

Application of the Random Forest Algorithm for Stock Purchase Recommendation in Indonesian State-Owned Banks

Penerapan Algoritma Random Forest untuk Rekomendasi Pembelian Saham di Bank-Bank Milik Negara Indonesia

Huzaeni¹, Fadhil Akram²

^{1,2}Politeknik Negeri Lhokseumawe, Indonesia

Abstract

The stock market of Indonesian state-owned enterprises (SOEs) presents significant investment potential but is characterized by high volatility and complex influencing factors, making it difficult for investors—especially beginners—to make informed decisions. This research proposes a comprehensive stock recommendation system based on the Random Forest algorithm, designed to analyze and predict SOE stock performance through the integration of historical price data, financial report indicators, and market sentiment analysis. The system aims to identify optimal stock purchase recommendations by capturing both quantitative and qualitative market dynamics. Model performance was assessed using the Mean Absolute Error (MAE) metric, demonstrating that the Random Forest-based approach effectively enhances prediction accuracy and reliability. The findings suggest that this model can serve as a valuable decision-support tool for investors, improving market insight, minimizing financial risk, and supporting more data-driven investment strategies.

Keywords: Random Forest, Stock Recommendation System, State-Owned Banks, Financial Report, Market Sentiment, Predictive Analytics, Investment Decision Support.

Abstrak

Pasar saham BUMN Indonesia menawarkan potensi investasi yang signifikan, namun ditandai dengan volatilitas tinggi dan faktor-faktor pengaruh yang kompleks, sehingga menyulitkan investor—terutama pemula—untuk mengambil keputusan yang tepat. Penelitian ini mengusulkan sistem rekomendasi saham komprehensif berbasis algoritma Random Forest, yang dirancang untuk menganalisis dan memprediksi kinerja saham BUMN melalui integrasi data harga historis, indikator laporan keuangan, dan analisis sentimen pasar. Sistem ini bertujuan untuk mengidentifikasi rekomendasi pembelian saham optimal dengan menangkap dinamika pasar kuantitatif dan kualitatif. Kinerja model dinilai menggunakan metrik Mean Absolute Error (MAE), yang menunjukkan bahwa pendekatan berbasis Random Forest secara efektif meningkatkan akurasi dan keandalan prediksi. Temuan menunjukkan bahwa model ini dapat berfungsi sebagai alat pendukung keputusan yang berharga bagi investor, meningkatkan wawasan pasar, meminimalkan risiko keuangan, dan mendukung strategi investasi yang lebih berbasis data.

Kata kunci: Random Forest, Sistem Rekomendasi Saham, BUMN, Laporan Keuangan, Sentimen Pasar, Analisis Prediktif, Pendukung Keputusan Investasi.

I. INTRODUCTION

The capital market plays a crucial role in supporting national economic growth by providing opportunities for the public to invest and gain profits through stock trading activities. However, stock prices are highly volatile and influenced by various macroeconomic factors as well as internal company conditions, which often makes it difficult for investors—particularly beginners—to make informed and timely investment decisions [1]. Therefore, it is essential to develop an intelligent analytical system capable of assisting investors in understanding market trends and providing accurate, data-driven stock purchase recommendations.

In the investment domain, fundamental analysis and technical analysis are the two main approaches commonly used to assess stock performance. Fundamental analysis focuses on evaluating a company's financial health and performance, while technical analysis emphasizes historical price patterns as a basis for predicting future trends [2]. Combining these two approaches can provide a more comprehensive understanding of stock market behavior, especially for state-owned enterprises (SOEs), which play a strategic role in maintaining national economic stability [3]. The availability of historical data from platforms such as Yahoo Finance also enables more efficient, accurate, and real-time analysis processes [4].

With the rapid advancement of Artificial Intelligence (AI) and data science, various machine learning techniques have been widely applied to predict stock prices with a high degree of accuracy. Among these methods, the Random Forest algorithm has gained significant attention due to its capability to handle complex data structures, reduce overfitting, and produce stable prediction results [5]. Previous studies have demonstrated that this method is effective for stock prediction using historical price data [6], financial reports [7], and market sentiment analysis derived from various online news sources [8]. Furthermore, the implementation of web-based systems allows investors to access prediction results and investment recommendations in real time and interactively [9].

Based on these considerations, this study aims to apply the Random Forest algorithm to develop a stock recommendation system for Indonesian state-owned enterprises (SOEs) by integrating historical data, financial data, and market sentiment. The objective is to assist investors in making more accurate and profitable investment decisions [10].

II. METHOD

A. Data

The data used in this study were obtained from two primary sources: Google News and Yahoo Finance. The Google News dataset was collected to capture market sentiment related to the banking sector of Indonesian state-owned enterprises (SOEs), while Yahoo Finance provided the historical stock price data and financial reports of the respective companies. These two data sources form the foundation for building and analyzing the prediction model. The collected dataset consists of time series data covering the period from April 23, 2020, to May 23, 2025.

B. System Illustration

The system illustration serves as an essential step in explaining the overall functionality and workflow of the application. This section presents several diagrams that provide a comprehensive understanding of the structure and operational flow of the state-owned enterprise (SOE) stock recommendation system.

- 1) Context Diagram: The context diagram illustrates the interaction between the system and external entities, providing a clear understanding of the system's boundaries and the flow of information. It serves as a high-level representation that defines how users and external data sources communicate with the state-owned enterprise (SOE) stock recommendation system.

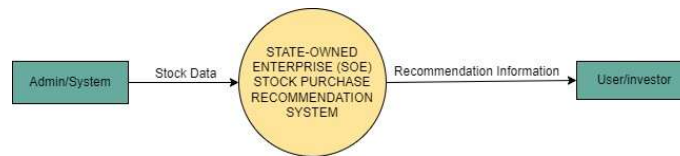


Figure 1. Context Diagram Illustration

Figure 1 presents the context diagram of the state-owned enterprise (SOE) stock recommendation system. The diagram illustrates the system's workflow, where the administrator inputs stock data, and the system processes it by integrating historical data, financial reports, and market sentiment. The processed information is then used to generate stock recommendations for investors, serving as a basis for informed investment decision-making.

- 2) DFD (Data Flow Diagram) Level 0: The Level 0 Data Flow Diagram (DFD) provides an overview of the main data flow and core processes within the system. It illustrates how data is managed, processed, and transformed from input to output. This diagram helps to visualize the overall functionality of the state-owned enterprise (SOE) stock recommendation system, including the flow of stock data, financial information, and market sentiment that are integrated and analyzed to generate investment recommendations for users..

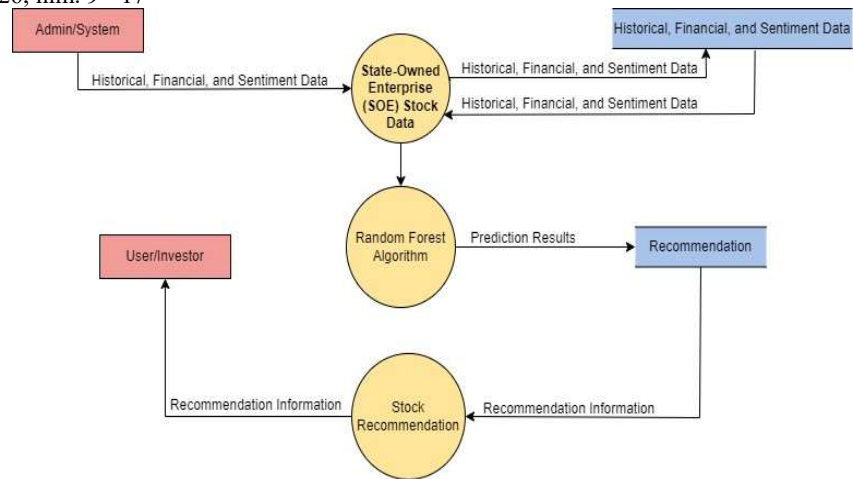


Figure 2. Level 0 Data Flow Diagram (DFD) Illustration

Figure 2 illustrates the Level 0 Data Flow Diagram of the state-owned enterprise (SOE) stock recommendation system based on the Random Forest algorithm. In this process, the administrator inputs historical stock data, financial reports, and market sentiment data. The system processes these inputs through the Random Forest algorithm to generate stock price predictions, which are then forwarded to the recommendation module. The final recommendations are presented to the user or investor as decision-support information for investment activities.

C. Method

This study employs the Random Forest Regressor algorithm to predict the stock prices of Indonesian state-owned enterprises (SOEs) by integrating three main data sources: historical stock data, financial reports, and market sentiment data.

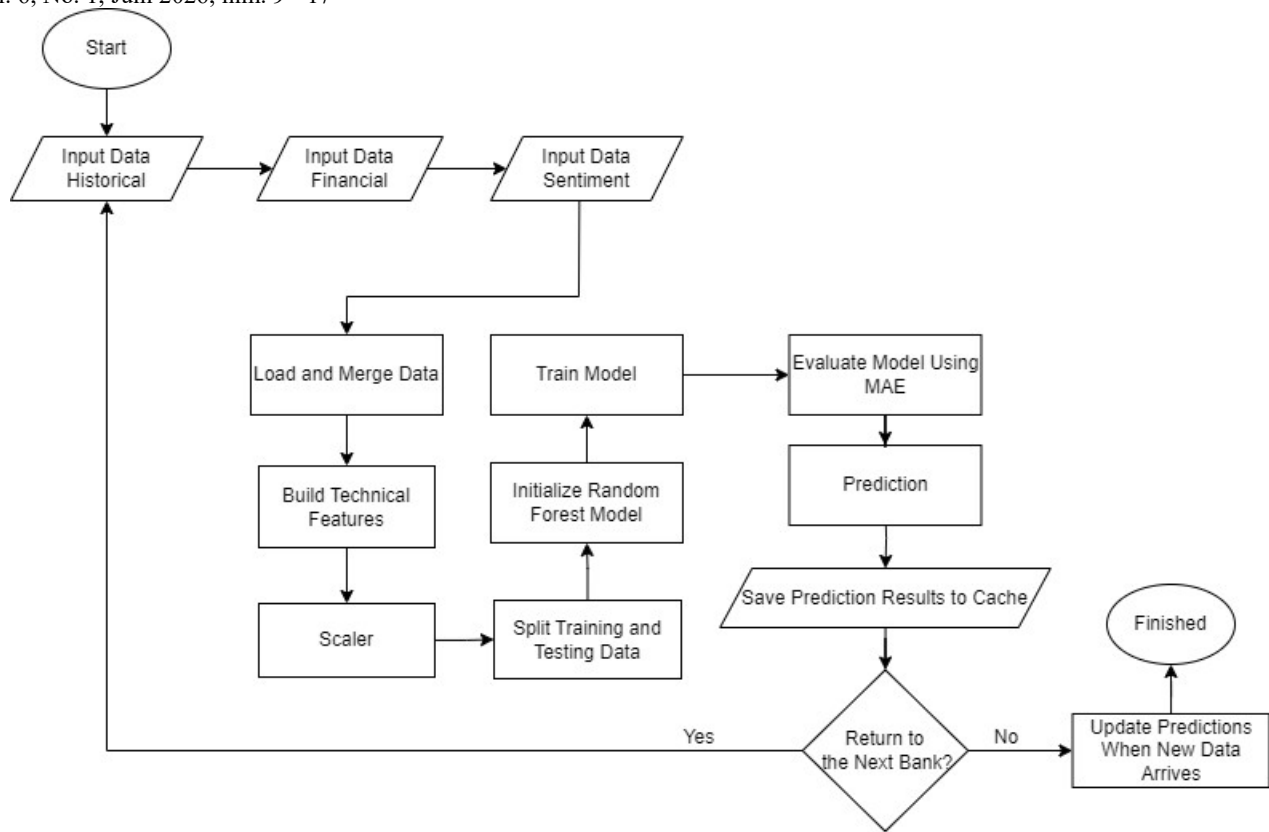


Figure 3. Method Illustration

Figure 3 illustrates the workflow of the Random Forest method applied in the stock price prediction system. The process begins with the collection of historical data, financial data, and market sentiment data. These datasets are then merged, transformed into technical features, and normalized to ensure suitability for modeling. Subsequently, the data are divided into training and testing sets to build and evaluate the Random Forest model, using the Mean Absolute Error (MAE) as the performance evaluation metric. The resulting predictions are stored in a cache and automatically updated whenever new data become available, enabling the system to provide investors with accurate and up-to-date stock recommendations.

D. Testing Illustration

The testing phase is conducted to evaluate the performance of the Random Forest model in predicting the stock prices of state-owned enterprises (SOEs). This process aims to assess how accurately the model can generate prediction results and determine its reliability in supporting investment decision-making.

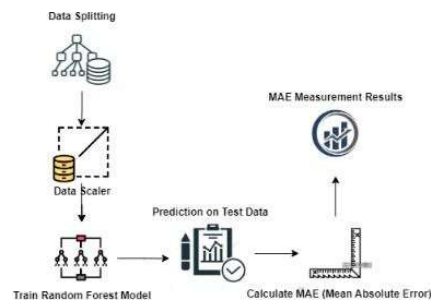


Figure 4. Testing Illustration

Figure 4 presents the evaluation process of the Random Forest model. The dataset is divided into 80% training data and 20% testing data, followed by a normalization process using a scaler to ensure data consistency. The model is trained using the training data and then applied to predict the testing data. The predicted results are compared with the actual values using the Mean Absolute Error (MAE) metric to measure the model's prediction accuracy.

III. RESULT AND DISCUSSION

A. Stock Prediction Results Based on Method and Data

Stock price prediction for the banking sector was conducted using historical data, which was visualized through candlestick charts. This analysis provides an overview of the price movement patterns of Bank BRI, BTN, BNI, and Mandiri. The prediction results indicate varying trends across the four issuers, with observable bullish and neutral tendencies depending on each bank's market behavior and historical performance.

1) Bank BRI

The stock movement pattern of Bank BRI demonstrates a relatively stable trend with a positive upward direction, indicating consistent performance and investor confidence in the market.

CANDLESTICK CHART PREDIKSI SAHAM BANK BRI

Tren BULLISH



Figure 5. Bank BRI Stock Chart

Figure 5 illustrates the long-term upward trend of Bank BRI's stock, indicating a bullish tendency. This reflects a positive market outlook and suggests stable growth potential for the company.

2) Bank BTN

The stock movement of Bank BTN exhibits relatively dynamic fluctuations over certain periods, reflecting varying market responses and short-term volatility in investor sentiment.



Figure 6. Bank BTN Stock Chart

Figure 6 illustrates the fluctuating price movement of Bank BTN's stock, which, despite its volatility, still demonstrates potential for recovery. The predictive indicators suggest a bullish trend, indicating a favorable short-term outlook for the stock's performance.

3) Bank BNI

The stock performance of Bank BNI demonstrates a relatively moderate and stable movement compared to other state-owned banks, indicating a balanced market response and lower volatility within the observed period.



Figure 7. BNI Bank Stock Chart

Figure 7 presents the stock trend of Bank BNI, which displays a relatively moderate and neutral movement. This indicates that market participants remain cautious, awaiting stronger momentum or external catalysts to define the subsequent price direction.

4) Bank Mandiri

The stock performance of Bank Mandiri demonstrates a consistent upward trend prior to experiencing a market correction. This pattern reflects strong investor confidence during the initial growth phase, followed by a natural adjustment as the market stabilizes.



Figure 8. Mandiri Bank Stock Chart

Figure 8 illustrates the stock trend of Bank Mandiri, which exhibits consistent growth leading up to a specific peak, followed by a corrective phase. The current prediction reflects a neutral sentiment, indicating that the stock price movement remains stable and balanced between buying and selling pressures.

B. Model Accuracy Evaluation Using Mean Absolute Error (MAE)

The evaluation was carried out using the Mean Absolute Error (MAE) metric to assess the predictive accuracy of the Random Forest model in forecasting the stock prices of four state-owned banks. This metric provides a quantitative measure of the average absolute difference between the predicted and actual stock prices, serving as an indicator of the model's overall performance and reliability.

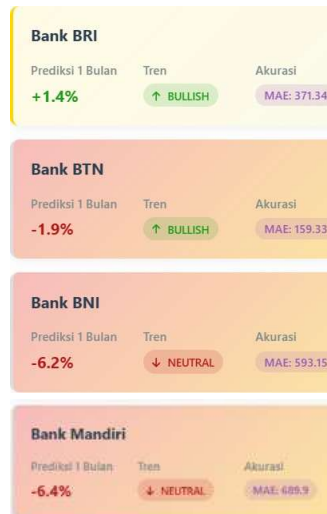


Figure 9. Model Accuracy Evaluation Based on Mean Absolute Error (MAE)

Figure 9 illustrates the results of the model accuracy evaluation using the Mean Absolute Error (MAE). The findings indicate that BTN's stock recorded the lowest MAE value of 159.33, suggesting superior predictive accuracy compared to BRI (371.34), BNI (593.15), and Mandiri (689.90). The lower MAE value observed in BTN reflects higher price stability and less volatility, while the greater fluctuations in BNI and Mandiri stocks contributed to higher prediction errors, indicating more dynamic and less predictable market behavior.

IV. CONCLUSION

The implementation of the Random Forest algorithm has demonstrated its effectiveness in developing a stock purchase recommendation system through the integration of historical stock data, financial reports, and sentiment analysis. This comprehensive data fusion enables the model to generate more accurate and reliable stock price predictions. The performance evaluation based on the Mean Absolute Error (MAE) indicates that Bank BTN achieved the lowest error value of 159.33, followed by BRI (371.34), BNI (593.15), and Mandiri (689.90). These results highlight the significant influence of data quality and market stability on prediction accuracy. In conclusion, the proposed system successfully provides price estimations that closely approximate real market values and delivers meaningful recommendations to assist investors in making well-informed and data-driven investment decisions.

REFERENCES

- [1] A. Runis Makkulau and I. Yuana, "YUME : Journal of Management Penerapan Analisa Fundamental dan Technical Analysis Sebagai Upaya Dalam Meningkatkan Keinginan Investasi Mahasiswa di Pasar Modal Syariah," YUME J. Manag., vol. 4, no. 3, pp. 165–180, 2021, doi: 10.37531/yume.vxix.432.
- [2] Z. Sun, "The Approaches to Predict Stock Prices by Random Forest," Appl. Comput. Eng., vol. 96, no. 1, pp. 162–168, Nov. 2024, doi: 10.54254/2755-2721/96/20241333.

- [3] A. Pratomo, R. Febrian Umbara, and A. A. Rohmawati, "Prediksi Pergerakan Harga Saham dengan Metode Random Forest Menggunakan Trend Deterministic Data Preparation (Studi Kasus Saham Perusahaan PT Astra International Tbk, PT Garuda Indonesia Tbk, dan PT Indosat Tbk)." [Online]. Available: <https://finance.yahoo.com/>
- [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] M. Iqbal Baihaqi, A. Syaripudin, and F. Agung Nugroho, "Implementation Of The Random Forest Algorithm In Stock Price Predictions Based On Historical Data Implementasi Algoritma Random Forest Pada Prediksi Harga Saham Berdasarkan Data Historis."
- [6] I. Of, Y. Finance, M. L. In, and E. O. F. Model, "Implementasi Sistem Prediksi Saham Real-Time dengan Integrasi Yahoo Finance API dan Machine Learning di Google Colab: Analisis dan Evaluasi Performa Model," vol. 15, no. 01, pp. 25–31, 2025.
- [7] A. M. Priyatno, L. Ningsih, and M. Noor, "Harnessing Machine Learning for Stock Price Prediction with Random Forest and Simple Moving Average Techniques," J. Eng. Sci. Appl., vol. 1, no. 1, pp. 1–8, 2024, doi: 10.69693/jesa.v1i1.1.
- [8] K. Valiant, Y. Lukito, and R. G. Santosa, "Sistem Prediksi Harga Saham LQ45 Dengan Random Forest Classifier," no. 2, 2019, doi: 10.21460/jutei.2019.32.187.
- [9] V. M. Tanto and T. A. Kurniawan, "Pengembangan sistem rekomendasi investasi saham berbasis web (studi kasus: reliance sekuritas malang)," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 6, no. 6, pp. 2813–2821, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [10] Y. Kaligis, "Analisis Tingkat Kesehatan Bank Dengan Menggunakan Metode Camel Pada Industri Perbankan Bumn Yang Terdaftar Di Bursa Efek Indonesia," J. Ris. Ekon. Manajemen, Bisnis dan Akunt., vol. 1, no. 3, pp. 263–272, 2013, doi: 10.35794/emba.v1i3.2015.