Figs. 5, 6, and 7, the forecasting results of the hybrid ARIMA-DRF perform better than the independent ARIMA and DRF models. This is because the fitted and forecasted values of the hybrid model fit the actual value. This outcome aligns with a previous study that indicates the superior performance of the hybrid ARIMA-RF model in comparison to both the ARIMA and RF models individually [8].

3.5 Forecasting ICI Data

The forecasting of ICI Data was conducted by employing the most effective performance approach. Based on the performance evaluation results shown in Table 3, the hybrid ARIMA-DRF model achieved the best forecasting accuracy compared to the individual ARIMA and DRF models. This model consistently produced the lowest error values across all accuracy measures, including MAPE, RMSE, and MAD.

Table 3. Performance Evaluation Results

Model	Accuracy Measures	Results
ARIMA	MAPE	9.12%
	RMSE	722.54
	MAD	627.63
DRF	MAPE	5.40%
	RMSE	435.56
	MAD	371.50
ARIMA-DRF	MAPE	4.17%
	RMSE	355.24
	MAD	276.40

The forecasting outcome of the ICI data using a hybrid ARIMA-DRF model is displayed in Table 4, while the visualization of the prediction values is presented in Fig. 8. According to Table 4 and Fig. 8, the forecasting results of the ICI data tend to fluctuate. The value of ICI data exhibits a downward trend from October 2023 to December 2023, followed by an upward trajectory in January 2024. This can be caused by other economic factors such as interest rates, money supply, and exports [21].

Table 4. Forecasting Results of ICI Data Using Hybrid ARIMA-DRF

Forecasting Period	Forecasting Value
October 2023	7003.244
November 2023	6992.089

Forecasting Period	Forecasting Value
December 2023	6919.166
January 2024	7077.378
February 2024	7031.922
March 2024	7079.697
April 2024	7031.031
May 2024	7071.114
June 2024	7009.597
July 2024	7032.512
August 2024	7079.194
September 2024	7031.031

Furthermore, Fig. 8 illustrates that, throughout the subsequent months, the ICI continues to oscillate, maintaining values within the 7000–7100 range. This fluctuating pattern reflects the dynamic nature of the stock market and highlights the hybrid model's ability to closely capture the expected variations in the ICI data. This pattern reflects sustained market volatility without very sharp trends up or down in the short term. The predictions obtained are not only a reference for estimating market movements quantitatively but also serve as a guideline in developing adaptive investment strategies in accordance with the characteristics of market fluctuations. Thus, the hybrid ARIMA-DRF model proposed in this study has the potential to support investment decisions that are more scalable and responsive to the dynamics of the Indonesian capital market. Overall, the visualization supports the forecasting results in Table 4 and emphasizes the model's effectiveness in handling nonstationary and nonlinear financial time series data.

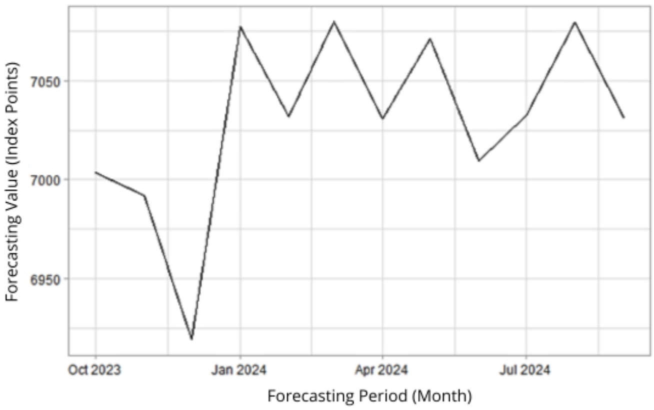

Figure 8. The Visualization of ICI Forecasting Values from October 2023 to September 2024

To provide a broader perspective, previous studies have explored the use of various hybrid models for financial time series forecasting. Pai and Lin (2005) proposed a hybrid ARIMA-SVM model that combines linear and nonlinear components, resulting in superior forecasting accuracy compared to the individual ARIMA and SVM models [11]. Similarly, Patel et al. (2015) compared hybrid models such as ARIMA-SVM, ARIMA-ANN, and ARIMA-RF, and concluded that hybrid approaches consistently outperform single models when applied to stock index prediction [9]. These studies highlight the effectiveness of hybrid models in capturing the complex dynamics present in financial markets.

The results obtained in this study are consistent with those findings. The hybrid ARIMA-DRF model demonstrated superior performance over the individual ARIMA and DRF models, as reflected by lower MAPE, RMSE, and MAD values. This supports the conclusion that combining linear and nonlinear modeling techniques enhances forecasting accuracy for nonstationary and nonlinear data, such as the Indonesia Composite Index. Therefore, the proposed hybrid ARIMA-DRF model offers a promising alternative for analyzing financial time series, in line with previous research outcomes [9], [11].

4. CONCLUSION

Indonesia Composite Index (ICI) data have a nonstationary pattern, and they also fluctuate over time. This study compares ARIMA, DRF, and hybrid ARIMA-DRF in modeling ICI data. The best ARIMA model for ICI data is ARIMA (1, 1, 0). The optimal hyperparameters for the DRF model are 600 (ntree) and 10 (node size). The hybrid ARIMA-DRF model is a combination of the best ARIMA model with the DRF model of the residuals. According to RMSE, MAE, and MAPE values from each model, the hybrid ARIMA-DRF outperforms the independent ARIMA and DRF models. Therefore, it can be concluded that the proposed hybrid ARIMA-DRF model can be a good alternative to analyze ICI data that have a nonstationary pattern and nonlinear characteristics. This model can also be an alternative to analyze other data with those characteristics. The forecasting results of ICI data using the hybrid ARIMA-DRF tend to increase and fluctuate over the forecasting periods (October 2023-September 2024). The results of this research offer valuable insights for investors, financial analysts, and policymakers. By showcasing the effectiveness of the hybrid ARIMA-DRF model in predicting ICI data, this study makes a methodological contribution to time series analysis. The proposed model has the potential to serve as a dependable tool for guiding investment decisions and analyzing market trends. Furthermore, the approach introduced in this study can be adapted to other financial time series with comparable characteristics, thereby improving forecasting precision in the financial sector.

However, this study has several limitations. Firstly, the focus is exclusively on ICI data, which may restrict the applicability of the findings to other financial indices or economic variables. Secondly, although the hybrid ARIMA-DRF model effectively captures both linear and nonlinear patterns, it does not consider external macroeconomic factors that could impact stock market fluctuations. Lastly, the model's validation was performed using historical data, and its predictive capability in real-time conditions has yet to be thoroughly evaluated. On the other hand, for future research, this hybrid ARIMA-DRF model can be integrated into a real-time Decision Support System (DSS) for investment firms, providing automatic alerts when significant shifts in the ICI trend are detected, thus enhancing proactive risk management.

Author Contributions

Andika Putri Ratnasari: leading the conceptualization, developing the methodology, formal analysis, implementing the software, performing the visualization, and preparing the original draft of the manuscript. Luthfia Hanun Yuli Arini: providing resources, assisting in the validation process, and participating in the review and editing of the manuscript. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgment

The authors would like to express their sincere gratitude to Universitas Negeri Yogyakarta (UNY) for the support and facilities provided during the implementation of this research.

Declarations

The authors declare no competing interests.

REFERENCES

[1] J. D. Cryer and K. Chan, *TIME SERIES ANALYSIS WITH APPLICATION IN R*. NEW YORK: SPRINGER SCIENCE+BUSINESS MEDIA, LLC, 2008. doi: https://doi.org/10.1007/978-0-387-75959-3

[2] D. Kobiela, D. Krefta, W. Król, and P. Weichbroth, “ARIMA VS LSTM ON NASDAQ STOCK EXCHANGE DATA,” *Procedia Comput. Sci.*, vol. 207, pp. 3830–3839, 2022. doi: https://doi.org/10.1016/j.procs.2022.09.445

[3] E. Dave, A. Leonardo, M. Jeanice, and N. Hanafiah, “FORECASTING INDONESIA EXPORTS USING A HYBRID MODEL ARIMA-LSTM,” *Procedia Comput. Sci.*, vol. 179, pp. 480–487, 2021. doi: https://doi.org/10.1016/j.procs.2021.01.031

[4] O. Fathi, “TIME SERIES FORECASTING USING A HYBRID ARIMA AND LSTM MODEL,” *Velv. Consult.*, pp. 1–7, 2019.

[5] F. Haselbeck, J. Killinger, K. Menrad, T. Hannus, and D. G. Grimm, “MACHINE LEARNING OUTPERFORMS CLASSICAL FORECASTING ON HORTICULTURAL SALES PREDICTIONS,” *Mach. Learn. with Appl.*, vol. 7, 2022. doi: 10.1016/j.mlwa.2021.100239. doi: https://doi.org/10.1016/j.mlwa.2021.100239

[6] S. Han, H. Kim, and Y. S. Lee, “DOUBLE RANDOM FOREST,” *Mach. Learn.*, vol. 109, no. 8, pp. 1569–1586, 2020. doi: https://doi.org/10.1007/s10994-020-05889-1

[7] A. P. Ratnasari, B. Susetyo, and K. A. Notodiputro, “COMPARISON OF DOUBLE RANDOM FOREST AND LONG SHORT-TERM MEMORY METHODS FOR ANALYZING ECONOMIC INDICATOR DATA,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, pp. 757–766, 2023. doi: https://doi.org/10.30598/barekengvol17iss2pp0757-0766

[8] N. Prawoto and B. A. Putra, “THE FACTORS AFFECTING THE COMPOSITE STOCK PRICE INDEX (CSPI) OF INDONESIA STOCK EXCHANGE,” *J. Ekon. dan Stud. Pembang.*, vol. 21, 2020. doi: 10.18196/jesp.21.1.5032. doi: https://doi.org/10.18196/jesp.21.1.5032

[9] M. Kumar and M. Thenmozhi, “FORECASTING STOCK INDEX RETURNS USING ARIMA-SVM, ARIMA-ANN, AND ARIMA-RANDOM FOREST HYBRID MODELS,” *Int. J. Banking, Account. Financ.*, vol. 5, no. 3, pp. 284–308, 2014. doi: https://doi.org/10.1504/IJBAAF.2014.064307

[10] N. Merh, V. Prakash Saxena, and K. Raj Pardasani, “A COMPARISON BETWEEN HYBRID APPROACHES OF ANN AND ARIMA FOR INDIAN STOCK TREND FORECASTING,” *Bus. Intell. J.*, pp. 23–43, 2010.

[11] F. P. Pai and C. S. Lin, “A HYBRID ARIMA AND SUPPORT VECTOR MACHINES MODEL IN STOCK PRICE FORECASTING,” *Int. J. Manag. Sci.*, vol. 33, pp. 497–505, 2005. doi: https://doi.org/10.1016/j.omega.2004.07.024

[12] S. Ray, A. Lama, P. Mishra, T. Biswas, S. S. Das, and B. Gurung, “AN ARIMA-LSTM MODEL FOR PREDICTING VOLATILE AGRICULTURAL PRICE SERIES WITH RANDOM FOREST TECHNIQUE,” *Appl. Soft Comput.*, vol. 149, 2023. doi: https://doi.org/10.1016/j.asoc.2023.110939

[13] J. Zhang, H. Liu, W. Bai, and X. Li, “A HYBRID APPROACH OF WAVELET TRANSFORM, ARIMA, AND LSTM MODEL FOR THE SHARE PRICE INDEX FUTURES FORECASTING,” *North Am. J. Econ. Financ.*, vol. 69, 2023. doi: https://doi.org/10.1016/j.najef.2023.102022

[14] C. H. Chien, A. J. C. Trappey, and C. C. Wang, “ARIMA-ADABOOST HYBRID APPROACH FOR PRODUCT QUALITY PREDICTION IN ADVANCED TRANSFORMER MANUFACTURING,” *Adv. Eng. Informatics*, vol. 57, 2023. doi: https://doi.org/10.1016/j.aei.2023.102055

[15] D. C. Montgomery, C. L. Jennings, and M. Kulahci, “INTRODUCTION TIME SERIES ANALYSIS AND FORECASTING,” P. 671, 2015.

[16] T. T. H. PHAN AND X. H. NGUYEN, “COMBINING STATISTICAL MACHINE LEARNING MODELS WITH ARIMA FOR WATER LEVEL FORECASTING,” *Adv. Water Resour.*, vol. 142, 2020. doi: https://doi.org/10.1016/j.advwatres.2020.103656

[17] W. Zha *et al.*, “FORECASTING MONTHLY GAS FIELD PRODUCTION BASED ON THE CNN-LSTM MODEL,” *Energy*, vol. 260, 2022. doi: https://doi.org/10.1016/j.energy.2022.124889

[18] G. I. Drewil and R. J. Al-Bahadili, “AIR POLLUTION PREDICTION USING LSTM DEEP LEARNING AND METAHEURISTICS ALGORITHMS,” *Meas. Sensors*, vol. 24, 2022. doi: https://doi.org/10.1016/j.measen.2022.100546

[19] I. Nasirtafreshi, “FORECASTING CRYPTOCURRENCY PRICES USING RECURRENT NEURAL NETWORK AND LONG SHORT-TERM MEMORY,” *Data Knowl. Eng.*, vol. 139, no. October 2021, p. 102009, 2022. doi: https://doi.org/10.1016/j.datak.2022.102009

[20] H. Prabowo, Suhartono, and D. D. Prastyo, “THE PERFORMANCE OF RAMSEY TEST, WHITE TEST, AND TERASVIRTA TEST IN DETECTING NONLINEARITY,” *INFERENSI*, vol. 3, no. 1, 2020. doi: https://doi.org/10.12962/j27213862.v3i1.6876

[21] A. J. Wibowo and R. Khoirudin, “DOES MACROECONOMIC FLUCTUATION MATTER FOR THE COMPOSITE STOCK PRICE INDEX?,” *J. Ekon. Pembang.*, vol. 20, no. 1, pp. 105–114, 2022. doi: https://doi.org/10.29259/jep.v20i1.17479