Penerapan Algoritma *XGBoost* untuk Prediksi Harga Saham *RTX Corporation* dalam Konteks Analisis Pasar

Michael Henokh[1)], Abertun Sagit Sahay[2)], Liyando Hermawan Hasibuan[3)]

[1)2)3)]Jurusan Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya
Jl. Yos Sudarso, Palangka Raya 73112
[1)]henokhsiga@gmail.com
[2)]abertun@it.upr.ac.id
[3)]liyando.hasibuan@it.upr.ac.id

Abstrak

Fluktuasi harga saham yang sangat dinamis di pasar modal menjadikan prediksi pergerakan harga sebagai tantangan kompleks bagi investor dan analis keuangan. Penelitian ini bertujuan untuk menerapkan dan mengevaluasi kinerja algoritma *Extreme Gradient Boosting* (*XGBoost*) dalam memprediksi harga saham harian *RTX Corporation* menggunakan pendekatan analisis teknikal. Penelitian ini menggunakan dataset historis dari *Yahoo Finance* periode 1962 hingga 2026 yang dibagi dengan rasio 85% untuk data latih dan 15% untuk data uji menggunakan *time-series split*. Model utama (*baseline*) dibangun dengan mengekstraksi 43 fitur indikator teknikal. Selain itu, penelitian ini menguji skenario penyederhanaan model melalui *feature selection*, mengembangkan model *hybrid* yang menggabungkan fitur teknikal dengan analisis sentimen berita berbasis algoritma *VADER* sebagai *Proof of Concept* (PoC), serta mendesain *dashboard* antarmuka interaktif menggunakan *Streamlit*. Hasil penelitian menunjukkan bahwa *XGBoost* mampu memprediksi harga saham dengan tingkat akurasi yang luar biasa. Model *baseline* menghasilkan *Mean Absolute Percentage Error (MAPE)* sebesar 1,2350%, *Root Mean Square Error (RMSE)* sebesar $1,50, dan skor R^2 mencapai 0,9972. Penerapan *feature selection* membuktikan bahwa penggunaan 5 fitur teratas mampu mengoptimalkan performa hingga menurunkan *MAPE* menjadi 1,19%. Pada eksperimen model *hybrid*, penambahan variabel sentimen berita terbukti meningkatkan *Directional Accuracy* secara drastis hingga 83,3%, meskipun secara statistik masih dibatasi oleh fenomena *small sample size* akibat limitasi akses data berita.

Kata kunci: Prediksi Harga Saham, *XGBoost*, Indikator Teknikal, *Feature Selection*, *Dashboard*, *RTX Corporation*.

Abstract

The highly dynamic stock price fluctuations in the capital market make price movement prediction a complex challenge for investors and financial analysts. This study aims to apply and evaluate the performance of the Extreme Gradient Boosting (XGBoost) algorithm in predicting the daily stock price of RTX Corporation using a technical analysis approach. This research utilizes a historical dataset from Yahoo Finance spanning from 1962 to 2026, which is divided into an 85% training set and a 15% testing set using a time-series split. The baseline model was built by extracting 43 technical indicator features. In addition, this study examines model simplification scenarios through feature selection, develops a hybrid model combining technical features with news sentiment analysis based on the VADER algorithm as a Proof of Concept (PoC), and designs an interactive dashboard interface using Streamlit. The research results demonstrate that

DOI:https://doi.org/10.47111/jointecoms.v6i3 *Received: 01-09-2026*
Accepted: 10-09-2026

XGBoost is capable of predicting stock prices with exceptional accuracy. The baseline model produced a Mean Absolute Percentage Error (MAPE) of 1.2350%, a Root Mean Square Error (RMSE) of $1.50, and an R^2 score reaching 0.9972. The application of feature selection proved that the use of the top 5 features was able to optimize performance, reducing the MAPE to 1.19%. In the hybrid model experiment, the addition of news sentiment variables was proven to drastically increase Directional Accuracy to 83.3%, although statistically, it is still constrained by the small sample size phenomenon due to limitations in news data access.

Keywords: *Stock Price Prediction, XGBoost, Technical Indicators, Feature Selection, Dashboard, RTX Corporation.*

1. Pendahuluan

Pasar modal merupakan salah satu instrumen ekonomi vital yang mencerminkan dinamika ekonomi suatu negara, di mana harga saham seringkali berfluktuasi sebagai respons terhadap berbagai faktor, baik internal maupun eksternal perusahaan. Kemampuan untuk memprediksi pergerakan harga saham dengan akurasi tinggi menjadi sangat krusial bagi investor, analis pasar, dan pembuat kebijakan untuk mengambil keputusan yang informatif dan strategis. Namun, kompleksitas pasar saham, yang dipengaruhi oleh data ekonomi, peristiwa geopolitik, sentimen pasar, dan kinerja internal perusahaan, menjadikannya tantangan besar dalam upaya prediksi [2].

Dalam menghadapi tantangan ini, perkembangan pesat dalam bidang kecerdasan buatan (AI) dan pembelajaran mesin (Machine Learning/ML) telah membuka peluang baru untuk menganalisis data pasar saham yang besar dan kompleks. Berbagai algoritma telah dikembangkan dan diterapkan untuk mendeteksi pola tersembunyi dan memprediksi tren harga saham. Salah satu algoritma yang menunjukkan kinerja superior dalam tugas prediksi adalah *Extreme Gradient Boosting* (*XGBoost*). *XGBoost* adalah implementasi open-source yang efisien dari algoritma Gradient Boosting yang dikenal karena skalabilitas dan kecepatannya yang superior, serta kemampuannya menangani missing values, memberikan feature importance untuk interpretasi, dan memanfaatkan regularization untuk mencegah overfitting [2].

Raytheon Technologies Corporation (RTX), sebagai perusahaan multinasional kedirgantaraan dan pertahanan terkemuka di dunia, mewakili studi kasus analisis ini. Kinerja keuangan RTX sangat relevan dengan isu-isu geopolitik, kontrak pemerintah, dan inovasi teknologi [2].

Penelitian ini berfokus pada penerapan algoritma *XGBoost* untuk mendeteksi pola dan memprediksi harga saham *RTX Corporation* menggunakan indikator teknikal sebagai fitur utama. Selain itu, penelitian ini juga mengintegrasikan fitur non-teknikal (analisis sentimen berita menggunakan algoritma *VADER*) ke dalam pemodelan prediktif *XGBoost* sebagai eksperimen hybrid, serta mengembangkan *dashboard* interaktif untuk visualisasi hasil prediksi secara real-time [4].

2. Tinjauan Pustaka

2.1. *XGBoost*

XGBoost (*Extreme Gradient Boosting*) adalah algoritma pembelajaran ensemble yang efisien, merupakan pengembangan dari Gradient Boosting Decision Tree (GBDT) dan dipandang sebagai model aditif yang terdiri dari pohon CART (Classification and Regression Tree) [2].

Proses pemodelan *XGBoost* melibatkan penambahan pohon secara iteratif untuk memperbaiki prediksi dari model sebelumnya [2]. Fungsi objektif *XGBoost* dirancang untuk meminimalkan loss prediksi sekaligus mengontrol kompleksitas model untuk mencegah overfitting, yang didefinisikan dalam Persamaan (1) [2].

$$obj = \sum l(y_i, \hat{y}_i) + \sum \Omega(f_k) \quad (1)$$

Di mana *l* adalah fungsi loss yang mengukur seberapa baik model memprediksi nilai aktual, dan Ω adalah istilah regularisasi yang digunakan untuk mengontrol kompleksitas model *f_k* [2].

2.2. Analisis Teknikal dalam Prediksi Harga Saham

Analisis teknikal berfokus pada pengamatan data historis harga saham dan volume perdagangan untuk mengidentifikasi pola pergerakan harga dan mengantisipasi tren masa depan [2]. Beberapa indikator teknikal umum yang digunakan sebagai fitur dalam model prediksi meliputi Simple Moving Average (SMA), Exponential Moving Average (EMA), Moving Average Convergence/Divergence (MACD), dan Relative Strength Index (RSI) [2].

2.3. Analisis Berita dan Sentimen dengan *VADER*

Analisis berita dan sentimen diposisikan sebagai komponen pendukung untuk memberikan konteks non-teknikal [2]. Algoritma sentimen yang digunakan adalah *VADER* (Valence Aware Dictionary and sEntiment Reasoner), yaitu pendekatan lexicon dan rule-based yang efektif untuk teks berita [2]. Skor sentimen dihitung menggunakan nilai compound pada rentang -1 (sangat negatif) sampai +1 (sangat positif) [3].

3. Metode Penelitian

3.1. Pengumpulan dan Pra-Pemrosesan Data

Dataset yang digunakan adalah data historis harga saham harian *RTX Corporation* yang diakses secara real-time dari *Yahoo Finance* melalui Application Programming Interface (API) menggunakan library yfinance [1]. Periode pengambilan data mencakup 2 April 1962 hingga waktu eksperimen dilakukan, dengan total 16.053 baris data setelah proses pembersihan (menghapus nilai NaN menggunakan dropna) [2]. Data kemudian dibagi menggunakan pendekatan time-series split dengan rasio 85% untuk data latih (Training Set, 13.645 records) dan 15% untuk data uji (Testing Set, 2.408 records) [3].

3.2. Rekayasa Fitur dan Pemodelan

Tahap rekayasa fitur melibatkan perhitungan 43 fitur teknikal ke dalam dataset yang terdiri dari 9 kategori: Moving Average (6 fitur), Price Position (4 fitur), Momentum (8 fitur), Volatilitas (8 fitur), Volume (4 fitur), Price Dynamics (3 fitur), Lag Features (3 fitur), Temporal (4 fitur), dan Returns/Volatility (3 fitur) [2]. Seluruh fitur dinormalisasi menggunakan RobustScaler sebelum proses training [2].

Hyperparameter tuning dilakukan menggunakan Grid Search dengan 5-fold cross-validation berbasis Time Series Split untuk mengoptimalkan performa model *XGBoost* [2]. Metrik utama yang digunakan untuk evaluasi model adalah *Mean Absolute Percentage Error (MAPE)* yang dihitung dengan Persamaan (2) dan *Root Mean Square Error (RMSE)* dengan Persamaan (3) [2].

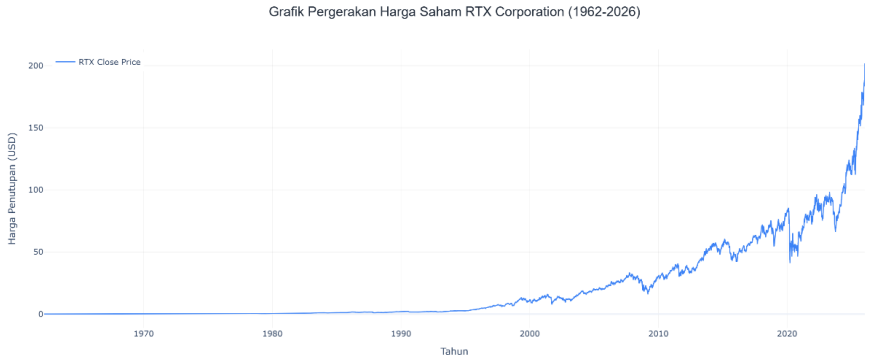
$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - y_i)^2} \quad (3)$$

4. Hasil dan Pembahasan

4.1. Deskripsi Data dan Hyperparameter Tuning

Data yang digunakan memiliki rata-rata harga penutupan $20.74 dengan standar deviasi $31.42, mengindikasikan volatilitas yang signifikan sepanjang periode 63+ tahun pengamatan [2].



Gambar 1. Grafik pergerakan harga saham *RTX*

Setelah proses *Grid Search*, diperoleh konfigurasi hyperparameter optimal untuk *XGBoost* seperti yang ditunjukkan pada Tabel 1.

Parameter	Nilai	Fungsi
n_estimators	500	Jumlah decision trees yang dibangun secara iteratif
max_depth	7	Kedalaman maksimum setiap tree
learning_rate	0.08	Tingkat pembelajaran (shrinkage factor)
subsample	0.8	Proporsi sampel untuk training setiap tree

Tabel 1. Konfigurasi Hyperparameter *XGBoost*

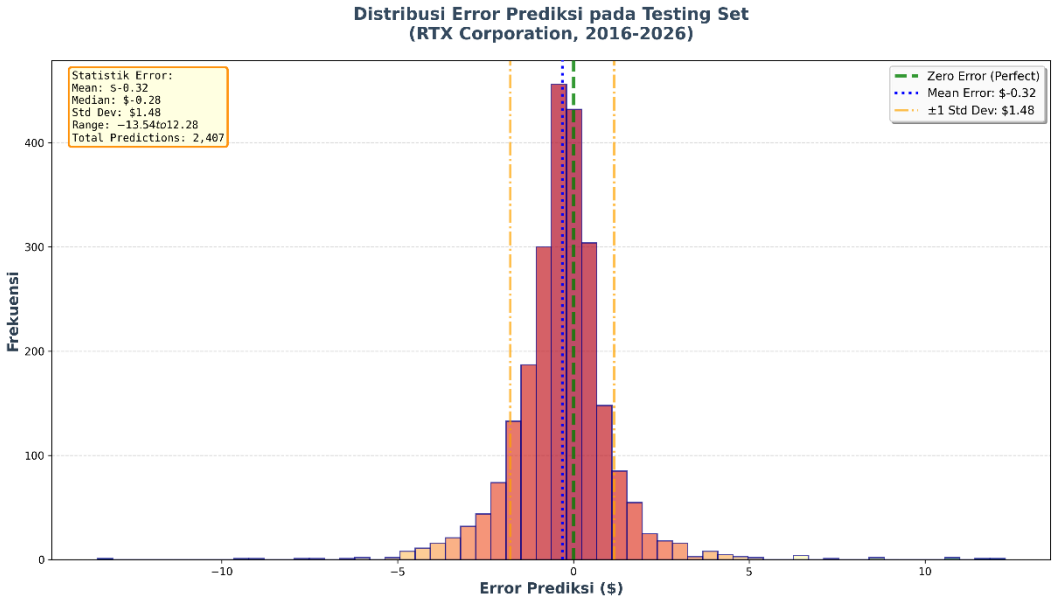
4.2. Evaluasi Performa Model Baseline

Model *XGBoost* dievaluasi menggunakan Testing Set yang merepresentasikan kondisi pasar pada periode 2016-2026 [2]. Hasil pengujian disajikan pada Tabel 2.

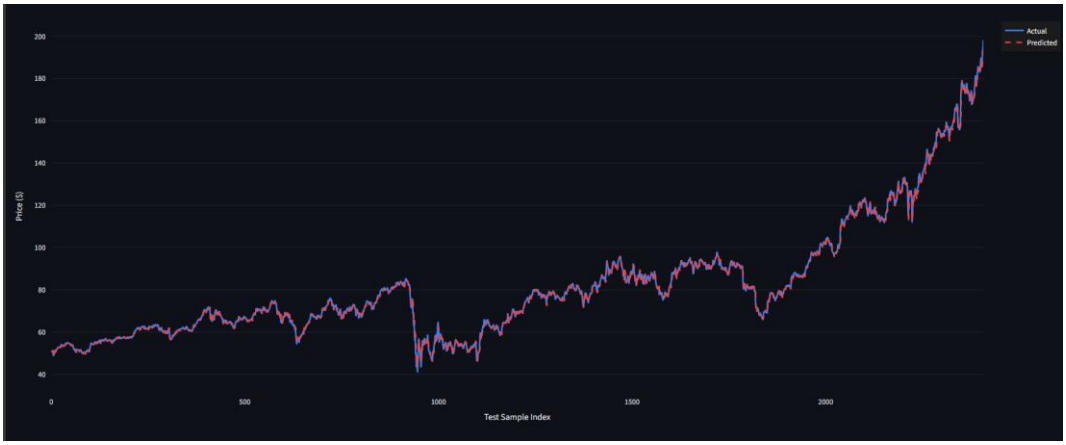
Metrik	Nilai	Interpretasi
MAPE	1.2350%	Error rata-rata 1.2350%, akurasi 98.765%
RMSE	$1.50	*Root Mean Square Error* dalam dollar
MAE	$0.99	Mean Absolute Error kurang dari $1
R^2 Score	0.9972	Model menjelaskan 99.72% variasi harga

Tabel 2. Hasil Evaluasi Model pada Testing Set

Nilai *MAPE* 1.2350% menunjukkan bahwa rata-rata galat prediksi model sangat kecil, yang berarti prediksi model sangat dekat dengan nilai aktual [2]. R^2 Score sebesar 0.9972 mengindikasikan bahwa model berhasil menjelaskan 99.72% variasi dalam harga saham RTX [2]. Distribusi error prediksi juga berpusat di sekitar nol, mengindikasikan tidak adanya *systematic bias* [5].



Gambar 2. Distribusi error prediksi



Gambar 3. Harga aktual vs harga prediksi

4.3. *Feature Selection* dan Feature Importance

Proses *feature selection* berbasis importance dari model *XGBoost* (dihitung berdasarkan *gain*) menunjukkan bahwa fitur yang merepresentasikan momentum, tren, volatilitas, dan memori jangka pendek memiliki pengaruh terbesar [5]. Lima fitur teratas ditunjukkan pada Tabel 3.

No	Nama Fitur	Nilai Importance
1	EMA_5	0.212255 (21.23%)
2	BB_Upper_20	0.207236 (20.72%)
3	EMA_20	0.198693 (19.87%)
4	Close_Lag_1	0.198220 (19.82%)
5	BB_Distance_20	0.183596 (18.36%)

Tabel 3. Daftar 5 fitur terpilih berdasarkan importance tertinggi

Perbandingan model baseline dengan model *feature selection* (5 fitur) menunjukkan peningkatan performa. Nilai *MAPE* berhasil ditekan menjadi 1.19% dan R^2 naik menjadi 0.9974, membuktikan bahwa penyederhanaan model ini efektif tanpa mengorbankan akurasi [2].

4.4. Eksperimen Model Hybrid

Sebagai Proof of Concept (PoC), penelitian ini menggabungkan 5 fitur teknikal terbaik dengan 9 fitur sentimen berita *VADER* [2]. Meskipun eksperimen ini meningkatkan *Directional Accuracy* menjadi 83.3%, model mengalami *overfitting* yang ditandai dengan nilai R^2 negatif (-1.2266) [2]. Hal ini disebabkan oleh keterbatasan akses API berita historis, sehingga data latih hanya terdiri dari 39 baris data, yang memicu fenomena *Curse of Dimensionality* [5].

4.5. Implementasi *Dashboard*

Untuk memfasilitasi visualisasi hasil prediksi secara interaktif, penelitian ini mengembangkan *dashboard* menggunakan framework *Streamlit* [2]. *Dashboard* menyediakan 8 tab utama untuk eksplorasi prediksi harga real-time, analisis performa, kalkulator investasi, dan visualisasi indikator teknikal [2].



Gambar 4. Antarmuka utama *dashboard*

5. Kesimpulan

Algoritma *XGBoost* terbukti sangat akurat dalam memprediksi harga saham *RTX Corporation* menggunakan 43 fitur teknikal dengan menghasilkan *MAPE* sebesar 1,2350%, *RMSE* $1,50, dan skor R^2 sebesar 0,9972 [2]. Melalui proses *feature selection*, penggunaan subset 5 fitur terpilih terbukti menyederhanakan model dan berhasil menekan nilai *MAPE* menjadi 1,19% [2]. Hal ini menunjukkan bahwa model sangat mengandalkan indikator tren dan volatilitas (seperti EMA dan Bollinger Bands) untuk memprediksi harga [2].

Eksperimen integrasi fitur non-teknikal (sentimen berita *VADER*) sebagai model hybrid menunjukkan peningkatan pada *Directional Accuracy* (83.3%), namun secara statistik mengalami kegagalan generalisasi akibat minimnya sampel data latih yang tersedia [2]. Seluruh hasil analisis ini telah diintegrasikan dengan baik ke dalam sebuah *dashboard* berbasis web untuk mendukung pemantauan pasar secara real-time [2].

Daftar Pustaka

[1] B. Jange, "Prediksi Harga Saham Bank BCA Menggunakan *XGBoost*," *ARBITRASE: Journal of Economics and Accounting*, vol. 3, no. 2, pp. 231–237, 2022.

[2] J. Monterio, "Raytheon Technologies Corporation: An Analysis of Leadership, Ethics, and Social Responsibility," *The Journal of Global Awareness*, vol. 3, no. 1, pp. 1–14, 2022.

[3] M. Zebua and E. Gea, "Pengaruh Faktor Eksternal Terhadap Harga Saham: Analisis Dengan Data Historis Dari *Yahoo Finance*," *Identik: Jurnal Ilmu Ekonomi, Pendidikan Dan Teknik*, vol. 02, no. 01, 2025.

[4] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Applied Soft Computing Journal*, vol. 90, 2020.

[5] S. Carta, A. Corriga, A. Ferreira, A. S. Podda, and D. R. Recupero, "A multi-layer and multi-ensemble stock trader using deep learning and deep reinforcement learning," *Applied Intelligence*, vol. 51, no. 2, pp. 889–905, 2021.

[6] Bijaksana Putra Negara, A. (2022). *Penerapan Algoritma Model Regresi pada Angka New Active Cases Covid-19 di Indonesia*. *01*(1). https://doi.org/10.26418/juara.v1i1.59941

[7] Arifin, S., & Febryanto, B. A. (2025). Analisis Sentimen Pengguna Aplikasi Merdeka Mengajar Menggunakan Metode *Vader* Lexicon. *SISFOTENIKA*, *15*(1), 1–11. https://doi.org/10.30700/sisfotenika.v15i1.513

[8] Fazlija, B., & Harder, P. (2022). Using Financial News Sentiment for Stock *Price* Direction *Prediction*. *Mathematics*, *10*(13). https://doi.org/10.3390/math10132156

[9] Hakkal, S., & Lahcen, A. A. (2024). *XGBoost* To Enhance Learner Performance *Prediction*. *Computers and Education: Artificial Intelligence*, *7*. https://doi.org/10.1016/j.caeai.2024.100254

[10] Ibrahem Ahmed Osman, A., Najah Ahmed, A., Chow, M. F., Feng Huang, Y., & El-Shafie, A. (2021). Extreme gradient *boosting* (*Xgboost*) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, *12*(2), 1545–1556. https://doi.org/10.1016/j.asej.2020.11.011

[11] Mohammed, R. H., & El-saieed, A. M. (2025). Chaotic billiards optimized *hybrid transformer* and *XGBoost* model for *robust* and sustainable *time series forecasting*. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-10641-7

[12] Nyoman, N., Pinata, P., Sukarsa, M., Kadek, N., & Rusjayanthi, D. (2020). Prediksi Kecelakaan Lalu Lintas di Bali dengan *XGBoost* pada Python. *Jurnal Ilmiah Merpati*, *8*(3). https://doi.org/10.24843/JIM.2020.v08.i03.p04