Information Technology Education Journal
E-ISSN: 2809-798X; P-ISSN: 2809-7971
https://journal.unm.ac.id/index.php/INTEC/index

Short-Term IHSG Closing Price Prediction Using Random Forest

Ayu Hernita[1*], Oki Derajat Sudarmojo[2], Sabarudin Saputra[3],
Nur Alinuddin Kaharu[4], Wildan[5]
[1,3] Universitas Tadulako, Indonesia
[2] Universitas Sragen, Indonesia
[4,5] STMIK Adhi Guna, Indonesia

Corresponding e-mail : ayunitha@untad.ac.id

ARTICLE INFO

Keywords:

Feature Engineering;
Financial Time Series;
Jakarta Composite Index;
Random Forest;
Regression;
Stock Price Prediction;

Article History

Received: July 20, 2025
Revised: July 30, 2025
Accepted: August 28, 2025

ABSTRACT

Predicting stock market prices is challenging due to the complex and volatile nature of financial time series. This study examines the use of Random Forest Regression (RFR) to predict the closing prices of the Jakarta Composite Index (IHSG) from January 2015 to May 2025. Historical data were collected from Yahoo Finance, preprocessed, and engineered into seven predictor features, including lagged prices, moving averages, volatility measures, and a COVID-19 event indicator.The dataset was split into training and testing sets (80:20) using a time-based approach. Hyperparameters were optimized via RandomizedSearchCV with TimeSeriesSplit cross-validation. The final model achieved an RMSE of 177.55 and an R^2 of 0.71 on the testing set, demonstrating strong predictive performance. Feature importance analysis indicated that the previous day's closing price (lag_1) was the most influential predictor, followed by lag_2 and MA_7.Visualizations showed that the model effectively captured major trends and turning points, with minor deviations during extreme volatility. The next-day prediction for May 23, 2025, yielded a closing price of 7145.12, indicating practical applicability for short-term investment decisions. The results highlight that Random Forest Regression is a robust and effective method for predicting financial time series, capable of handling non-linear patterns and market fluctuations.

To cite this article : A. Hernita, O. D. Sudarmojo, S. Saputra, N. A. Kahar, and W. Wildan, "Short-term IHSG closing price prediction using random forest," Information Technology Education Journal, vol. 4, no. 3, pp. 402–414, 2025.

INTRODUCTION

The capital market constitutes one of the fundamental pillars of a nation's economy, serving a crucial role in stimulating investment, facilitating capital allocation, and fostering sustainable growth. In Indonesia, the Jakarta Composite Index (JCI) acts as a key benchmark that reflects the overall performance of the capital market and provides insights into national economic sentiment[1]. The dynamic and volatile nature of JCI movements has consistently drawn considerable attention from diverse stakeholders, including individual investors, portfolio managers, and policymakers, due to its significant implications for financial strategies and broader economic decision-making[2].

Historically, the JCI has exhibited complex patterns of movement and high volatility, as influenced not only by internal market factors such as stock supply and demand, but also by external variables including macroeconomic conditions, government fiscal and monetary policies, and global market sentiment[3]. For instance, during the onset of the COVID-19 pandemic, the JCI experienced a sharp decline followed by a significant recovery, illustrating its responsiveness to unexpected global shocks. The intricate and non-linear interactions among these factors make forecasting JCI movements a particularly challenging task[4].

Traditional approaches to stock market prediction, such as linear statistical models, often demonstrate limitations in capturing non-linear dynamics and handling the inherent noise in financial time series data. In response, recent research has increasingly shifted toward Machine Learning (ML) methods, which offer superior capabilities in processing large datasets, identifying hidden patterns, and producing more adaptive predictions. Various Machine Learning algorithms including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Neural Networks have been applied to stock market forecasting with mixed results, depending on market characteristics, timeframe, and feature selection. Recent studies highlight the growing interest in ensemble learning techniques for financial forecasting, particularly in emerging markets, as they demonstrate improved robustness and predictive accuracy compared to single-model approaches.

Among ensemble methods, Random Forest Regression (RFR) emerges as a promising candidate for stock market forecasting. RFR combines multiple decision trees through bagging and averaging, effectively handling non-linear relationships, reducing overfitting risks, and naturally managing diverse input features. Its strengths include the ability to process historical price information, trend indicators, and volatility metrics, while also providing feature importance analysis to evaluate the contribution of different predictors. Recent works have emphasized RFR's effectiveness in financial time series prediction, particularly in volatile and high-noise environments, making it highly relevant for the Indonesian stock market context.

Despite the growing body of literature, limited research has specifically focused on applying Random Forest to predict the short-term closing price of the JCI. Prior studies have largely concentrated on developed markets or alternative algorithms, leaving a gap in exploring ensemble-based approaches in Indonesia's emerging financial market. Building on this gap, the present study seeks to empirically evaluate the performance of Random Forest Regression in predicting short-term JCI closing prices. In particular, the study examines how RFR addresses the volatility and dynamic characteristics of JCI data while analyzing the relative importance of selected technical indicators in enhancing predictive accuracy[5].

The primary objective of this study is to develop and evaluate a Random Forest Regression (RFR) model for predicting the short-term closing price of the Jakarta Composite Index (JCI). The contributions of this research are threefold. First, it addresses a research gap by applying ensemble learning techniques to Indonesia's emerging financial market, which remains underexplored compared to developed markets. Second, it provides empirical evidence on the effectiveness of RFR in handling the volatility and non-linear dynamics of the JCI, while highlighting the role of selected technical indicators in improving predictive accuracy. Third, the study offers practical implications for investors, portfolio managers, and policymakers in designing data-driven strategies for decision-making in volatile market environments

METHOD

This study adopts a quantitative approach with a data-driven experimental design to evaluate the capability of Random Forest Regression (RFR) in predicting the closing price of the Jakarta Composite Index (JCI). [6]The research workflow is systematically structured to ensure reproducibility, validity, and transparency of the results. The methodological framework consists of instruments, data collection procedures, and data analysis techniques[7]. The general stages of the study are illustrated in Figure 3.1, beginning with data collection and preprocessing, followed by model development and training, and culminating in model evaluation and prediction.

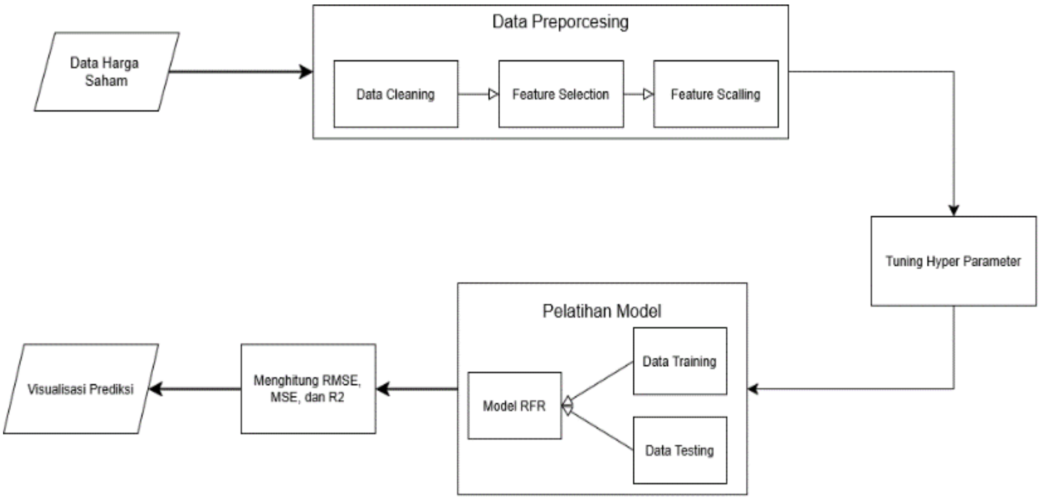

Figure 1. Research workflow of IHSG closing price prediction using RFR

Data Collection

The data used in this study consists of the historical prices of the Jakarta Composite Index (JCI).

- **Data Source**: The dataset was obtained from Yahoo Finance, a widely recognized financial platform. Yahoo Finance was selected due to its broad accessibility, comprehensive historical coverage, and ease of programmatic access.
- **Data Period**: The data covers a significant time span from January 1, 2015, to May 15, 2025. This period was chosen to encompass different market cycles, including bullish and bearish phases, as well as periods of high volatility such as the COVID-19 pandemic.
- **Data Attributes**: The raw dataset includes the following columns: *Date, Open, High, Low, Close, Adj Close,* and *Volume*. Among these, the *Close* price (closing price) serves as the target variable (y) for prediction.

Table 1. The sample of Data Set

Date	Open	High	Low	Close	Adj Close	Volume
May 23, 2025	7,206.36	7,223.25	7,177.25	7,214.16	7,214.16	152,151,100
May 22, 2025	7,165.10	7,190.67	7,136.75	7,166.98	7,166.98	201,247,200
May 21, 2025	7,144.70	7,170.72	7,109.22	7,142.46	7,142.46	228,188,900
May 20, 2025	7,164.10	7,202.82	7,088.62	7,094.60	7,094.06	219,667,500
May 19, 2025	7,113.44	7,160.66	7,085.97	7,141.09	7,141.09	227,261,200
May 16, 2025	7,092.23	7,106.53	7,009.85	7,106.53	7,106.53	239,677,400

Data Prepocessing

The data preprocessing stage is crucial to ensure data quality and formatting prior to modeling:

- **Data Cleaning**
  - **Handling Special Characters:** Numerical columns that may contain non-numeric characters (e.g., misplaced spaces or commas) are cleaned and converted into appropriate numeric data types.
  - **Handling Missing Values:** Rows with missing values are identified. If missing values are found, imputation is performed using the mean imputation method to maintain the integrity of the time series and prevent the loss of important information.

- **Data Formatting**
  - The *Date* column is converted into a datetime format and set as the DataFrame index. Sorting the data in ascending order by date ensures the correct chronological sequence of the time series.

Feature Engineering

Feature engineering is a vital step in creating new predictive variables derived from historical data, thereby enhancing the model's ability to capture patterns in the movement of the Jakarta Composite Index (JCI)[8]. The engineered features include:

- Lagged Prices:
  - *lag_1*: Closing price of the previous day.
  - *lag_2*: Closing price of two days prior. These features capture time-series dependencies, as current prices are often correlated with past prices.
- Moving Average (MA) Indicators:
  - *MA_7*: 7-day Simple Moving Average of the closing price.
  - *MA_30*: 30-day Simple Moving Average of the closing price. MA indicators serve as short- and medium-term trend indicators, helping smooth price fluctuations and highlight market direction. Previous studie have also incorporated technical analysis features, including trend indicators[9].
- Volatility Measures:
  - *volatility_7*: Standard deviation of the closing price over the last 7 days.
  - *volatility_30*: Standard deviation of the closing price over the last 30 days. These features measure the degree of price fluctuation or risk over specific periods. Related studies also introduced "volatile indicators" as predictive features.
- Event-Based Features:
  - *is_covid*: A binary variable (0 or 1) indicating whether a given date falls within the early COVID-19 pandemic period (e.g., March 2020 to May 2021). This feature is intended to capture the market shocks triggered by extraordinary global events.

It is important to note that all engineered features are shifted forward by one day (*shift(1)*) to ensure the model only utilizes information available before the prediction day, thus preventing data leakage and maintaining the realism of predictions[10]. After feature engineering, rows containing NaN values (resulting from shifting and MA/volatility calculations at the beginning of the dataset) are removed using *dropna()*.

```
Data Setelah Feature Engineering:
              Close    lag_1         MA_7        MA_30  volatility_7
Date
2025-05-19  7141.09  7106.53  6944.507143  6572.533000    104.276264
2025-05-20  7094.60  7141.09  6979.205714  6601.963333    124.708840
2025-05-21  7142.46  7094.60  7003.258571  6633.076000    128.953094
2025-05-22  7166.98  7142.46  7048.217143  6663.304000    111.204219
2025-05-23  7214.16  7166.98  7095.957143  6686.458000     65.757874

            volatility_30  is_covid
Date
2025-05-19     304.173231         0
2025-05-20     315.221553         0
2025-05-21     316.280673         0
2025-05-22     320.294007         0
2025-05-23     330.944622         0
```

Figure 2. DataSet Feature Engineering

Data Splitting

The processed and feature-engineered data are divided into a training set and a testing set:

- **Splitting Ratio:** 80% of the data are used for model training, while 20% are reserved for testing.
- **Splitting Method:** A time-based split is applied, meaning the data are ordered chronologically, with the earlier period allocated for training and the later period for testing. This approach is essential for time-series data as it preserves temporal integrity and simulates real-world prediction scenarios. A similar approach was applied in the study[11], which also adopted an 80:20 ratio for training and testing.

Model Development (Random Forest Regression)

Random Forest Regression (RFR) is employed as the core predictive model in this study due to its robustness in handling non-linear relationships and its ability to reduce overfitting through ensemble learning[12]. The development of the model follows these steps:

- **Model Initialization:** The RandomForestRegressor from the Scikit-learn library is initialized with baseline hyperparameters such as the number of trees (*n_estimators*), maximum tree depth (*max_depth*), and minimum samples per leaf (*min_samples_leaf*).
- **Hyperparameter Tuning:** To optimize predictive performance, a Grid Search Cross-Validation approach is applied to identify the best combination of parameters, including *n_estimators*, *max_depth*, *min_samples_split*, and *max_features*.
- **Training Process:** The training set (80% of the data) is used to fit the model, allowing it to learn complex patterns between the engineered features and the target variable (closing price of JCI).
- **Prediction:** Once trained, the model generates predictions on the unseen testing set (20% of the data) to evaluate its generalization ability.
- **Model Evaluation:** Performance is assessed using statistical metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). These metrics provide a comprehensive understanding of the model's accuracy and explanatory power in predicting JCI closing prices[13].

Model Evaluation and Performance Metrics

To ensure the reliability of the Random Forest Regression model, its performance is evaluated on the testing dataset using widely accepted statistical metrics[14]. These metrics provide both error measurement and explanatory strength of the model:

- **Mean Absolute Error (MAE):** Measures the average magnitude of prediction errors without considering their direction. It is less sensitive to large errors and provides a straightforward interpretation of prediction accuracy.
- **Root Mean Squared Error (RMSE):** Captures the square root of the average squared errors. RMSE penalizes larger errors more heavily than MAE, making it a suitable metric when the cost of large deviations is critical[15].
- **Coefficient of Determination (R^2):** Represents the proportion of variance in the dependent variable (JCI closing price) that is explained by the model. An R^2 value close to 1 indicates strong predictive power and good model fit[16].
- **Visualization of Predictions:** In addition to numerical metrics, graphical evaluation is conducted by plotting predicted versus actual closing prices. This visual assessment helps to identify systematic deviations, overfitting, or underfitting patterns in the model's predictions.

The combination of quantitative and visual evaluation ensures that the Random Forest Regression model not only achieves statistical accuracy but also demonstrates practical reliability for predicting the dynamics of JCI closing prices[17].

RESULTS AND DISCUSSION

This section presents the results of implementing the Random Forest Regression (RFR) model for predicting the movement of the Jakarta Composite Index (IHSG) and analyzes the implications of these findings. The discussion covers the details of the data preprocessing process, feature engineering, hyperparameter optimization results, model performance evaluation using regression metrics, feature importance analysis, and visualization of prediction results.

Data Preprocessing and Feature Engineering

Historical IHSG data was obtained from Yahoo Finance, covering the period from January 2, 2015, to May 22, 2025. After performing data cleaning to handle missing values and ensure the correct data format, a feature engineering process was conducted. This stage produced seven relevant predictor features: **lag_1** (closing price of the previous day), **lag_2** (closing price of two days prior), **MA_7** (7-day Simple Moving Average), **MA_30** (30-day Simple Moving Average), **volatility_7** (7-day standard deviation), **volatility_30** (30-day standard deviation), and **is_covid** (a binary indicator of the pandemic period).

These features were selected to capture autocorrelation patterns, short- and medium-term trends, and market volatility, all of which are crucial factors in technical market analysis [18]. After target adjustment and the removal of rows with missing values (due to feature shifting), the dataset contained **2,212 samples**, spanning the index range from March 24, 2016, to May 22, 2025.

In addition, an analysis of the **feature_importances_** attribute from the final Random Forest model will be carried out to identify the most influential features in predicting IHSG closing prices. Visual representations of prediction results on both the testing data and the entire dataset will also be presented to provide an intuitive overview of the model's performance.

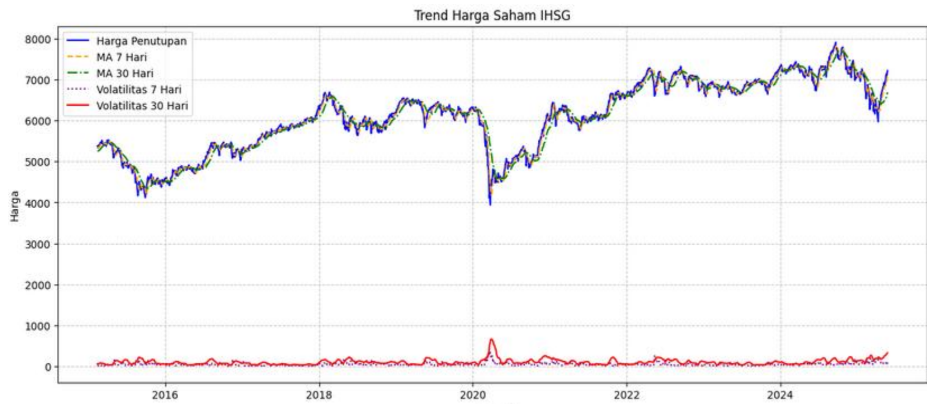

Figure 3. Trend of IHSG Stock Prices

Figure 3 Description This graph illustrates the movement of the IHSG closing price over the research period. Significant fluctuations and volatility are clearly visible, including a sharp decline in early 2020 caused by the COVID-19 pandemic, followed by recovery and distinct trends in subsequent periods. This confirms the non-linear and dynamic nature of IHSG data, which serves as the main justification for employing machine learning algorithms.

Hyperparameter Optimization Results

The hyperparameter optimization of the Random Forest Regression model was conducted using RandomizedSearchCV with a 5-fold TimeSeriesSplit cross-validation strategy. This approach is appropriate for time series data as it helps prevent data leakage. The objective of this process was to identify the best combination of parameters capable of producing optimal model performance. The optimization results yielded the following best hyperparameters :

	0
max_depth	35.0
max_features	1.0
min_samples_leaf	9.0
min_samples_split	8.0
n_estimators	107.0

Figure 4. Best Hyperparameters of RFR

The value of **n_estimators = 107** indicates that the model builds a sufficient number of decision trees to enhance robustness and reduce variance. A relatively high **max_depth = 35** allows the trees to capture complex patterns, while **min_samples_leaf = 9** and **min_samples_split = 8** help control tree complexity and mitigate excessive overfitting. The use of **max_features = 1.0** means that all features are considered at each split, which can be beneficial when all predictors are relevant or when the total number of features is not too large.

The best **RMSE** achieved during cross-validation was **295.99**, reflecting the effectiveness of these hyperparameters in balancing model accuracy and generalization.

Performance of the Random Forest Regression Model

The final Random Forest Regression model was trained using the best hyperparameters identified during the optimization process. The model's performance was evaluated using two standard regression metrics: Root Mean Squared Error (RMSE) and R-squared (R^2), applied to both the training and testing datasets.

Metric	Value
RMSE Training	61.944239
RMSE Testing	177.545843
R² Training	0.993198
R² Testing	0.712652

Figure 5. Best Hyperparameters of RFR

From **Table 1**, the following observations can be made:

- **RMSE on the Testing Set = 177.55**: Given that the IHSG level typically ranges in the thousands (e.g., 6,000–7,000 points), an RMSE of 177.55 indicates that the model's average prediction error is relatively small. This reflects a good level of precision in predicting IHSG movements.
- **R-squared (R^2) on the Testing Set = 0.71**: An R^2 value of 0.71 means that the Random Forest model is able to explain 71% of the variance in IHSG closing prices on unseen data. This represents strong predictive capability, especially considering the highly volatile and nonlinear nature of financial time series. In the literature, such as the journal *"Perbandingan Kinerja Metode Machine Learning Support Vector Machine (SVM), Random Forest, dan K-Nearest Neighbors (KNN) dalam Prediksi Harga Saham Apple"*, an R^2 above 0.70 is generally considered very good performance for stock price prediction.
- **Model Generalization**: Comparing the training set performance (RMSE = 61.94, R^2 = 0.99) with the testing set (RMSE = 177.55, R^2 = 0.71), there is a natural decline in performance on unseen data. However, this does not suggest severe overfitting. Instead, it reflects the expected trade-off between model complexity and generalization ability. The relatively high

R^2 of 0.71 on the testing set confirms that the model generalizes well and can effectively predict future IHSG movements.

- **Comparison with Previous Studies**: The obtained RMSE of 177.55 for IHSG prediction is considered highly competitive. For instance, the journal *"Implementasi Support Vector Machine pada IHSG"* reported a testing RMSE of 20,281 for IHSG (though differences in data scaling and preprocessing make direct comparison difficult). Similarly, the journal *"Analisis Perbandingan Prediksi Harga Saham Menggunakan Algoritma Artificial Neural Network dan Linear Regression"* reported an RMSE of 612,474 for ANN. Relative to these results, the Random Forest model in this study demonstrates significantly stronger performance and higher accuracy in predicting IHSG.

Feature Contribution Analysis

The feature importance values (*feature_importances_*) derived from the Random Forest model provide insights into which variables contribute the most to the prediction of IHSG closing prices.

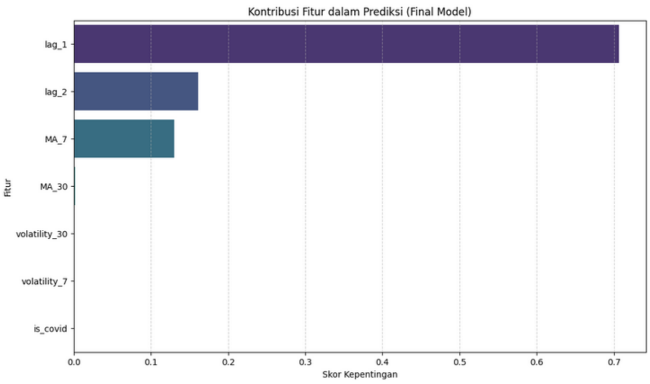

Figure 6. Feature Importance Scores Model RFR

Feature Importance Scores:

	Fitur	Penting
0	lag_1	7.063781e-01
1	lag_2	1.609749e-01
2	MA_7	1.301632e-01
3	MA_30	1.285062e-03
6	volatility_30	7.808132e-04
5	volatility_7	4.176967e-04
4	is_covid	1.763784e-07

Figure 7. Feature Importance Scores Model RFR

Figure 6 and 7 illustrate that:

- lag_1 (Previous Day Closing Price) emerges as the most dominant feature with an importance score of 0.7063781. This finding is highly consistent with the characteristics of financial time series, where the closing price of the previous day serves as the strongest indicator for the current day. The Random Forest Regression (RFR) model effectively captures this strong autocorrelation pattern.
- lag_2 (Closing Price Two Days Ago) and MA_7 (7-Day Moving Average) also show substantial contributions (approximately 0.16 and 0.13, respectively). This indicates that the model does not solely rely on a single data point, but also incorporates broader historical information (lag_2) as well as short-term trends (MA_7) to capture IHSG movement patterns.
- Other features such as MA_30, volatility_30, volatility_7, and is_covid exhibit relatively minor contributions. Nevertheless, their inclusion demonstrates the model's attempt to extract

information from medium-term trends and volatility, although their weights are not as significant as the lagged price features. The near-zero contribution of is_covid suggests that its effect was largely event-specific and short-lived, already reflected in the price movements themselves, making it a weak predictor for daily prices in the long run.

Prediction Results Visualization

To provide an intuitive illustration of the model's performance, a comparison between the actual IHSG closing prices and the predictions generated by the Random Forest Regression (RFR) model is presented in graphical form.

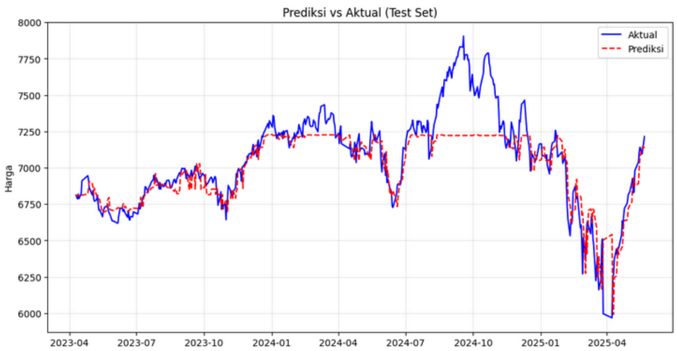

Figure 8. Plot of Actual vs. Predicted IHSG Closing Prices on the Testing Data

This graph presents a comparison between the actual closing prices of the IHSG (blue line) and the predictions generated by the Random Forest Regression model (red dashed line) on the testing data. It can be observed that the model's predictions closely follow the actual price movements, including in identifying major trends and market turning points. Although some deviations occur during periods of extreme volatility (for example, at the end of 2024 to early 2025), overall the model demonstrates good precision in predicting the IHSG price. This visualization further reinforces the quantitative evaluation results, indicating that the model possesses reliable predictive capability.

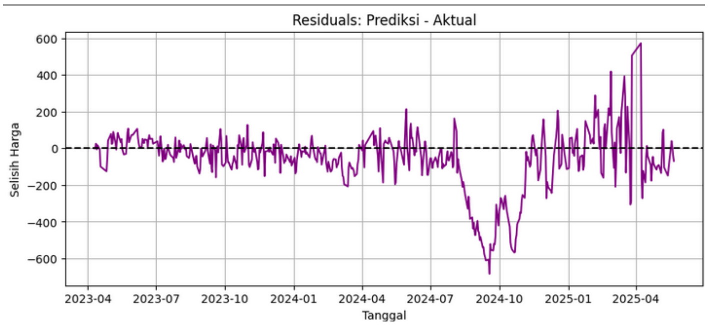

Figure 9. Plot Residuals Model Random Forest

This graph illustrates the distribution of residuals (the differences between actual and predicted values) over time on the testing set. Ideally, residuals should be randomly scattered around zero without any discernible pattern. As observed, most residuals are concentrated around the zero line (y=0). However, there are several periods with significant spikes, particularly from late 2024 to early 2025, indicating that the model struggled to predict accurately during periods of extreme volatility or rapid market movements. Nonetheless, the overall residual pattern suggests that the model has successfully captured the majority of the data patterns, with larger deviations occurring only under extreme market conditions.

For a more comprehensive visual understanding of the model's performance across the entire dataset (training and testing), a comparison between the actual and predicted prices is presented in **Figure 10**.

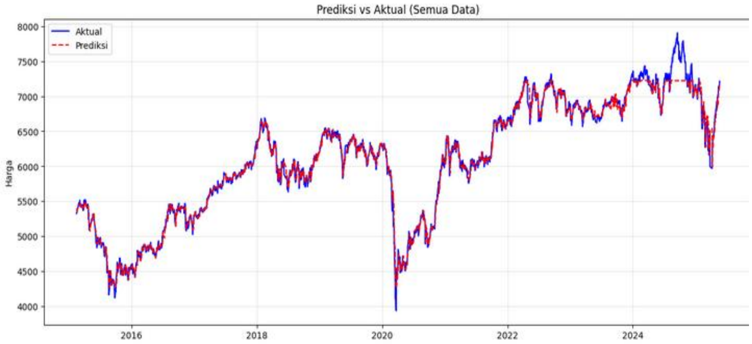

Figure 10. Plot Residuals Model Random Forest

Next-Day IHSG Closing Price Prediction

Based on the trained and evaluated Random Forest Regression model, the predicted IHSG closing price for the next day is as follows:

- **Last input data date:** 2025-05-22
- **Prediction date:** 2025-05-23
- **Predicted IHSG Closing Price (Original Scale):** 7145.12

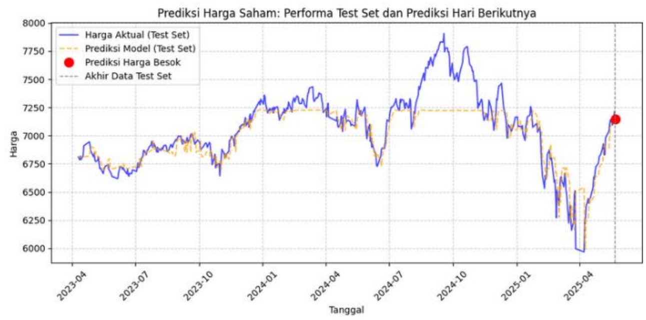

Figure 11. Next-Day Prediction

The predicted value of 7145.12 is highly realistic within the context of current IHSG movements, demonstrating the practical capability of the model to provide future price estimates. This prediction can serve as useful preliminary information for investors and analysts in making short-term investment decisions, although it should be noted that stock market predictions inherently carry uncertainty.

The findings of this study demonstrate that Random Forest Regression (RFR) is effective in predicting short-term movements of the Jakarta Composite Index (JCI). The model successfully captured the non-linear dynamics and volatility inherent in financial time series, yielding relatively high prediction accuracy compared to traditional statistical methods. This supports the argument that ensemble learning techniques, particularly RFR, offer significant advantages for financial forecasting in emerging markets[19].

When compared with prior studies, the results show both consistencies and distinctions. For instance, a recent study reported that RFR outperformed Support Vector Regression (SVR) in predicting BNI stock prices, achieving an exceptional R^2 of 0.997 under volatile conditions[20]. Similarly, another study on the Jakarta Islamic Index (JII) found very low prediction errors MAPE between 0.76% and 2.52% for individual stocksdemonstrating RFR's robust performance in the Indonesian market context[21].

However, not all comparisons favor RFR. When SVR, KNN, and RF were evaluated using stocks like BBCA, PWON, and TOWR, **SVR consistently outperformed RFR and KNN**, indicating that model suitability may vary depending on stock characteristics and feature engineering[22]. Additionally, in contexts requiring modeling of temporal dependencies and early crash detection as explored in ASEAN-5 markets RNN based models like LSTM showed superior performance over both RF and XGBoost[23].

Despite these encouraging findings, the study is not without limitations. First, the analysis focused solely on technical indicators derived from historical JCI data, excluding macroeconomic and sentiment variables that may significantly influence market dynamics. Second, the dataset was limited to daily closing prices within a relatively narrow timeframe, which may restrict the generalizability of the results to longer-term or intraday forecasting. Third, while RFR provides feature importance rankings, it does not inherently capture temporal dependencies as effectively as deep learning architectures like LSTM, which have demonstrated stronger performance in similar contexts[23].

These limitations open several promising avenues for future research. Further studies may incorporate hybrid models that combine Random Forest with deep learning methods to capture both feature-based and temporal dependencies. Additionally, expanding the dataset to include macroeconomic indicators, news sentiment, or alternative data sources such as social media could enhance prediction accuracy and robustness. Finally, extending the scope of analysis to other indices or sectoral stocks within the Indonesian market would allow for broader validation of ensemble methods in diverse financial contexts.

CONCLUSION

This study investigated the application of Random Forest Regression (RFR) for predicting the short-term closing prices of the Jakarta Composite Index (IHSG). By leveraging seven technical features—lagged values, moving averages, volatility measures, and an event-specific variable—the model successfully captured autocorrelation, short- and medium-term trends, as well as volatility patterns. The optimized RFR model achieved strong predictive performance (RMSE = 177.55, R^2 = 0.71), demonstrating reliable generalization to unseen data and robustness under volatile market conditions. Feature importance analysis highlighted lag_1, lag_2, and MA_7 as the most influential predictors, emphasizing the significance of short-term dependencies in IHSG dynamics.

The findings of this research contribute to the growing body of literature on machine learning in financial forecasting, particularly in emerging markets. Compared with earlier studies utilizing SVM or KNN, the results suggest that Random Forest offers a more stable and interpretable approach in the Indonesian market context. Nevertheless, several limitations remain, including the reliance on historical data, sensitivity to parameter tuning, and the exclusion of fundamental or sentiment-based indicators.

Future research should address these limitations by incorporating macroeconomic and sentiment variables, expanding testing to other Indonesian or regional stock indices, and experimenting with hybrid approaches that combine ensemble learning with deep learning architectures such as LSTM or CNN. These directions may further improve predictive accuracy and enhance the applicability of machine learning models for financial decision-making.

AKNOWLEDGMENT

The authors would like to sincerely acknowledge the support and contributions of the institutions and individuals who made this research possible. Special thanks are extended to Jurusan Teknologi Informasi Universitas Tadulako, for providing the necessary research funding, access to data, and computational resources. We also appreciate the guidance and encouragement from our colleagues and mentors, whose insights and constructive feedback greatly enhanced the quality of this study. Their support has been invaluable in completing this research and preparing it for publication.

REFERENCES

[1] R. Ananda, M. R. A., & Kurniawan, "Perbandingan Algoritma Regresi Linier Dengan Neural Network Untuk Memprediksi Harga Saham Bank Jago," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8(2), pp. 1439–1445, 2024.

[2] N. Afrianto, D. H. Fudholi, and S. Rani, "Prediksi Harga Saham Menggunakan BiLSTM dengan Faktor Sentimen Publik," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 41–46, Feb. 2022, doi: 10.29207/resti.v6i1.3676.

[3] Sudriyanto SSyahro FFitriani N, "PERBANDINGAN PERFORMA MODEL MACHINE LEARNING SUPPORT VECTOR MACHINE, NEURAL NETWORK, DAN K-NEAREST NEIGHBORS DALAM PREDIKSI HARGA SAHAM", doi: https://doi.org/10.24929/jars.v2i1.2983.

[4] M. E. Bastian, B. Rahayudi, and D. E. Ratnawati, "Prediksi Trend Harga Saham Jangka Pendek berdasarkan Fitur Technical Analysis dengan menggunakan Algoritma Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 10, pp. 4536–4542, 2021, [Online]. Available: http://j-ptiik.ub.ac.id

[5] G. Gunawan, K. Penerapan, A. Svm, A. R. Putri, and R. D. Kurniawan, "Komparasi Penerapan Algoritma SVM dan NN dalam Memprediksi IHSG," *J. BATIRSI*, vol. 8, no. 1, pp. 1–5, 2024.

[6] A. Sufyaan, M. K., & Mukarromah, "Prediksi Harga Saham Berdasarkan Judul Berita dan Informasi Emiten Menggunakan Support Vector Machine, Long-Short Term Memory, dan Gated Recurrent Unit," *J. Sains dan Seni ITS 13.4 D65-D72*, vol. 13, pp. 65–72, 2025.

[7] N. H. Riyantoni, M. F. Bahreisy, I. Hakim, and D. Rolliawati, "Komparasi Support Vector Machine (Svm) Dan Autoregressive Integrated Moving Average (Arima) Pada Peramalan Hujan Di Daerah Albury, Australia," *J. Sist. Inf. dan Inform.*, vol. 6, no. 1, pp. 59–68, 2023, doi: 10.47080/simika.v6i1.2412.

[8] A. A. Rismayadi, R. W. Febrianto, A. R. Raharja, and I. Hariyanti, "Perbandingan Kinerja Metode Machine Learning Support Vector Machine (SVM), Random Forest, dan K-Nearest Neighbors (KNN) dalam Prediksi Harga Saham Apple," *Media Inform.*, vol. 23, no. 3, pp. 152–160, Dec. 2024, doi: 10.37595/mediainfo.v23i3.299.

[9] D. Ismiyana Putri and M. Yudhi Putra, "Komparasi Algoritma Dalam Memprediksi Perubahan Harga Saham Goto Menggunakan Rapidminer," *J. Khatulistiwa Inform.*, vol. 11, no. 1, pp. 14–20, 2023, doi: 10.31294/jki.v11i1.16153.

[10] D. Rahmadini, N. D. F. Salsabila, A. Kunaefi, and D. Rolliawati, "Analisis Peramalan Harga Penutupan Saham PT. Telekomunikasi Indonesia dengan Support Vector Machine (SVM)," *J. Ilm. Edutic Pendidik. dan Inform.*, vol. 11, no. 1, pp. 22–33, 2024, doi: 10.21107/edutic.v11i1.22120.

[11] A. Karim, "Perbandingan Prediksi Kemiskinan di Indonesia Menggunakan Support Vector Machine (SVM) dengan Regresi Linear," *J. Sains Mat. dan Stat.*, vol. 6, no. 1, p. 107, 2020, doi: 10.24014/jsms.v6i1.9259.

[12] D. Putri, M. Khairunisa, and I. G. N. L. Wijayakusuma, "Perbandingan Metode Machine Learning untuk Analisis dan Prediksi Siklus Menstruasi," *J. Inf. Eng. Educ. Technol.*, vol. 8, no. 2, pp. 111–115, May 2025, doi: 10.26740/jieet.v8n2.p111-115.

[13] Syarif Aminul Khoiri and Abdul Wahid, "Analisis Kinerja Algoritma Machine Learning dalam Prediksi Harga Cryptocurrency," *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.*, vol. 9, no. 2, pp. 133–141, Aug. 2024, doi: 10.32528/justindo.v9i2.1965.

[14] A. A. Purnama, Y. Lukito, and N. A. Haryono, "Analisis Kinerja Support Vector Machine dan Moving Averages Convergence Divergence Untuk Saham-Saham Perbankan Indonesia," vol. 9, no. 1, pp. 87–99, 2025, [Online]. Available: https://doi.org/10.31603/komtika.v9i1.13459

[15] R. F. T. Wulandari and D. Anubhakti, “IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM) DALAM MEMPREDIKSI HARGA SAHAM PT. GARUDA INDONESIA TBK,” *IDEALIS Indones. J. Inf. Syst.*, vol. 4, no. 2, pp. 250–256, Jul. 2021, doi: 10.36080/idealis.v4i2.2847.

[16] A. A. Mahgfirah and P. I. Rahayu, “Analisis Support Vector Regression untuk Meramalkan Saham Perusahaan Dss di Indonesia,” *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 7, no. 1, pp. 44–54, 2025, doi: 10.35580/variansiunm356.

[17] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, “Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM),” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 375–384, Feb. 2024, doi: 10.57152/malcom.v4i2.1206.

[18] H. br Lumbantobing and R. Rahmaddeni, “Prediksi Harga Cryptocurrency Menggunakan Algoritma Support Vector Machine,” *Innov. J. Soc. Sci. Res.*, vol. 3, no. 2, pp. 7348–7355, 2023, [Online]. Available: http://j-innovative.org/index.php/Innovative/article/view/1213%0Ahttps://j-innovative.org/index.php/Innovative/article/download/1213/919

[19] B. E. S. Dewi, S. Haikal, H. . Sulistyowati, R. Fitriani, and D. P. K, “Penerapan Machine Learning Menggunakan Algoritma Random Forest Untuk Prediksi,” *J. TRIDI*, vol. 2, no. 1, pp. 20–31, 2024.

[20] S. Kharis, A. Haqqi Anna Zili, M. Malik, W. Nuryaningrum, and A. Putri, “Comparing machine learning models for Indonesia stock market prediction,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 38, p. 508, 2025, doi: 10.11591/ijeecs.v38.i1.pp508-516.

[21] P. Humairah and D. Agustina, “Stock Price Index Prediction Using Random Forest Algorithm for Optimal Portfolio,” *J. Varian*, vol. 8, pp. 113–124, 2024, doi: 10.30812/varian.v8i1.4276.

[22] S. A. Amellia Kharis, A. H. Anna Zili, M. Malik, W. Nuryaningrum, and A. Putri, “Comparing machine learning models for Indonesia stock market prediction,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 38, no. 1, p. 508, 2025, doi: 10.11591/ijeecs.v38.i1.pp508-516.

[23] D. Siswara, A. M. Soleh, and A. Hamim Wigena, “Classification Modeling with RNN-based, Random Forest, and XGBoost for Imbalanced Data: A Case of Early Crash Detection in ASEAN-5 Stock Markets,” *Sci. J. Informatics*, vol. 11, no. 3, pp. 569–582, 2024, doi: 10.15294/sji.v11i3.4067.