



ANALYSIS OF FEATURE IMPORTANCE ON THE CLOSING PRICE OF BBRI SHARES USING RANDOM FOREST

Warlina

Informatics, Faculty of Information Technology, Sebelas April University, Indonesia

Email: 220660121185@student.unsap.ac.id

(Article received: 20 Januari 2026; Revision: 20 Februari 2026; published: 15 April 2026)

Abstract

This study aims to analyze the factors influencing the closing price fluctuations of PT Bank Rakyat Indonesia (Persero) Tbk (BBRI) stocks using historical data from 2015–2024 through a Random Forest-based machine learning algorithm. It further evaluates the contribution of each predictor variable using the Permutation Importance method. The methodology includes data collection, cleaning, and feature engineering (Close_Lag1, Open_Lag1, High_Low_Range, Volume_Lag1). Data normalization was performed using MinMaxScaler, followed by chronological splitting (80% training, 20% testing). The experiment using RandomForestRegressor ($n_estimators=150$, $max_depth=10$) yielded a Root Mean Squared Error (RMSE) of 491.96 and an R-squared (R^2) of 0.0624. Although the statistical prediction capability is limited, the Permutation Importance analysis successfully identified 'Low' and 'High' prices as the most significant predictors. These findings highlight the complexity of the capital market and suggest the necessity of exploring macroeconomic variables and advanced time-series models like LSTM in future research to improve prediction accuracy.

Keywords: BBRI, Feature Importance, Machine Learning, Random Forest, Stock Price.

ANALISIS FEATURE IMPORTANCE TERHADAP HARGA PENUTUPAN SAHAM BBRI MENGGUNAKAN RANDOM FOREST

Abstrak

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi fluktuasi harga penutupan saham PT Bank Rakyat Indonesia (Persero) Tbk (BBRI) menggunakan data historis periode 2015–2024 melalui penerapan algoritma machine learning berbasis Random Forest, serta mengevaluasi kontribusi masing-masing variabel prediktor menggunakan metode Permutation Importance. Tahapan metodologi meliputi pengumpulan data, pembersihan dataset, hingga rekayasa fitur (feature engineering) seperti Close_Lag1, Open_Lag1, High_Low_Range, dan Volume_Lag1. Normalisasi data dilakukan menggunakan teknik MinMaxScaler dengan pembagian data secara kronologis (80% data latih dan 20% data uji). Eksperimen menggunakan RandomForestRegressor dengan parameter $n_estimators=150$ dan $max_depth=10$ menghasilkan nilai Root Mean Squared Error (RMSE) sebesar 491,96 dan R-squared (R^2) sebesar 0,0624. Meskipun kemampuan prediksi secara statistik terbatas, analisis Permutation Importance berhasil mengidentifikasi fitur 'Low' dan 'High' sebagai prediktor paling signifikan. Hasil ini menekankan tingginya kompleksitas pasar modal dan menyarankan pentingnya eksplorasi variabel makroekonomi serta penggunaan model time series yang lebih canggih seperti LSTM untuk penelitian mendatang.

Kata kunci: BBRI, Feature Importance, Harga Saham, Machine Learning, Random Forest.

1. PENDAHULUAN

Pasar saham merupakan instrumen investasi yang memiliki tingkat ketidakpastian tinggi karena pergerakan harganya dipengaruhi oleh berbagai faktor ekonomi, kondisi internal perusahaan, serta perilaku sentimen investor [1]. Dinamika harga yang fluktuatif ini menjadikan prediksi harga saham sebagai tantangan krusial bagi investor dan manajer portofolio dalam mengelola risiko investasi [2]. Salah satu indikator utama dalam pengambilan keputusan investasi adalah harga penutupan saham, yang merepresentasikan nilai akhir transaksi pada suatu periode perdagangan [3]. Prediksi harga penutupan yang akurat memungkinkan investor menentukan

strategi jual-beli secara lebih rasional, terutama pada pasar yang volatil [4].

Perkembangan teknologi informasi mendorong pemanfaatan metode *Machine Learning* sebagai pendekatan alternatif dalam prediksi harga saham karena keunggulannya dalam menangkap pola *non-linear* dibandingkan metode statistik tradisional [2] [5]. Di antara berbagai algoritma, *Random Forest* menjadi metode populer karena kemampuannya menggabungkan banyak pohon keputusan (*decision trees*) untuk menghasilkan prediksi yang stabil [1] serta ketahanannya terhadap *overfitting* pada data berdimensi tinggi [6].

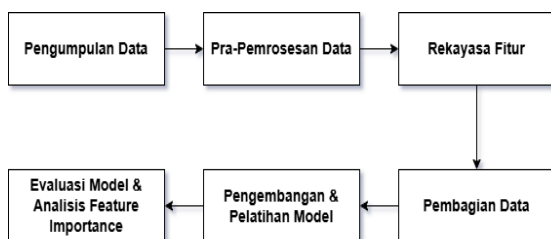
Beberapa penelitian terdahulu menunjukkan performa *kompetitif Random Forest*. Urrochman dkk. [1] menemukan bahwa *Random Forest* memberikan hasil lebih stabil dibandingkan *Support Vector Regression* (SVR). Senada dengan itu, Wu [6] menunjukkan bahwa meskipun *Long Short-Term Memory* (LSTM) unggul dalam dependensi jangka panjang, *Random Forest* tetap kompetitif dengan kompleksitas komputasi yang lebih rendah. Studi lain oleh Valiant, Lukito, dan Santosa [3] pada pasar Indonesia membuktikan efektivitas algoritma ini dalam memprediksi saham LQ45.

Namun, penerapan *Machine Learning* di bidang keuangan sering terkendala masalah interpretabilitas (*black box*), di mana sulit menjelaskan kontribusi variabel terhadap hasil prediksi [5][7]. Khaidem dkk. menyoroti pentingnya identifikasi fitur yang berpengaruh terhadap pergerakan harga untuk membantu interpretasi model. Oleh karena itu, penelitian ini tidak hanya berfokus pada akurasi prediksi, tetapi juga menganalisis peran setiap fitur menggunakan metode *Feature Importance*.

Penelitian ini bertujuan menerapkan algoritma *Random Forest* untuk memprediksi harga penutupan saham PT Bank Rakyat Indonesia (Persero) Tbk (BBRI) dan menganalisis kontribusi fitur (*feature importance*) sebagai dasar pendukung keputusan investasi. Saham perbankan dipilih karena karakteristiknya yang dinamis dan peran sentralnya dalam indeks pasar domestik.

2. METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif dengan alur kerja data science yang terstruktur untuk prediksi deret waktu finansial [8][9]. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan, rekayasa fitur, pelatihan model, pengembangan & pelatihan model, pembagian data, evaluasi model & analisis *Feature Importance*.



Gambar 1. Alur Metodologi Penelitian

2.1 Pengumpulan dan Pra-pemrosesan Data

Data yang digunakan adalah data historis harian saham BBRI periode 2015 hingga 2024 yang mencakup atribut *Open*, *High*, *Low*, *Close*, dan *Volume*. Penggunaan data historis ini merupakan pendekatan umum untuk menggambarkan perilaku pasar secara empiris [9][11]. Pra-pemrosesan dilakukan dengan menangani nilai hilang dan *noise*

untuk meningkatkan stabilitas model, sebagaimana direkomendasikan dalam studi terdahulu [12].

2.2 Rekayasa Fitur (Feature Engineering)

Untuk meningkatkan kemampuan model menangkap pola temporal, dilakukan penambahan fitur baru meliputi:

- Lag Features*: *Close_Lag1* dan *Open_Lag1* untuk menangkap ketergantungan harga sebelumnya.
- Volatility Feature*: *High_Low_Range* untuk merepresentasikan volatilitas intrahari.
- Volume Lag*: *Volume_Lag1* untuk melihat pengaruh aktivitas perdagangan sebelumnya. Pemanfaatan fitur berbasis lag dan rentang harga terbukti efektif merepresentasikan volatilitas data saham [4][13]. Seluruh fitur dinormalisasi menggunakan *Min-Max Scaling* agar berada pada rentang seragam [14]. *Dataset* dibagi secara kronologis (80% latih, 20% uji) untuk menjaga urutan temporal dan mencegah kebocoran informasi [15].

2.3 Model dan Evaluasi

Model dibangun menggunakan *Random Forest Regressor* dengan parameter $n_estimators=150$ dan $max_depth=10$. Algoritma ini dipilih karena kemampuannya menangani hubungan *nonlinier* dan mengurangi risiko *overfitting* [16][17]. Evaluasi model menggunakan metrik *Root Mean Squared Error* (RMSE) dan Koefisien Determinasi (R^2) [18]. Analisis fitur dilakukan menggunakan *Permutation Feature Importance* karena mampu memberikan interpretasi yang lebih objektif dibandingkan metode bawaan model [15].

3. HASIL DAN PEMBAHASAN

3.1 Karakteristik Data

Hasil statistik deskriptif menunjukkan harga penutupan saham BBRI memiliki standar deviasi yang besar, mengindikasikan volatilitas tinggi. Tingginya volatilitas ini merupakan karakteristik umum pasar saham yang banyak dilaporkan dalam penelitian berbasis machine learning [19][20]. Informasi tren jangka panjang ini menjadi dasar penting dalam penentuan fitur yang relevan [12].

Statistik deskriptif data historis saham BBRI (2015-2024) disajikan pada Tabel 1. Data menunjukkan harga penutupan memiliki nilai rata-rata 5.987,89 dengan standar deviasi 3.508,19.

Tabel 1. Statistik Deskriptif Harga Saham BBRI

	Open	High	Low	Close	Volume
coun	2194	2194	2194	2194	2194

t					
mean	5988.54	6051.87	5925.4	5987.89	1.29812e+08
std	3500.02	3532.82	3478.03	3508.19	1.05087e+08
min	2250	2270	2160	2170	6.3575e+06
25%	3922.5	3960	3870	3910	6.33133e+07
50%	4490	4530	4480	4480	1.09559e+08
75%	5600	5650	5575	5600	1.67164e+08
max	16450	16825	16425	16450	8.72607e+08

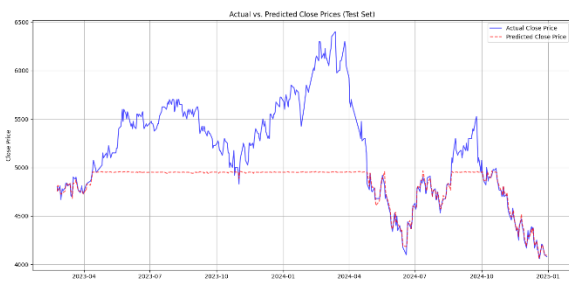
Visualisasi tren harga penutupan selama periode penelitian dapat dilihat pada Gambar 2, yang memperlihatkan fluktuasi harga yang signifikan dan adanya fase bullish serta koreksi.



Gambar 2. Tren Harga Saham BBRI (2015-2024)

3.2 Evaluasi Kinerja Model

Berdasarkan pengujian, model menghasilkan RMSE sebesar 491,96 dan R^2 sebesar 0,0624. Nilai R^2 yang rendah menunjukkan bahwa model hanya mampu menjelaskan sebagian kecil variasi harga. Visualisasi prediksi menunjukkan model cenderung mengikuti tren umum namun kurang mampu menangkap fluktuasi tajam. Perbandingan visual antara nilai prediksi model dan harga aktual ditampilkan pada Gambar 3.



Gambar 3. Perbandingan Harga Aktual vs Prediksi pada Data Uji

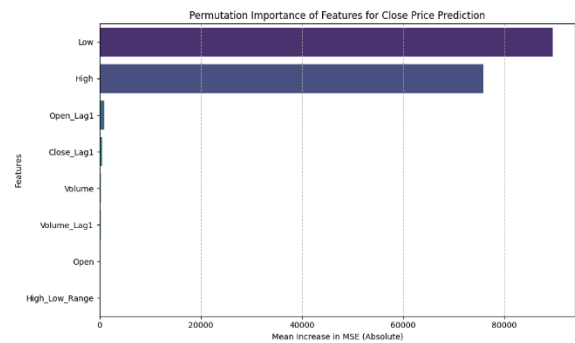
Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa *Random Forest* memiliki keterbatasan dalam menangkap

ketergantungan temporal jangka panjang dibandingkan model *deep learning* [1][21][22].

3.3 Analisis Feature Importance

Meskipun akurasi terbatas, *analisis Permutation Importance* berhasil mengidentifikasi variabel kunci. Fitur *Low* dan *High* teridentifikasi sebagai prediktor paling signifikan, sesuai dengan karakteristik pasar di mana rentang harga harian mencerminkan sentimen pasar. Fitur historis (*Open_Lag1*, *Close_Lag1*) juga menunjukkan relevansi kuat. Sementara itu, fitur Volume dan *Volume_Lag1* memiliki pengaruh lebih rendah namun tetap berkontribusi, sejalan dengan temuan bahwa indikator teknikal yang tepat dapat meningkatkan kinerja model [23][19].

Hasil analisis *Permutation Feature Importance* untuk mengidentifikasi variabel yang paling berpengaruh disajikan dalam Gambar 4 dan dirincikan nilainya pada Tabel 2.



Gambar 4. Visualisasi Pentingnya Fitur (Feature Importance)

Tabel 2. Hasil Permutation Feature Importance

Fitur	Rata-rata Kenaikan MSE
Low	89546.3
High	75860.9
Open_Lag1	986.393
Close_Lag1	479.991
Volume	316.813
Volume_Lag1	212.858
Open	89.1116
High_Low_Range	28.7021

4. DISKUSI

Berdasarkan hasil yang diperoleh pada bab sebelumnya, bagian ini membahas interpretasi temuan dan kaitannya dengan teori.

4.1 Analisis Akurasi Model

Nilai RMSE 491,96 dan R^2 0,0624 menunjukkan bahwa performa model *Random Forest* belum optimal. Seperti terlihat pada Gambar 3, garis prediksi cenderung "halus" dan gagal menangkap fluktuasi tajam. Hal ini sejalan dengan penelitian Wang yang menyatakan bahwa *Random Forest* memiliki keterbatasan dalam menangkap dependensi temporal jangka panjang dibandingkan model *Deep Learning* seperti LSTM [22]. Volatilitas pasar yang tinggi menyebabkan model kesulitan memprediksi lonjakan harga hanya berdasarkan data historis [12].

4.2 Interpretasi Faktor Penentu Harga

Fitur *Low* dan *High* teridentifikasi sebagai prediktor paling dominan (Tabel 2). Hal ini sesuai karakteristik pasar di mana rentang harga harian mencerminkan sentimen pasar. Menariknya, fitur Volume memiliki pengaruh yang relatif kecil, yang mengindikasikan bahwa dalam model ini, posisi harga (*price action*) lebih superior dibandingkan data likuiditas. Meskipun demikian, fitur *Lag* tetap memberikan kontribusi, menegaskan adanya aspek ketergantungan waktu pada data saham [11].

5. KESIMPULAN

Penerapan Random Forest Regressor untuk prediksi harga saham BBRI menghasilkan nilai RMSE 491,96 dan R^2 0,0624. Hasil ini mengindikasikan bahwa model belum optimal dalam menangkap volatilitas pasar yang ekstrem. Namun, analisis feature importance menegaskan bahwa harga terendah (*Low*) dan tertinggi (*High*) harian adalah fitur paling dominan. Disarankan penggunaan model time series yang lebih canggih seperti LSTM atau integrasi faktor eksternal (sentimen berita) untuk penelitian mendatang guna meningkatkan akurasi.

6. DAFTAR PUSTAKA

- [1] M. Y. Urrochman, H. Asy'ari, and F. A. Hizham, "Performance Comparison of Random Forest Regression, Svr Models in Stock Price Prediction," *J. Pilar Nusa Mandiri*, vol. 21, no. 1, pp. 16–23, 2025, doi: 10.33480/pilar.v21i1.6072.
- [2] T. Li, "Forecasting stock price trends with machine learning techniques," *Appl. Comput. Eng.*, vol. 46, no. 1, pp. 171–175, 2024, doi: 10.54254/2755-2721/46/20241307.
- [3] K. Valiant, Y. Lukito, and R. G. Santosa, "Sistem Prediksi Harga Saham LQ45 Dengan Random Forest Classifier," *J. Terap. Teknol. Inf.*, vol. 3, no. 2, pp. 127–136, 2019, doi: 10.21460/jutei.2019.32.187.
- [4] P. Humairah and D. Agustina, "Stock Price Index Prediction Using Random Forest Algorithm for Optimal Portfolio," *J. Varian*, vol. 8, no. 1, pp. 113–124, 2024, doi: 10.30812/varian.v8i1.4276.
- [5] Z. Guo, "Stock Market Price Prediction Using Machine Learning Models," *Adv. Econ. Manag. Polit. Sci.*, vol. 45, no. 1, pp. 102–111, 2023, doi: 10.54254/2754-1169/45/20230266.
- [6] H. Wu, "Comparison of Random Forest and LSTM in Stock Prediction," *Adv. Econ. Manag. Polit. Sci.*, vol. 86, no. 1, pp. 28–34, 2024, doi: 10.54254/2754-1169/86/20240936.
- [7] Y. Liu, "Stock Prediction Analysis Based on LogisticRegression and Random Forest," *MedScien*, vol. 1, no. 3, pp. 31–35, 2025, doi: 10.61173/kwhazd65.
- [8] E. Fitri and D. Riana, "Analisa Perbandingan Model Prediction Dalam Prediksi Harga Saham Menggunakan Metode Linear Regression, Random Forest Regression Dan Multilayer Perceptron," *METHOMIKA J. Manaj. Inform. dan Komputerisasi Akunt.*, vol. 6, no. 1, pp. 69–78, 2022, doi: 10.46880/jmika.vol6no1.pp69-78.
- [9] M. A. Lubis and Samsudin, "Using the Random Forest Method in Predicting Stock Price Movements," *Journal.Ittelkom-Pwt.Ac.Id/Index.Php/Dinda*, vol. 5, no. 2, pp. 258–267, 2025.
- [10] M. Zikri, F. Riana, and G. Fitri Laxmi, "Perbandingan Metode Long Short-Term Memory dan Double Random Forest dalam Prediksi Harga Penutupan Saham," *Krea-Tif J. Tek. Inform.*, vol. 12, no. 1, pp. 44–54, 2024, doi: 10.32832/krea-tif.v12i1.17468.
- [11] A. Rasyid, A. B. Muharam, and A. Solichin, "Prediksi Harga Saham Syariah Indonesia Berdasarkan Analisis Fundamental, Teknikal Dan Bandarmology Menggunakan Metode Random Forest," *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 10, no. 2, pp. 1663–1677, 2025, doi: 10.29100/jipi.v10i2.7855.
- [12] M. Jannah, "Tinjauan Algoritma Machine Learning untuk Prediksi Harga Saham: Studi Literatur," *J. Penelit. Multidisiplin Bangsa*, vol. 1, no. 7, pp. 608–613, 2024, doi: 10.59837/jpnmb.v1i7.121.
- [13] M. Faridita, "PREDIKSI HARGA PENUTUPAN BEBERAPA SAHAM MENGGUNAKAN RANDOM FOREST DAN ARTIFICIAL NEURAL NETWORK," *Encycl. Qual. Life Well-Being Res.*, pp. 273–273, 2024, doi: 10.1007/978-3-031-17299-1_300217.
- [14] C. Anindita, W. Laelatuji, and R. Rusmini, "Logistic Regression with Min-Max Scaling and TF-IDF for App Classification and Recommendation on Google Play Store," *J. Ilmu Komput. dan Inform.*, vol. 5, no. 1, pp. 79–90, 2025, doi: 10.54082/jiki.288.

- [15] M. Rusydi Hanan, L. Muflikhah, and F. A. Bachtiar, "Prediksi Nefropati Menggunakan Algoritma Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 5, pp. 2548–964, 2025, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [16] A. Nugroho and D. Harini, "Teknik Random Forest untuk Meningkatkan Akurasi Data Tidak Seimbang," *JSITIK J. Sist. Inf. dan Teknol. Inf. Komput.*, vol. 2, no. 2, pp. 128–140, 2024, doi: 10.53624/jsitik.v2i2.379.
- [17] T. Misriati and R. Aryanti, "Optimalisasi Random Forest dan Support Vector Machine dengan Hyperparameter GridSearchCV untuk Analisis Sentimen Ulasan PrimaKu," *J. Inf. Syst. Res.*, vol. 5, no. 4, pp. 1333–1341, 2024, doi: 10.47065/josh.v5i4.5347.
- [18] J. Manajemen, S. Informasi, and D. Aulia, "Perbandingan Algoritma Regresi Linier Berganda dan Random Forest Regression untuk Prediksi Konsentrasi Particulate Matter 2, 5," vol. 5, no. September, pp. 1060–1067, 2025.
- [19] A. Faqih and T. Sugihartono, "Perbandingan Algoritma XGBoost dan LSTM dalam Prediksi Harga Saham Tesla Menggunakan Data Tahun 2025," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 6, pp. 1563–1573, 2025, doi: 10.52436/1.jpti.836.
- [20] R. A. Gunawan, "PREDIKSI HARGA SAHAM PT. BANK BCA Tbk. DENGAN MENGGUNAKAN MACHINE LEARNING LONG SHORT-TERM MEMORY PERIODE 2014-2024," *Pros. Work. Pap. Ser. Manag.*, vol. 17, no. 1, pp. 1–5, 2025, doi: 10.25170/wpm.v17i1.7064.
- [21] V. T. Dhanalakshmi, C. Viola, and S. Mary, "Stock Price Prediction Using Random Forest Regression," *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 13, no. 5, 2023, doi: 10.56726/irjmets39128.
- [22] X. Wang, "Stock Price Prediction: A Comparative Study of Random Forest and LSTM Models," *Highlights Sci. Eng. Technol.*, vol. 107, pp. 117–123, 2024, doi: 10.54097/70a8b947.
- [23] A. Rebwar Mala Nabi, S. Saeed, R. M. Nabi, S. A. Saeed, and A. M. W. Abdi, "Feature Engineering for Stock Price Prediction," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 12s, pp. 2486–2496, 2020, [Online]. Available: <https://www.researchgate.net/publication/342339410>

