

SKRIPSI

**PENERAPAN ALGORITMA *XGBOOST* UNTUK PREDIKSI HARGA
SAHAM *RTX Corporation (RTX)* DALAM KONTEKS ANALISIS PASAR**



**DISUSUN OLEH:
MICHAEL HENOKH
213030503111**

**JURUSAN/PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS PALANGKA RAYA
2026**

**PENERAPAN ALGORITMA XGBOOST UNTUK DETEKSI POLA DAN PREDIKSI HARGA
SAHAM RTX Corporation (RTX) DALAM KONTEKS ANALISIS PASAR**

SKRIPSI

Sebagai salah satu syarat untuk menyelesaikan Program Strata-I pada Jurusan Teknik Informatika
Fakultas Teknik Universitas Palangka Raya

Oleh

MICHAEL HENOKH

213030503111

Telah dipertahankan didepan tim penguji, pada :

Hari/Tanggal : Senin, 13 Juli 2026

Waktu : 15:00-16:30 WIB

Tempat : Ruang Ujian I

1. **NOVERA KRISTIANTI, S.T., M.T.**
NIP. 19931116 202203 2 012

 (Ketua Penguji)

2. **ABERTUN SAGIT SAHAY, S.T., M.Eng.**
NIP. 19751212 200312 1 002

 (Pembimbing Utama/Pertama)

3. **LIYANDO HERMAWAN
HASIBUAN, S.T., M.Kom.**
NIP. 19960722 202406 1 001
Pendamping/Kedua)

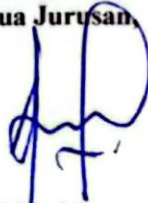
 (Pembimbing

4. **NAHUMI
NUGRAHANINGSIH, S.T., M.T., Ph.D.**
NIP. 19791009 200801 2 016

 (Anggota)

Mengetahui :

**Jurusan / Program Studi Teknik Informatika
Fakultas Teknik Universitas Palangka Raya
Ketua Jurusan**



ARIELA LESTARI, S.Kom., M.Cs., Ph.D.
NIP. 19800322 200501 2 004

**PENERAPAN ALGORITMA XGBOOST UNTUK PREDIKSI HARGA SAHAM RTX
Corporation (RTX) DALAM KONTEKS ANALISIS PASAR**

SKRIPSI

Sebagai salah satu syarat untuk menyelesaikan Program Strata-1 pada Jurusan Teknik
Informatika Fakultas Teknik Universitas Palangka Raya

Oleh

MICHAEL HENOKH

NIM. 213030503111

Disetujui untuk diajukan di Seminar Akhir Skripsi,
Palangka Raya, Senin 13 Juli 2026

Pembimbing I,



Abertun Sagit Sahay, S.T., M.Eng.
19751212 200312 1 002

Pembimbing II,



Liyando Hermawan Hasibuan, S.T., M.Kom.
19960722 202406 1 001

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS PALANGKA RAYA**

2026

PERNYATAAN

Dengan ini saya menyatakan sebenar – benarnya bahwa dalam Skripsi ini tidak terdapat karya ilmiah yang pernah diajukan untuk memperoleh gelar kesarjanaan disuatu Perguruan Tinggi, serta tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam Skripsi ini dan disebutkan dalam Tinjauan Pustaka

Palangka Raya, 13 Juli 2026



Michael Henokh

213030503111

RIWAYAT PENYUSUN

Data Diri

Nama : Michael Henokh
NIM : 213030503111
Fakultas : Teknik
Jurusan/Program Studi : Teknik Informatika
Jenjang : Strata 1 (S-1)
Jenis Kelamin : Laki-Laki
Tempat, Tanggal Lahir : Bandung, 3 Juli 2003
Agama : Kristen Protestan
Anak Ke : 1 (Satu)
Alamat : Jl.Aries No.232
No Telepon/HP : 081250456616



Data Orang Tua

Nama Ayah : Daniel Sinaga
Pekerjaan Ayah : Pendeta
Nama Ibu : Dorcas Yunika
Pekerjaan Ibu : Ibu Rumah Tangga
Alamat Orang Tua : Jl.Aries No.232
No. HP Orang Tua : 081250902447

Riwayat Pendidikan

SD : SDK Santo Don Bosco (Tahun Lulus 2015)
SMP : SMPN 2 Palangka Raya (Tahun Lulus 2018)
SMA : SMAN 2 Palangka Raya (Tahun Lulus 2021)

Palangka Raya, 13 Juli 2026

Michael Henokh

213030503111

HALAMAN PERSEMBAHAN

Segala puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat, karunia, dan kelancaran yang diberikan sehingga penyusunan skripsi ini dapat diselesaikan dengan baik.

Dengan segala kerendahan hati, karya sederhana ini penulis persembahkan dan haturkan rasa terima kasih yang mendalam kepada:

1. Tuhan Yesus Kristus, atas segala kasih karunia, penyertaan, kekuatan, dan pengharapan yang tak pernah berkesudahan. Terima kasih atas setiap jalan keluar di saat buntu dan kekuatan di saat lelah, sehingga penulis mampu menyelesaikan garis akhir perkuliahan ini.
2. Kedua Orang Tua tercinta, yang selalu menjadi sumber kekuatan utama bagi penulis. Terima kasih yang tak terhingga atas setiap doa yang dipanjatkan tanpa henti, pengorbanan yang tak ternilai harganya, serta dukungan moril maupun materiel yang selalu diberikan. Karya ini merupakan bentuk kecil dari rasa bakti dan cinta penulis untuk kalian.
3. Bapak dan Ibu Dosen Pembimbing, atas segala kesabaran, waktu, ilmu, serta arahan yang telah diberikan. Terima kasih atas setiap revisi dan masukan berharga yang terus menuntun penulis agar dapat menyajikan hasil penelitian yang terbaik.
4. Teman-teman seperjuangan, khususnya rekan-rekan mahasiswa Teknik Informatika, terima kasih atas segala solidaritas, bantuan, diskusi panjang, dan momen berbagi keluh kesah. Kehadiran kalian membuat perjalanan penyusunan skripsi ini terasa lebih ringan dan penuh makna.
5. Semua Pihak yang Tidak Dapat Disebutkan Satu per Satu, yang telah memberikan bantuan baik secara langsung maupun tidak langsung dalam proses penyelesaian skripsi ini.

Akhir kata, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada diri sendiri. Terima kasih sudah bertahan melewati masa-masa sulit, tidak menyerah pada keadaan, mau terus belajar dari kesalahan, dan berjuang keras menyelesaikan apa yang telah dimulai hingga akhirnya bisa mencapai titik ini.

KATA PENGANTAR

Puji dan syukur senantiasa penulis panjatkan ke hadirat Tuhan Yesus Kristus, atas segala kasih karunia, rahmat, dan penyertaan-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul "Penerapan Algoritma *XGBoost* untuk Prediksi Harga Saham *RTX Corporation (RTX)* dalam Konteks Analisis Pasar" dengan baik dan tepat waktu. Skripsi ini disusun sebagai salah satu syarat mutlak untuk memperoleh gelar Sarjana Teknik (S.T) pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya.

Penulis menyadari sepenuhnya bahwa dalam proses penyusunan skripsi ini, penulis banyak mendapatkan bimbingan, arahan, motivasi, serta bantuan dari berbagai pihak, baik secara moril maupun materiel. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah membantu, baik secara langsung maupun tidak langsung selama proses penelitian dan penulisan skripsi ini.

Penulis menyadari sepenuhnya bahwa skripsi ini masih jauh dari kata sempurna dikarenakan keterbatasan ilmu, pengalaman, serta wawasan yang penulis miliki. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari berbagai pihak demi perbaikan dan kesempurnaan penelitian ini di masa yang akan datang.

Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat, wawasan, serta inspirasi bagi para pembaca, khususnya bagi pengembangan ilmu pengetahuan di bidang Teknik Informatika.

Palangka Raya, 13 Juli 2026

Michael Henokh

213030503111

PENERAPAN ALGORITMA *XGBOOST* UNTUK PREDIKSI HARGA SAHAM *RTX Corporation (RTX)* DALAM KONTEKS ANALISIS PASAR

Michael Henokh

Jurusan Teknik Informatika Fakultas Teknik Universitas Palangka Raya
Kampus Tunjung Nyaho Jl. Yos Sudarso Palangka Raya 73112
henokhsiga@gmail.com

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*, sebuah perusahaan kedirgantaraan dan pertahanan multinasional, 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 (didominasi oleh fitur *momentum* dan tren seperti *EMA* dan *Bollinger Bands*) 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. Seluruh hasil analisis teknikal dan prediksi ini telah diintegrasikan dengan baik ke dalam sebuah *dashboard* berbasis *web* untuk mendukung pemantauan pasar secara *real-time*.

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

***Application of the XGBoost Algorithm for RTX Corporation (RTX) Stock Price
Prediction in the Context of Market Analysis***

Michael Henokh

*Department of Information Technology, Faculty of Engineering, University
of Palangka Raya Campus
Tunjung Nyaho Jl. Yos Sudarso Palangka Raya 73112
henokhsiga@gmail.com*

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, a multinational aerospace and defense company, 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 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 (dominated by momentum and trend features such as EMA and Bollinger Bands) 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. All results of this technical analysis and prediction have been well-integrated into a web-based dashboard to support real-time market monitoring.

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

DAFTAR ISI

ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
1.7 Jadwal Penelitian	7
BAB II LANDASAN TEORI	9
2.1 Tinjauan Pustaka.....	9
2.2 <i>RTX Coporation</i>	11
2.3 <i>XGBoost</i>	12
2.4 Justifikasi Pemilihan <i>XGBoost</i> dibandingkan Algoritma Lain.....	13
2.5 <i>Yahoo Finance</i>	14
2.6 Analisis Teknikal dalam Prediksi Harga Saham.....	15
2.7 Analisis Data Deret Waktu (<i>Time Series Analysis</i>).....	16
2.8 Rekayasa Fitur (<i>Feature Engineering</i>)	17
2.9 Analisis Berita dan Sentimen dengan <i>VADER</i>	17
BAB III METODOLOGI PENELITIAN.....	19
3.1 Data Penelitian.....	19
3.2 Metode Penelitian	21
3.2.1 Pengumpulan Data.....	21

3.2.2	<i>PraProcessing Data</i>	21
3.2.3	Rekayasa Fitur	22
3.2.4	Analisis Berita dan Sentimen (Fitur Non-Teknikal)	23
3.2.5	<i>Hyperparameter Tuning</i>	25
3.2.6	Pelatihan Model <i>XGBoost</i>	28
3.2.7	Prediksi	29
3.2.8	Evaluasi Model	30
3.2.9	Penataan Fitur dan Persiapan Model Analisis Non-Teknikal	31
3.2.10	Implementasi <i>Dashboard</i>	31
3.2.11	<i>Feature Selection</i>	33
BAB IV HASIL DAN PEMBAHASAN		35
4.1	Deskripsi Data	35
4.1.1	Pembagian Dataset	37
4.2	Rekayasa Fitur (<i>Feature Engineering</i>)	38
4.2.1	Indikator Teknikal yang Digunakan	38
4.2.2	<i>Praprocessing Data</i>	42
4.3	Hasil <i>Training</i> Model <i>XGBoost</i>	42
4.3.1	Konfigurasi Hyperparameter	42
4.3.2	Proses Training	43
4.4	Evaluasi Performa Model	44
4.4.1	Metrik Evaluasi	44
4.4.2	Analisis <i>Error Distribution</i>	47
4.4.3	Sample Prediksi	50
4.4.4	Hasil <i>Feature Selection</i>	52
4.4.5	Perbandingan Model <i>Baseline</i> dan Model <i>Feature Selection</i>	53
4.5	Analisis <i>Feature Importance</i>	54
4.5.1	Kontribusi Fitur terhadap Prediksi	54
4.5.2	Implikasi <i>Feature Selection</i> terhadap Interpretasi Model	55
4.6	Eksperimen Model <i>Hybrid</i> : Integrasi Fitur Non-Teknikal	56
4.6.1	Hasil Evaluasi Kinerja Model <i>Hybrid</i>	57
4.7	Implementasi <i>Dashboard</i> Aplikasi Interaktif	59
4.7.1	Arsitektur Sistem <i>Dashboard</i>	59

4.7.2	Fitur-Fitur Utama <i>Dashboard</i>	63
4.7.3	Antarmuka Pengguna dan Interaktivitas.....	65
4.8	<i>Wireframe Dashboard</i> Aplikasi	67
4.9	<i>Screenshot Dashboard</i> Aplikasi.....	68
BAB V KESIMPULAN DAN SARAN.....		75
5.1	Kesimpulan.....	75
5.2	Saran	76

DAFTAR GAMBAR

Gambar 3.1 Cuplikan Data harga saham harian <i>RTX Corporation</i> dari <i>Yahoo Finance</i>	20
Gambar 3.2 Tahapan Metode Penelitian	21
Gambar 4.1 Grafik Pergerakan Harga Saham <i>RTX</i> (1962-2026)	37
Gambar 4.2 Visualisasi Pembagian Dataset <i>Training</i> dan <i>Testing</i>	38
Gambar 4.4 Distribusi <i>Error</i> Prediksi.....	47
Gambar 4.5 Harga Aktual vs Harga Prediksi	49
Gambar 4.6 Antarmuka Utama <i>Dashboard</i> Aplikasi Prediksi Harga Saham <i>RTX</i>	69
Gambar 4.7 <i>Sidebar Dashboard</i>	70
Gambar 4.8 Visualisasi Indikator Teknikal pada <i>Dashboard</i>	71
Gambar 4.9 Visualisasi Indikator Teknikal pada <i>Dashboard</i> (2)	71
Gambar 4.10 Evaluasi Performa Model pada <i>Testing Set</i>	72
Gambar 4.11 Evaluasi Performa Model pada <i>Testing Set</i> (2).....	73

DAFTAR TABEL

Tabel 1.1 Jadwal Penelitian Skripsi	7
Tabel 4.1 Statistik Deskriptif Data Harga Saham <i>RTX</i>	36
Tabel 4.2 Daftar 43 Fitur yang Digunakan dalam Model	38
Tabel 4.3 Konfigurasi <i>Hyperparameter XGBoost</i>	42
Tabel 4.4 Hasil Evaluasi Model pada <i>Testing Set</i>	44
Tabel 4.5 Sample Prediksi pada 5 Hari Terakhir <i>Testing Set</i>	50
Tabel 4.6 Daftar fitur yang dipilih berdasarkan urutan <i>importance</i> tertinggi.	53
Table 4.7 Perbandingan Performa Model	53
Tabel 4.8 Kategori Fitur dan Kontribusi Kolektifnya	54
Tabel 4.9 Hasil Evaluasi Model <i>Hybrid</i>	57

BAB I

PENDAHULUAN

1.1 Latar Belakang

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 (Zebua & Gea, 2025). 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. Informasi mengenai saham seringkali tidak lengkap, rumit, tidak pasti, dan tidak jelas, yang semakin mempersulit prediksi kinerja ekonomi di masa depan (Olatunji Akinrinola et al., 2024).

Prediksi harga saham adalah salah satu tugas paling menantang dalam peramalan keuangan karena sifat pasar saham yang fluktuatif (*volatile*) dan *non-linear* (Olatunji Akinrinola et al., 2024). Tantangan ini menuntut adanya metode yang akurat untuk dapat mendeteksi pola dan memprediksi pergerakan harga, terutama untuk perusahaan yang beroperasi di sektor kompleks seperti kedirgantaraan dan pertahanan, di mana banyak faktor eksternal dapat memengaruhi stabilitas harga (Zebua & Gea, 2025).

Pencarian untuk prediksi pasar saham yang akurat telah menjadi titik fokus penelitian keuangan selama beberapa dekade. Evolusi model prediksi telah beralih dari metode statistik dasar ke teknik *machine learning* yang canggih, mencerminkan kemajuan dalam kekuatan komputasi dan metodologi analisis data (Olatunji Akinrinola et al., 2024).

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 dikembangkan oleh Tianqi Chen dan Carlos Guestrin. Algoritma ini dikenal karena skalabilitas dan kecepatannya yang superior dibandingkan solusi populer lainnya, bahkan mampu menangani miliaran contoh data dalam pengaturan terdistribusi atau terbatas memori. *XGBoost* mendominasi kumpulan data terstruktur atau tabular pada masalah pemodelan prediktif klasifikasi dan regresi, dibuktikan dengan penggunaannya yang konsisten oleh pemenang kompetisi di platform ilmu data kompetitif *Kaggle*. Meskipun dirancang untuk data tabular, *XGBoost* juga dapat digunakan untuk peramalan deret waktu. Banyak penelitian telah dilakukan untuk memprediksi pasar saham menggunakan *XGBoost* di berbagai negara seperti Cina, Amerika Serikat, Iran, dan Korea Selatan. Penelitian sebelumnya menunjukkan bahwa indikator teknikal dapat meningkatkan akurasi peramalan saham ketika digunakan sebagai input. Dalam penelitian Jange (2022), menunjukkan bahwa penggunaan indikator teknikal dalam rekayasa fitur secara signifikan menurunkan nilai *Mean Absolute Percentage Error (MAPE)* pada prediksi harga saham Bank BCA menggunakan *XGBoost*, mencapai akurasi 4.01% (Jange, 2022).

Pemilihan *XGBoost* didasarkan pada keunggulan teknis yang relevan: kemampuan menangani *missing values*, *feature importance* untuk interpretasi, dan *regularization* mencegah *overfitting*. *Systematic review* oleh Sezer et al. (2020) terhadap 160 penelitian mengidentifikasi *XGBoost* sebagai algoritma dengan performa paling konsisten untuk *financial forecasting*, dengan keunggulan pada efisiensi dan interpretability. Carta et al. (2021) mengkonfirmasi *XGBoost* memberikan balance optimal antara *accuracy*, *speed*, dan *interpretability*,

sementara Jange (2022) membuktikan efektivitasnya dengan *MAPE* 4.01% untuk prediksi saham Indonesia menggunakan indikator teknikal, menjadikan *XGBoost* pilihan tepat untuk penelitian ini yang memerlukan akurasi tinggi dan transparansi model.

Raytheon Technologies Corporation (RTX), sebagai perusahaan multinasional kedirgantaraan dan pertahanan terkemuka di dunia, mewakili studi kasus analisis ini. Beroperasi dalam lingkungan global yang ditandai dengan persaingan dan konfrontasi yang meningkat, serta perkembangan teknologi yang belum pernah terjadi sebelumnya, kinerja keuangan *RTX* sangat relevan dengan isu-isu geopolitik dan inovasi teknologi. Perusahaan ini terlibat dalam penelitian dan pengembangan produk serta teknologi canggih, termasuk mesin pesawat, sistem pertahanan udara, dan drone, yang semuanya memiliki implikasi besar terhadap stabilitas pasar dan sentimen investor. Fluktuasi harga saham *RTX* tidak hanya dipengaruhi oleh faktor ekonomi umum, tetapi juga oleh kebijakan pertahanan, kontrak pemerintah, dan isu-isu etika yang terkait dengan industri senjata (Monterio, 2022).

Penelitian ini akan berfokus pada penerapan algoritma *XGBoost* untuk mendeteksi pola dan memprediksi harga saham *RTX Corporation* dalam konteks analisis pasar. Dengan memanfaatkan kapabilitas *XGBoost* dalam menganalisis data *time series* yang kompleks dan mempertimbangkan karakteristik unik perusahaan seperti yang diuraikan dalam analisis sebelumnya, diharapkan penelitian ini dapat memberikan model prediksi yang lebih akurat dan pemahaman yang lebih baik tentang faktor-faktor pendorong harga saham di pasar.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah: Seberapa akurat algoritma *XGBoost* dalam memprediksi harga saham *RTX Corporation* menggunakan indikator teknikal sebagai fitur utama dalam konteks analisis pasar?

1.3 Batasan Masalah

Agar penelitian ini tetap fokus dan terarah, maka batasan masalah ditetapkan sebagai berikut:

1. Data harga saham yang digunakan adalah data historis *RTX Corporation* yang diakses secara *real-time* dari *Yahoo Finance* API. Pengambilan data mencakup seluruh periode historis yang tersedia sejak 2 April 1962 hingga waktu eksperimen dilakukan. Karakteristik data *real-time* memungkinkan model selalu dilatih dengan informasi pasar terkini, meningkatkan relevansi prediksi.
2. Indikator teknikal yang digunakan sebagai fitur dalam model prediksi mencakup berbagai kategori analisis teknikal, meliputi: *Simple Moving Average (SMA)* untuk periode 5, 10, dan 20 hari, *Exponential Moving Average (EMA)* untuk periode 5, 10, dan 20 hari, *Moving Average Convergence/Divergence (MACD)* beserta *Signal Line* dan *Histogram*, *Relative Strength Index (RSI)*, *Bollinger Bands*, *Stochastic Oscillator*, *Average True Range (ATR)*, serta indikator *volume* dan fitur turunan harga lainnya. Total fitur yang digunakan adalah 43 fitur teknikal.
3. Pembagian dataset untuk model prediksi ini akan menggunakan rasio 85% untuk data latih (*Training Set*) dan 15% untuk data uji (*Testing Set*).
4. Metrik evaluasi kinerja model prediksi akan menggunakan *Mean Absolute Percentage Error (MAPE)* dan/atau *Root Mean Square Error (RMSE)*, bergantung pada karakteristik data dan tujuan evaluasi.
5. Analisis fundamental perusahaan seperti laporan keuangan, rasio *P/E*, *EPS*, dan proyeksi pendapatan tidak termasuk dalam lingkup penelitian ini, kecuali yang secara implisit tercermin dalam pergerakan harga historis.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Menerapkan algoritma *XGBoost* untuk mengidentifikasi pola pada data historis harga saham *RTX Corporation*.
2. Membangun dan mengevaluasi model prediksi harga saham *RTX Corporation* menggunakan algoritma *XGBoost* dan indikator teknikal, serta mengukur tingkat akurasi.
3. Mengidentifikasi indikator teknikal yang paling berpengaruh dalam memprediksi harga saham *RTX Corporation* menggunakan *XGBoost*.
4. Mengintegrasikan fitur non-teknikal (analisis sentimen berita menggunakan algoritma *VADER*) ke dalam pemodelan prediktif *XGBoost* untuk menganalisis dan memprediksi reaktivitas pergerakan harga saham *RTX* dalam jangka pendek.
5. Mengembangkan *dashboard* interaktif untuk visualisasi hasil prediksi dan analisis model secara *real-time*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan wawasan dan alat bantu yang potensial untuk pengambilan keputusan investasi yang lebih baik di saham *RTX Corporation*, dengan menawarkan metode prediksi yang akurat.
2. Menambah literatur ilmiah dalam bidang aplikasi pembelajaran mesin, khususnya *XGBoost*, untuk prediksi harga saham, serta memberikan studi kasus konkret mengenai perusahaan di sektor kedirgantaraan dan pertahanan.
3. Hasil penelitian ini dapat menjadi dasar untuk pengembangan model prediksi yang lebih kompleks, misalnya dengan mengintegrasikan analisis sentimen, faktor fundamental, atau membandingkan kinerja *XGBoost* dengan algoritma *deep learning* lainnya seperti *LSTM*, seperti yang disarankan oleh penelitian sebelumnya.

1.6 Sistematika Penulisan

Penulisan laporan skripsi ini disusun dalam lima bab, dengan menggunakan sistematika penulisan berikut.

BAB I PENDAHULUAN

Pada bab ini menguraikan latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, sistematika penulisan dan jadwal skripsi.

BAB II LANDASAN TEORI

Pada bab ini berisi mengenai teori dan materi yang diuraikan dan di jelaskan oleh para ahli, serta digunakan untuk mendukung penelitian analisis ini.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan langkah-langkah sistematis yang dilakukan dalam penelitian, mulai dari pengumpulan data hingga proses analisis, guna menghasilkan temuan yang valid dan reliabel.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian, yaitu temuan atau data yang diperoleh setelah proses pengumpulan dan analisis data. Selanjutnya, pembahasan menginterpretasikan dan menjelaskan hasil tersebut.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang merupakan jawaban ringkas atas permasalahan penelitian, serta saran untuk pengembangan lebih lanjut dan perbaikan di masa mendatang.

BAB II LANDASAN TEORI

Pada bab ini berisi mengenai teori dan materi yang diuraikan dan di jelaskan oleh para ahli, serta digunakan untuk mendukung penelitian analisis ini meliputi tinjauan pustaka, *RTX Corporation*, *XGBoost*, Analisis Teknikal, Evaluasi Kinerja.

2.1 Tinjauan Pustaka

1. *Prediksi Harga Saham Bank BCA Menggunakan XGBoost (Jange, 2022)*
Penelitian ini bertujuan untuk memprediksi harga saham Bank BCA menggunakan algoritma *XGBoost*. Data yang digunakan adalah data harga saham selama empat tahun (2017-2020), dengan indikator teknikal seperti *Simple Moving Average (SMA)*, *Exponential Moving Average (EMA)*, *Moving Average Convergence/Divergence (MACD)*, dan *Relative Strength Index (RSI)* sebagai fitur. Hasil penelitian menunjukkan bahwa indikator teknikal, khususnya *Exponential Moving Average*, sangat memengaruhi hasil prediksi. Model *XGBoost* mencapai akurasi prediksi yang baik dengan *MAPE* sebesar 4.01%, meskipun sedikit kurang baik pada Maret 2020 akibat pandemi *Covid-19*. Studi ini menyimpulkan bahwa penggunaan indikator teknikal dalam rekayasa fitur dan penyetelan *hyperparameter* secara signifikan dapat menurunkan nilai *MAPE*, dan membandingkan *XGBoost* dengan *Prophet*, menunjukkan *XGBoost* melampaui kinerja *Prophet*.
2. *Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python (Nyoman et al., 2020)*
Penelitian ini fokus pada peramalan kecelakaan lalu lintas di Bali menggunakan algoritma *XGBoost* yang diimplementasikan dengan bahasa pemrograman *Python*. Data yang digunakan adalah data tahunan dari Badan Pusat Statistik Provinsi Bali (1996-2019), mencakup jumlah kejadian, orang meninggal dunia, luka berat, dan luka ringan. Hasil peramalan diukur menggunakan *RMSE (Root Mean Square Error)*. Model *XGBoost* menunjukkan kinerja yang sangat baik pada kategori jumlah orang meninggal (*RMSE 4.92*) dan luka berat (*RMSE 4.11*). Penelitian ini menegaskan kemampuan *XGBoost* dalam peramalan data deret waktu dan

pentingnya *RMSE* sebagai metrik evaluasi, di mana nilai *RMSE* yang rendah menunjukkan model mampu mengikuti pola data dan memprediksi pola selanjutnya.

3. *XGBoost To Enhance Learner Performance Prediction* (Hakkal & Lahcen, 2024) Jurnal ini mengeksplorasi penggunaan *XGBoost* untuk meningkatkan prediksi kinerja pelajar dalam *Intelligent Tutoring Systems (ITS)*. Meskipun berbeda domain, penelitian ini menguatkan kapabilitas *XGBoost* sebagai algoritma *boosting* berbasis pohon yang skalabel, adaptif, dan berkinerja tinggi, mampu memberikan hasil *state-of-the-art* di berbagai aplikasi ilmu data. Jurnal ini menyoroti bagaimana *XGBoost* berhasil meningkatkan kinerja model berbasis *logistic regression* (seperti *PFA* dan *DAS3H*) pada beberapa *dataset* riil, menunjukkan peningkatan signifikan dalam metrik *AUC*. Penelitian ini juga membahas pentingnya penyetelan *hyperparameter* seperti *learning rate (eta)*, *max_depth*, *gamma*, *num_boost_round*, dan *reg_lambda* untuk optimasi model dan pencegahan *overfitting*, meskipun mencatat adanya *trade-off* dengan waktu komputasi, terutama pada *dataset* besar.
4. *Raytheon Technologies Corporation: An Analysis of Leadership, Ethics, and Social Responsibility* (Monterio, 2022) Jurnal ini menganalisis *Raytheon Technologies Corporation (RTX)* dalam konteks struktur kepemimpinan, isu dan praktik etika, hubungan pemangku kepentingan, dan tanggung jawab sosial perusahaan. *RTX*, sebagai perusahaan kedirgantaraan dan pertahanan multinasional terkemuka, beroperasi dalam lingkungan global yang dinamis dan berteknologi maju. Jurnal ini menguraikan bagaimana *RTX* menghadapi kompleksitas dalam operasinya, termasuk dilema etika terkait penjualan senjata dan kepatuhan terhadap kontrol ekspor A.S., yang mana pelanggaran di masa lalu telah menyebabkan denda. Hal ini menunjukkan bahwa kinerja dan reputasi *RTX* tidak hanya bergantung pada faktor ekonomi, tetapi juga pada isu-isu geopolitik dan pertimbangan etika. Meskipun tidak berfokus pada prediksi harga saham, jurnal ini memberikan pemahaman mendalam

tentang ekosistem operasional *RTX* yang kompleks, yang secara tidak langsung dapat memengaruhi persepsi pasar dan harga sahamnya.

5. *Pengaruh Faktor Eksternal terhadap Harga Saham: Analisis dengan Data Historis dari Yahoo Finance* (Zebua & Gea, 2025) Jurnal ini menyelidiki dampak faktor eksternal terhadap harga saham dengan menganalisis data historis dari *Yahoo Finance*. Tujuan studi ini adalah untuk meneliti secara komprehensif bagaimana peristiwa global, berita internasional, dan indikator makroekonomi memengaruhi dinamika pasar saham. Studi ini menggunakan pendekatan metode campuran, menggabungkan analisis deret waktu kuantitatif dan metodologi studi peristiwa kualitatif, untuk menganalisis pergerakan harga saham di berbagai sektor selama peristiwa global signifikan antara 2018-2023. Penelitian ini menggunakan teknik statistik canggih, seperti regresi berganda, metodologi studi peristiwa, dan analisis korelasi, untuk mengidentifikasi dan mengukur hubungan antara faktor eksternal dan variasi harga saham.

2.2 *RTX Coporation*

Raytheon Technologies Corporation (RTC), yang juga dikenal sebagai *RTX*, adalah perusahaan multinasional terkemuka dalam industri kedirgantaraan dan pertahanan yang bermarkas di Waltham, Massachusetts, Amerika Serikat. Perusahaan ini dibentuk pada April 2020 melalui penggabungan historis antara *Raytheon Company* dan *United Technologies Corporation*. (Monterio, 2022)

Penggabungan ini menghasilkan sebuah entitas yang terdiri dari empat segmen bisnis utama: *Collins Aerospace*, *Pratt & Whitney*, *Raytheon Intelligence & Space*, dan *Raytheon Missiles & Defense*. Melalui segmen-segmen ini, *RTX* memproduksi berbagai teknologi canggih, termasuk mesin pesawat, avionik, sistem pertahanan udara, rudal berpemandu, satelit, dan *drone*, yang melayani pelanggan komersial, militer, dan pemerintah (Monterio, 2022)

Sebagai salah satu produsen senjata internasional terkemuka, kinerja dan reputasi *RTX* tidak hanya bergantung pada faktor ekonomi, tetapi juga sangat dipengaruhi oleh isu-isu geopolitik dan pertimbangan etika (Monterio, 2022)

2.3 *XGBoost*

XGBoost (*Extreme Gradient Boosting*) adalah algoritma pembelajaran *ensemble* yang telah dikenal luas dan efisien, banyak diterapkan karena keunggulannya yang beragam. Algoritma ini merupakan pengembangan dari *Gradient Boosting Decision Tree* (*GBDT*) dan dapat dipandang sebagai model aditif yang terdiri dari pohon *CART* (*Classification and Regression Tree*). (Zhang et al., 2022)

XGBoost adalah salah satu implementasi mesin penguat gradien yang dikenal sebagai salah satu algoritma berkinerja terbaik yang digunakan untuk pembelajaran terawasi. Algoritma ini dapat digunakan untuk masalah regresi maupun klasifikasi. *XGBoost* disukai oleh para ilmuwan data karena kecepatan eksekusinya yang tinggi di luar komputasi inti (Ibrahim Ahmed Osman et al., 2021)

Keunggulan *XGBoost* yang menjadikannya pilihan kuat dalam berbagai aplikasi yaitu:

1. Fleksibilitas Tinggi: *XGBoost* sangat adaptif terhadap berbagai jenis masalah dan *dataset*.
2. Prediktabilitas Kuat: Model yang dibangun dengan *XGBoost* cenderung memiliki kemampuan prediksi yang tinggi.
3. Kemampuan Generalisasi Kuat: Algoritma ini mampu belajar dari data pelatihan dan menggeneralisasi dengan baik pada data baru yang belum pernah dilihat sebelumnya.
4. Skalabilitas Tinggi: *XGBoost* dirancang untuk menangani *dataset* besar secara efisien, menjadikannya pilihan yang baik untuk era "data masif".
5. Efisiensi Pelatihan Model: Proses pelatihan model pada *XGBoost* dikenal cepat dan efisien.

6. Ketahanan (*Robustness*): Algoritma ini memiliki ketahanan yang baik terhadap *noise* dan anomali dalam data.

Proses pemodelan *XGBoost* melibatkan penambahan pohon secara iteratif untuk memperbaiki prediksi dari model sebelumnya. Fungsi objektif *XGBoost* dirancang untuk meminimalkan *loss* prediksi sekaligus mengontrol kompleksitas model untuk mencegah *overfitting*. Fungsi objektif *XGBoost* didefinisikan sebagai:

$$obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

di mana $obj = \sum_{i=1}^n l(y_i, \hat{y}_i)$ adalah fungsi *loss*, yang mengukur seberapa baik model memprediksi nilai aktual y_i dan $\sum_{k=1}^K \Omega(f_k)$ adalah istilah regularisasi yang digunakan untuk mengontrol kompleksitas model f_k dan menghindari *overfitting*. Kompleksitas pohon didefinisikan dengan mempertimbangkan jumlah *leaf node* (T) dan bobot setiap *node* (w). (Zhang et al., 2022)

Selain itu, *XGBoost* juga memiliki berbagai *hyperparameter* yang dapat disesuaikan untuk mengoptimalkan kinerja model. *Hyperparameter tuning* dilakukan dengan mencoba berbagai nilai yang berbeda pada parameter sebuah algoritma. Parameter yang bisa dikatakan optimal adalah parameter yang menghasilkan kinerja model terbaik (Yulianti et al., 2022). Parameter yang optimal dalam kasus regresi adalah parameter yang memberikan hasil evaluasi seperti *Root Mean Squared Error (RMSE)* yang terendah (Bijaksana Putra Negara, 2022).

2.4 Justifikasi Pemilihan *XGBoost* dibandingkan Algoritma Lain

Pemilihan algoritma *XGBoost* dalam penelitian ini didasarkan pada temuan-temuan terbaru pada literatur, yang menunjukkan bahwa model *boosting* berbasis pohon tetap kompetitif untuk data keuangan bertipe tabular dan *time series* berbasis rekayasa fitur. Pada studi komparatif prediksi saham, *XGBoost*

dilaporkan mampu memberikan kinerja yang baik ketika dibandingkan dengan pendekatan linear maupun model *machine learning* lain, terutama ketika data diperkaya oleh fitur teknikal dan dilakukan penyetelan *hyperparameter* yang memadai (Yusuf Andrika & Rahardi, 2025)

Selain itu, penelitian pada periode terbaru juga menunjukkan bahwa *XGBoost* banyak digunakan sebagai komponen inti maupun komponen *hybrid* untuk meningkatkan akurasi prediksi deret waktu finansial, termasuk pada skenario perbandingan dengan ARIMA dan LSTM (Wu, 2023)

Meskipun model *deep learning* seperti LSTM dan arsitektur *Transformer* semakin banyak digunakan, sejumlah studi terkini tetap menempatkan *XGBoost* sebagai *baseline* kuat atau bagian dari model hibrida pada *forecasting* finansial (Mohammed & El-saieed, 2025; Raudys & Goldstein, 2022). Hal ini mengindikasikan bahwa pada konteks data historis saham dengan struktur fitur teknikal yang jelas, *XGBoost* masih merupakan pilihan metodologis yang rasional dan kuat untuk penelitian prediksi harga saham.

2.5 Yahoo Finance

Yahoo Finance adalah sebuah *platform* daring terkemuka yang menyediakan data historis harga saham. *Platform* ini meliputi berbagai pasar saham global, berita internasional, dan indikator makroekonomi secara komprehensif. Karena kredibilitas dan ketersediaan datanya, *Yahoo Finance* sering digunakan sebagai sumber data utama dalam penelitian akademis dan analisis keuangan (Zebua & Gea, 2025)

Data yang disediakan oleh *Yahoo Finance*, termasuk harga saham harian, sangat vital untuk penelitian yang bertujuan menganalisis pergerakan harga saham dan pengaruh faktor eksternal terhadapnya. *Platform* ini memungkinkan para peneliti untuk menyelidiki secara komprehensif bagaimana peristiwa global, berita internasional, dan indikator makroekonomi memengaruhi dinamika pasar saham. Dalam konteks penelitian ini, data historis dari *Yahoo*

Finance menjadi dasar untuk menganalisis pergerakan harga saham *RTX Corporation* dan mendeteksi pola yang dipengaruhi oleh berbagai faktor eksternal. (Zebua & Gea, 2025)

2.6 Analisis Teknikal dalam Prediksi Harga Saham

Analisis teknikal saham memiliki peran krusial dalam membantu para investor dan pelaku pasar mengambil keputusan yang lebih baik di pasar saham yang dinamis. Dalam lingkungan yang penuh dengan fluktuasi harga dan ketidakpastian, analisis teknikal memainkan peran penting dalam mengidentifikasi pola-pola pergerakan harga saham, mengantisipasi tren masa depan, dan mengidentifikasi momen optimal untuk melakukan transaksi jual beli. Analisis teknikal berfokus pada pengamatan data historis harga saham dan *volume* perdagangan (Pipin et al., 2023).

Dalam penelitian ini, beberapa indikator teknikal umum yang akan digunakan sebagai fitur dalam model prediksi harga saham *RTX Corporation* adalah:

1. ***Simple Moving Average (SMA)***: *SMA* dihitung sebagai rata-rata sederhana dari harga penutupan aset selama periode waktu tertentu. Ini digunakan untuk menghaluskan fluktuasi harga dan mengidentifikasi tren. (Jange, 2022)
2. ***Exponential Moving Average (EMA)***: *EMA* adalah rata-rata bergerak yang memberikan bobot lebih besar pada harga terbaru, sehingga lebih responsif terhadap perubahan harga dibandingkan *SMA*. Penelitian Jange (2022) menunjukkan bahwa indikator teknikal *Exponential Moving Average* sangat memengaruhi hasil prediksi harga saham. (Jange, 2022)
3. ***Moving Average Convergence/Divergence (MACD)***: *MACD* adalah indikator *momentum* yang mengikuti tren dan menunjukkan hubungan antara dua *moving average* harga aset. Umumnya dihitung dengan mengurangi *EMA* periode panjang dari *EMA* periode pendek, dan seringkali dilengkapi dengan garis sinyal. (Jange, 2022)

4. ***Relative Strength Index (RSI)***: *RSI* adalah osilator *momentum* yang mengukur kecepatan dan perubahan pergerakan harga. Ini digunakan untuk mengidentifikasi kondisi *overbought* atau *oversold* dalam suatu aset, dengan skala nilai antara 0 dan 100. (Jange, 2022)

Penggunaan indikator-indikator teknikal ini diharapkan dapat mengekstraksi pola-pola relevan dari data historis harga saham, yang kemudian akan dimanfaatkan oleh algoritma *XGBoost* untuk meningkatkan akurasi prediksinya.

2.7 Analisis Data Deret Waktu (*Time Series Analysis*)

Data deret waktu (*time series*) adalah serangkaian titik data yang diamati, diukur, dan direkam secara berurutan dalam interval waktu yang teratur. Karakteristik utama dari data *time series* adalah adanya ketergantungan temporal (*temporal dependence*), di mana nilai pada suatu waktu dipengaruhi oleh nilai-nilai pada waktu sebelumnya (historis).

Dalam konteks pasar keuangan, harga saham merupakan contoh klasik dari data *time series* non-stasioner yang memiliki tingkat volatilitas tinggi dan sifat non-linear. Analisis data deret waktu finansial bertujuan untuk mengekstraksi statistik yang bermakna dan pola-pola karakteristik dari data historis untuk memprediksi nilai di masa depan.

Dalam pemodelan *machine learning* seperti *XGBoost*, data *time series* tidak dapat diproses secara langsung dengan asumsi bahwa setiap baris observasi independen (*Independent and Identically Distributed*). Oleh karena itu, pendekatan khusus seperti rekayasa fitur (*feature engineering*) diperlukan untuk mengekstraksi elemen temporal (seperti *lag features* atau *momentum*) agar algoritma yang berbasis tabular dapat memahami memori temporal dari urutan data historis tersebut. (Sezer et al., 2020)

2.8 Rekayasa Fitur (*Feature Engineering*)

Feature Engineering atau rekayasa fitur adalah proses penerapan pengetahuan domain (*domain knowledge*) terhadap data mentah untuk menciptakan fitur-fitur baru (*variabel prediktor*) yang membuat algoritma machine learning dapat bekerja lebih efektif. Kualitas fitur yang dihasilkan sering kali memiliki dampak yang lebih besar terhadap akurasi model prediktif dibandingkan dengan kompleksitas algoritma itu sendiri.

Dalam analisis data finansial, data mentah seperti harga Pembukaan (*Open*), Tertinggi (*High*), Terendah (*Low*), Penutupan (*Close*), dan *Volume* (sering disingkat *OHLCV*) memiliki batas representasi informasi jika diumpankan secara mentah ke dalam model. Rekayasa fitur berfungsi untuk mentransformasi metrik dasar ini menjadi representasi matematis yang menangkap dinamika pasar yang lebih kompleks.

Algoritma berbasis *decision tree* seperti *XGBoost* sangat diuntungkan oleh proses rekayasa fitur yang terstruktur. Hal ini dikarenakan *XGBoost* beroperasi dengan membuat partisi/cabang berdasarkan nilai-nilai fitur yang disediakan. Melalui rekayasa fitur yang tepat, representasi interaksi antar-variabel menjadi lebih eksplisit, sehingga model dapat mengisolasi *noise* dan menemukan *hyperplane* optimal yang menghasilkan prediksi dengan tingkat *error* (seperti *MAPE* atau *RMSE*) yang diminimalkan. (Carta et al., 2021)

2.9 Analisis Berita dan Sentimen dengan *VADER*

Analisis berita dan sentimen dalam penelitian ini diposisikan sebagai komponen pendukung *dashboard* untuk memberikan konteks non-teknikal terhadap kondisi pasar. Penelitian ini juga menyiapkan eksperimen tambahan berupa model *hybrid* yang menggabungkan fitur teknikal dan fitur sentimen berita. Pada skenario *hybrid* ini, data sentimen tidak dipakai sebagai pengganti indikator teknikal, tetapi sebagai lapisan informasi tambahan yang diolah secara terstruktur sebelum prediksi dilakukan. (Fazlija & Harder, 2022).

Sumber data berita diperoleh dari layanan *NewsAPI* yang menampilkan artikel terkait *RTX Corporation* secara *near real-time*. Algoritma sentimen yang digunakan adalah *VADER (Valence Aware Dictionary and sEntiment Reasoner)*, yaitu pendekatan lexicon dan rule-based yang efektif untuk teks berita pendek hingga menengah.

Skor sentimen dihitung menggunakan nilai compound pada rentang -1 sampai +1, kemudian diklasifikasikan ke dalam kategori positif, netral, atau negatif berdasarkan nilai ambang tertentu. Peran utama modul ini adalah menyediakan informasi interpretatif tambahan bagi pengguna *dashboard* agar hasil prediksi harga dapat dibaca bersama dinamika sentimen berita pasar (Arifin & Febryanto, 2025).

Penelitian ini tetap konsisten pada ruang lingkup utama, yaitu prediksi harga saham berbasis fitur teknikal menggunakan *XGBoost*, sedangkan analisis sentimen berfungsi sebagai lapisan analisis pendamping.

BAB III

METODOLOGI PENELITIAN

Pada bab ini, dijelaskan tahapan-tahapan sistematis dalam pelaksanaan penelitian. Metodologi penelitian yang digunakan mencakup beberapa fase utama, mulai dari pengumpulan data, pra-pemrosesan data, rekayasa fitur, perancangan model *XGBoost*, pelatihan, prediksi, hingga evaluasi model.

3.1 Data Penelitian

Dataset yang digunakan dalam penelitian ini adalah data historis harga saham harian *RTX Corporation* (ticker symbol: *RTX*) yang diakses secara langsung dari *Yahoo Finance* melalui *Application Programming Interface* (API) menggunakan library *yfinance*.

Yahoo Finance dipilih sebagai sumber data karena kredibilitasnya sebagai platform finansial terkemuka yang menyediakan data pasar saham global secara gratis dan akurat. Platform ini telah banyak digunakan dalam penelitian akademis karena ketersediaan data historis yang lengkap dan terupdate secara *real-time*.

Proses pengambilan data dilakukan dengan spesifikasi sebagai berikut:

1. Sumber Data: *Yahoo Finance* API (*finance.yahoo.com*)
2. Ticker Symbol: *RTX* (*RTX Corporation*)
3. Periode Pengambilan: 2 April 1962 – tanggal eksperimen berlangsung.
4. Interval Data: Harian (*1 day*).
5. Total Records: banyaknya data perdagangan harian.

Atribut data yang diperoleh dari API meliputi:

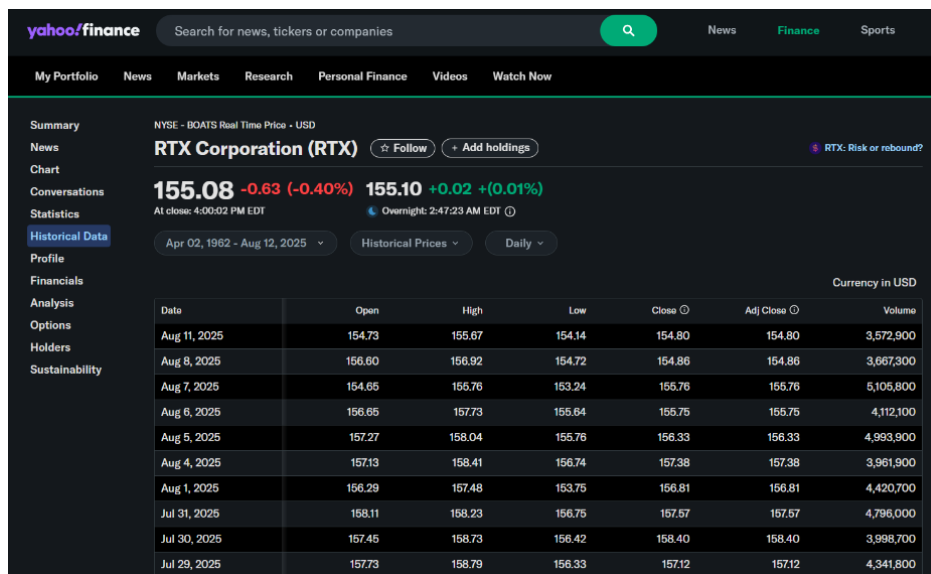
1. Date: Tanggal transaksi perdagangan.
2. Open: Harga pembukaan (*opening price*).
3. High: Harga tertinggi dalam sehari (*daily high*).
4. Low: Harga terendah dalam sehari (*daily low*).
5. Close: Harga penutupan yang telah disesuaikan untuk *stock splits* dan dividen (*adjusted closing price*).

6. *Volume*: *Volume* perdagangan harian (jumlah saham yang diperdagangkan).

Penggunaan *API* untuk pengambilan data memiliki beberapa keunggulan:

1. Data selalu *up-to-date* sesuai dengan waktu pengambilan.
2. Reproducibility: Peneliti lain dapat mengakses data yang sama untuk periode yang identik.
3. Konsistensi: Tidak ada perbedaan format atau inkonsistensi data.
4. Efisiensi: Tidak memerlukan proses manual *download* dan *update* data.

Data yang diperoleh kemudian langsung diproses dalam memori (*in-memory processing*) tanpa perlu penyimpanan *intermediary* dalam bentuk file, meningkatkan efisiensi *computational workflow*.



Summary
News
Chart
Conversations
Statistics
Historical Data
Profile
Financials
Analysis
Options
Holders
Sustainability

NYSE - BOATS Real Time Price - USD
RTX Corporation (RTX) ☆ Follow + Add holdings RTX: Risk or rebound?

155.08 -0.63 (-0.40%) **155.10** +0.02 (+0.01%)
At close: 4:00:02 PM EDT Overnight: 2:47:23 AM EDT

Apr 02, 1962 - Aug 12, 2025 Historical Prices Daily

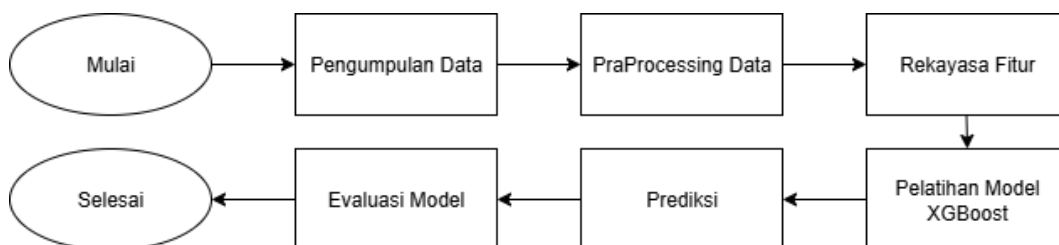
Currency in USD

Date	Open	High	Low	Close	Adj Close	Volume
Aug 11, 2025	154.73	155.67	154.14	154.80	154.80	3,572,900
Aug 8, 2025	156.60	156.92	154.72	154.86	154.86	3,667,300
Aug 7, 2025	154.65	156.76	153.24	155.76	155.76	5,105,800
Aug 6, 2025	156.65	157.73	155.64	155.75	155.75	4,112,100
Aug 5, 2025	157.27	158.04	155.76	156.33	156.33	4,993,900
Aug 4, 2025	157.13	158.41	156.74	157.38	157.38	3,961,900
Aug 1, 2025	156.29	157.48	153.75	156.81	156.81	4,420,700
Jul 31, 2025	158.11	158.23	156.75	157.57	157.57	4,796,000
Jul 30, 2025	157.45	158.73	156.42	158.40	158.40	3,998,700
Jul 29, 2025	157.73	158.79	156.33	157.12	157.12	4,341,800

Gambar 3.1 Cuplikan Data harga saham harian *RTX Corporation* dari *Yahoo Finance*

3.2 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode pemodelan prediktif untuk menganalisis dan memodelkan data deret waktu. Tujuannya adalah untuk mendeteksi pola dan memprediksi pergerakan harga saham *RTX Corporation*. Metodologi ini dirancang dalam serangkaian tahapan sistematis yang memastikan proses penelitian berjalan secara terstruktur dan menghasilkan temuan yang valid dan dapat diandalkan. Alur penelitian dapat digambarkan pada gambar 3.3 berikut:



Gambar 3.2 Tahapan Metode Penelitian

3.2.1 Pengumpulan Data

Tahap awal ini merupakan fondasi dari seluruh penelitian. Data yang dikumpulkan adalah data historis harga saham *RTX Corporation* (*RTX*) dari *platform* finansial seperti *Yahoo Finance*. Data ini bersifat kuantitatif dan memiliki format deret waktu, yang mencakup atribut penting seperti tanggal, harga pembukaan (*open*), harga tertinggi (*high*), harga terendah (*low*), harga penutupan (*close*), dan *volume* perdagangan harian. Pengumpulan data yang lengkap dan akurat menjadi kunci untuk memastikan kualitas model prediksi yang akan dibangun.

3.2.2 *PraProcessing* Data

Data yang diambil langsung dari *Yahoo Finance* API melalui *library yfinance* telah dalam format yang terstruktur dan bersih. *Library yfinance* secara otomatis menangani *adjustment* untuk *stock splits* dan

dividends, sehingga harga penutupan (*Close*) yang digunakan adalah *adjusted closing price* yang mencerminkan nilai aktual investasi.

Data mentah yang dikumpulkan seringkali memiliki ketidaksempurnaan, seperti nilai yang hilang (*missing values*) atau format yang tidak konsisten. Oleh karena itu, tahap pra-pemrosesan data sangat penting untuk mempersiapkan data agar siap untuk pemodelan. Dalam tahap ini, data akan dibersihkan, dinormalisasi, dan diorganisasi. Setelah data bersih, dataset akan dibagi menjadi dua bagian utama dengan menetapkan pembagian rasio dataset sebesar 85% untuk data latih (*Training Set*) dan 15% untuk data uji (*Testing Set*). Rasio 85:15 dipilih untuk memberikan data historis yang lebih banyak bagi model dalam mempelajari pola jangka panjang, sementara *testing set* 15% masih cukup substansial untuk validasi performa model pada berbagai kondisi pasar.

1. **Data Latih (*Training Set*):** Digunakan untuk melatih model *XGBoost* agar dapat mempelajari pola dan hubungan yang ada dalam data historis.
2. **Data Uji (*Testing Set*):** Digunakan untuk menguji kinerja model yang telah dilatih pada data yang belum pernah dilihat sebelumnya. Pembagian ini penting untuk mengukur akurasi model dan memastikan bahwa model tidak mengalami *overfitting*.

3.2.3 Rekayasa Fitur

Rekayasa fitur adalah proses mengubah data mentah menjadi fitur yang lebih informatif untuk meningkatkan kinerja model. Dalam konteks prediksi harga saham, fitur yang paling relevan adalah indikator teknikal. Tahap ini melibatkan perhitungan dan penambahan 43 fitur teknikal ke dalam dataset, yang terdiri dari: *Moving Averages* (*SMA* dan *EMA* untuk berbagai periode), indikator *momentum*

(*MACD*, *RSI*, *Stochastic Oscillator*), indikator volatilitas (*Bollinger Bands*, *ATR*), indikator *volume*, fitur dinamika harga, *lag features*, dan fitur temporal. Indikator-indikator ini berfungsi sebagai *variabel prediktor* yang membantu model *XGBoost* memahami tren, *momentum*, dan volatilitas pasar yang ada dalam data. Penelitian sebelumnya menunjukkan bahwa penggunaan indikator teknikal ini dapat secara signifikan meningkatkan akurasi model dalam memprediksi harga saham.

Selain 43 fitur teknikal yang digunakan pada model utama, penelitian ini juga menyiapkan eksperimen tambahan berupa model *hybrid* yang menggabungkan fitur teknikal dan fitur sentimen berita. Pada skenario *hybrid* ini, data sentimen tidak dipakai sebagai pengganti indikator teknikal, tetapi sebagai lapisan informasi tambahan yang diolah secara terstruktur sebelum prediksi dilakukan.

Pendekatan ini dipilih agar model tetap mempertahankan kekuatan utama dari fitur teknikal, sekaligus memperoleh konteks dari berita terbaru yang berpotensi memengaruhi arah harga saham dalam jangka pendek.

3.2.4 Analisis Berita dan Sentimen (Fitur Non-Teknikal)

Analisis berita dan sentimen diterapkan sebagai komponen monitoring independen untuk menyajikan konteks pasar non-teknikal secara *real-time*. Pada penelitian ini, hasil sentimen juga dapat digunakan sebagai variabel pelatihan model prediksi harga.

Proses ekstraksi sentimen dilakukan melalui tahapan berikut:

1. **Pengumpulan Berita:** Berita terkait *RTX Corporation* ditarik menggunakan *NewsAPI* secara *near real-time*. Karena limitasi akses *API* untuk akun pengembang yang memblokir penarikan berita lebih dari 30 hari ke belakang,

rentang data latih pada model *hybrid* disesuaikan secara proporsional.

2. **Sentiment Scoring dengan VADER:** Teks berita yang telah melalui tahap *praprocessing* dianalisis menggunakan algoritma *VADER (Valence Aware Dictionary and sEntiment Reasoner)*. Algoritma berbasis leksikon ini dipilih karena kemampuannya menangani teks pendek dan konteks finansial. *VADER* menghasilkan *compound score* pada rentang -1 (sangat negatif) hingga +1 (sangat positif).
3. **Agregasi Fitur Harian:** Skor sentimen individual diagregasi menjadi fitur matriks harian, menghasilkan sembilan variabel non-teknikal baru: *News_Count*, *News_Positive*, *News_Negative*, *News_Neutral*, *News_Sentiment_Sum*, *News_Sentiment_Avg*, *News_Bullish_Ratio*, *News_Bearish_Ratio*, dan *News_Has_News*.
4. **Penyatuan Matriks (Dataset Merger):** Kesembilan variabel sentimen tersebut digabungkan (*merge*) dengan dataset teknikal berdasarkan kesamaan tanggal publikasi berita dan hari bursa aktif.

Hasil analisis sentimen kemudian diagregasi menjadi fitur harian atau fitur ringkas untuk keperluan prediksi, seperti jumlah berita, nilai sentimen rata-rata, jumlah berita positif, jumlah berita negatif, rasio bullish, rasio bearish, dan indikator apakah berita tersedia pada periode tersebut.

Jika berita tidak dapat diambil pada saat prediksi, maka nilai sentimen dianggap netral sehingga sistem tetap dapat berjalan tanpa memutus proses prediksi.

3.2.5 Hyperparameter Tuning

Hyperparameter tuning merupakan tahap krusial dalam pengembangan model *machine learning* untuk mencapai performa optimal. Berbeda dengan parameter model yang dipelajari selama proses *training*, *hyperparameter* adalah konfigurasi yang harus ditetapkan sebelum proses pelatihan dimulai dan secara signifikan mempengaruhi kemampuan model dalam mempelajari pola data serta menggeneralisasi pada data baru (Carta et al., 2021).

Dalam penelitian ini, *hyperparameter tuning* dilakukan untuk mengoptimalkan performa model *XGBoost* melalui pendekatan sistematis sebagai berikut:

1. **Identifikasi *Hyperparameter* Kritis:** Berdasarkan karakteristik algoritma *XGBoost*, parameter yang paling berpengaruh terhadap performa model diidentifikasi:
 - a. *n_estimators*: Jumlah pohon keputusan yang akan dibangun. Nilai terlalu rendah menyebabkan *underfitting*, sedangkan terlalu tinggi meningkatkan risiko *overfitting* dan *computational cost*.
 - b. *max_depth*: Kedalaman maksimum setiap pohon. Parameter ini mengontrol kompleksitas model, di mana nilai lebih besar memungkinkan model menangkap pola lebih kompleks namun meningkatkan risiko *overfitting*.
 - c. *learning_rate (eta)*: Tingkat pembelajaran yang mengontrol kontribusi setiap pohon terhadap prediksi final. Nilai lebih kecil membuat proses *learning* lebih *robust* namun memerlukan lebih banyak pohon.
 - d. *min_child_weight*: Bobot minimum yang diperlukan untuk membuat partisi baru. Parameter ini berfungsi

sebagai *regularization* untuk mencegah model terlalu sensitif terhadap *noise*.

- e. *subsample*: Proporsi *data training* untuk membangun setiap pohon. *Subsampling* membantu mencegah *overfitting* dengan memperkenalkan *randomness*.
- f. *colsample_bytree*: Proporsi fitur yang dipilih secara random untuk membangun setiap pohon, memberikan efek *regularization* tambahan.
- g. *gamma*: *Minimum loss reduction* yang diperlukan untuk membuat *split* pada *leaf node*. Nilai lebih besar membuat model lebih konservatif.
- h. *reg_alpha* dan *reg_lambda*: L1 dan L2 *regularization terms* yang mengontrol kompleksitas model untuk mencegah *overfitting*.

2. Metode **Tuning**: *Grid Search* dengan *Cross-Validation*

Penelitian ini menggunakan *Grid Search* dengan *5-fold cross-validation* untuk menemukan kombinasi *hyperparameter* optimal. *Grid Search* adalah metode *exhaustive search* yang mencoba semua kombinasi nilai parameter yang telah didefinisikan dalam *search space*.

Search space yang digunakan:

- a. *n_estimators*: [100, 200, 300, 500]
- b. *max_depth*: [3, 4, 5, 6, 7]
- c. *learning_rate*: [0.01, 0.05, 0.08, 0.1]
- d. *min_child_weight*: [5, 10, 15, 20]
- e. *subsample*: [0.6, 0.8, 1.0]
- f. *colsample_bytree*: [0.6, 0.8, 1.0]
- g. *gamma*: [0, 0.1, 0.3, 0.5]
- h. *reg_alpha*: [0, 0.5, 1.0]
- i. *reg_lambda*: [0.5, 1.0, 2.0]

3. **Cross-Validation Strategy:** Mengingat data yang digunakan adalah *time series*, *cross-validation* dilakukan dengan mempertahankan urutan temporal data menggunakan *Time Series Split*. Metode ini membagi data *training* menjadi 5 *folds* secara berurutan, di mana setiap *fold training* selalu menggunakan data masa lalu untuk memprediksi data masa depan, mencegah data *leakage* yang dapat terjadi jika menggunakan *random split*.
4. **Metrik Evaluasi untuk Tuning:** Setiap kombinasi *hyperparameter* dievaluasi menggunakan *Mean Absolute Percentage Error (MAPE)* pada validation set. *MAPE* dipilih sebagai metrik utama karena interpretasinya yang intuitif dalam bentuk persentase *error* dan kemampuannya memberikan evaluasi yang konsisten pada berbagai level harga (Jange, 2022).
5. **Pemilihan Konfigurasi Optimal:** Kombinasi *hyperparameter* dengan *MAPE* terendah pada *cross-validation* dipilih sebagai konfigurasi *final model*. Selain *MAPE*, metrik tambahan seperti *RMSE*, R^2 , dan *directional accuracy* juga dicatat untuk memberikan gambaran komprehensif tentang performa model.
6. **Validasi Final:** Setelah *hyperparameter* optimal ditemukan, model dengan konfigurasi tersebut dilatih ulang menggunakan seluruh data *training* dan dievaluasi pada *testing set* yang terpisah untuk memastikan model dapat menggeneralisasi dengan baik pada data yang belum pernah dilihat.

Proses *hyperparameter tuning* ini mengikuti *best practices* dalam *machine learning* untuk *financial forecasting*, memastikan model *XGBoost* mencapai keseimbangan optimal antara *bias* dan *variance* (Carta et al., 2021).

3.2.6 Pelatihan Model *XGBoost*

Setelah *hyperparameter* optimal ditentukan melalui proses tuning pada tahap sebelumnya, model *XGBoost* dilatih menggunakan seluruh data *training* yang telah diproses dan di-augmentasi dengan fitur teknikal. Proses pelatihan model melibatkan tahapan berikut:

1. **Inisialisasi Model:** Model *XGBoost* diinisialisasi dengan *hyperparameter* optimal yang telah ditentukan melalui *Grid Search*. Konfigurasi ini mencakup semua parameter yang telah di-tune, termasuk jumlah pohon (*n_estimators*), kedalaman pohon (*max_depth*), *learning rate*, dan *parameter regularization*.
2. **Feature Scaling:** Sebelum *training*, seluruh fitur (teknikal) dinormalisasi menggunakan *RobustScaler*. *RobustScaler* dipilih karena ketahanannya terhadap *outliers*, yang umum terjadi pada data finansial yang memiliki volatilitas tinggi. *Scaler* di-fit hanya pada data *training* untuk mencegah data *leakage*, kemudian transformasi yang sama diterapkan pada data *testing* (Carta et al., 2021).
3. **Pelatihan Berbasis Gradient Boosting:** *XGBoost* membangun serangkaian pohon keputusan secara sekuensial melalui *gradient boosting framework*. Setiap pohon yang baru ditambahkan bertujuan untuk memperbaiki residual *error* yang dibuat oleh *ensemble* pohon sebelumnya. Proses iteratif ini terus berlanjut hingga mencapai jumlah pohon yang ditentukan (*n_estimators*) atau hingga *improvement* dalam *loss function* menjadi minimal.
4. **Early Stopping Mechanism:** Untuk mencegah *overfitting* dan meningkatkan efisiensi *computational*, *early stopping mechanism* diterapkan. Jika performa model pada validasi

set tidak mengalami *improvement* selama sejumlah iterasi tertentu, proses *training* dihentikan secara otomatis meskipun belum mencapai jumlah maksimal pohon. Mekanisme ini memastikan model tidak *overfit* pada data *training*.

5. **Regularization:** *XGBoost* menerapkan L1 (*reg_alpha*) dan L2 (*reg_lambda*) *regularization* pada *leaf weights* untuk mengontrol kompleksitas model. Parameter *gamma* juga digunakan untuk mengontrol minimum *loss reduction* yang diperlukan untuk membuat *split* baru pada *tree nodes*. Kombinasi *regularization* ini efektif mencegah *overfitting* pada data *time series* yang *prone to noise* (Carta et al., 2021).
6. **Monitoring Training Process:** Selama proses *training*, *loss function* dan *prediction error* dimonitor untuk memastikan model belajar dengan baik. *Learning curves* yang menampilkan performa pada *training set* versus *validation set* divisualisasikan untuk mendeteksi potensi *overfitting* atau *underfitting*.

Hasil dari tahap ini adalah model *XGBoost* terlatih yang telah mempelajari pola kompleks dari fitur teknikal, siap digunakan untuk melakukan prediksi pada data *testing* dan data masa depan.

3.2.7 Prediksi

Setelah model *XGBoost* selesai dilatih dan dioptimalkan, model tersebut akan digunakan untuk memprediksi harga saham pada data uji. Tahap ini merupakan aplikasi praktis dari model, di mana model akan menghasilkan serangkaian nilai yang merupakan perkiraan harga saham di masa depan berdasarkan pola yang telah dipelajari dari data historis.

3.2.8 Evaluasi Model

Evaluasi kinerja model prediksi merupakan tahapan krusial untuk menilai seberapa efektif model yang telah dibangun dalam mencapai tujuan penelitian. Metrik evaluasi menyediakan cara kuantitatif untuk mengukur perbedaan antara nilai yang diprediksi oleh model dengan nilai aktual. Untuk masalah regresi, seperti prediksi harga saham, beberapa metrik spesifik sangat relevan:

1. **Root Mean Square Error (RMSE):** *RMSE* adalah ukuran statistik yang menghitung rata-rata besarnya kesalahan prediksi. Metrik ini dihitung sebagai akar kuadrat dari *Mean Squared Error (MSE)*. *RMSE* memberikan bobot yang lebih besar pada kesalahan yang lebih besar karena adanya proses pengkuadratan, sehingga sensitif terhadap outlier dalam data. Nilai *RMSE* yang lebih rendah menunjukkan bahwa model memiliki akurasi prediksi yang lebih baik, karena prediksi model lebih dekat dengan nilai aktual.

Rumus matematis untuk *RMSE* adalah:

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

Di mana A_i adalah nilai aktual pada data ke- i , y_i adalah nilai prediksi pada data ke- i dan n adalah jumlah data. (Nyoman et al., 2020)

2. **Mean Absolute Percentage Error (MAPE):** *MAPE* adalah metrik evaluasi yang mengukur rata-rata dari persentase kesalahan absolut antara nilai aktual dan nilai ramalan. Semakin rendah nilai *MAPE*, semakin tinggi tingkat akurasi model peramalan tersebut (Vivas et al., 2020). Keunggulan

utama *MAPE* terletak pada interpretasinya yang intuitif, karena hasilnya dinyatakan dalam persentase, membuatnya mudah dipahami oleh berbagai kalangan. (Khalis Sofi et al., 2021)

Rumus matematis untuk *MAPE* adalah:

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

Di mana A_t adalah nilai aktual pada waktu t , F_t adalah nilai prediksi pada waktu t dan n adalah jumlah data.

3.2.9 Penataan Fitur dan Persiapan Model Analisis Non-Teknikal

Sebelum data dimasukkan ke model *hybrid*, baris data terakhir dari dataset teknikal disusun ulang agar kolom awal mengikuti urutan wajib, yaitu *Price*, *Close*, *High*, *Low*, *Open*, dan *Volume*. Kolom *Price* dibentuk dengan menduplikasi nilai *Close* untuk menjaga format input yang konsisten dengan pipeline implementasi.

Setelah urutan kolom disesuaikan, fitur teknikal dan fitur sentimen digabungkan ke dalam satu vektor input dan kemudian ditransformasi dengan *scaler* yang sama seperti pada model *hybrid* saat pelatihan. Dengan cara ini, proses inferensi live menggunakan struktur fitur yang identik dengan struktur fitur saat *training*.

3.2.10 Implementasi *Dashboard*

Untuk memfasilitasi visualisasi hasil prediksi dan analisis model secara interaktif, penelitian ini mengembangkan *dashboard* aplikasi menggunakan *framework Streamlit*. *Dashboard* menyediakan antarmuka *user-friendly* yang memungkinkan eksplorasi hasil prediksi tanpa memerlukan pengetahuan *programming*.

1. Visualisasi Data dan Prediksi

- a. Grafik interaktif harga saham historis dengan indikator teknikal.
- b. Perbandingan visual antara harga aktual dan hasil prediksi model.
- c. *Zoom, pan, dan hover information* untuk analisis detail.

2. Analisis Performa Model

- a. *Display* metrik evaluasi (*MAPE, RMSE, R², MAE, Directional Accuracy*).
- b. Visualisasi distribusi *error* prediksi.
- c. *Residual analysis plots*.

3. Feature Importance Visualization

- a. Ranking dan visualisasi kontribusi setiap fitur terhadap prediksi.
- b. Identifikasi indikator teknikal yang paling berpengaruh.

4. Sentiment Analysis Monitoring

- a. *Time series visualization* sentimen berita *RTX Corporation*.
- b. Korelasi antara sentimen dan pergerakan harga.
- c. News feed dengan sentiment scores.

5. Interactive Prediction

- a. *Input* tanggal spesifik untuk *generate* prediksi harga.
- b. *Multi-step ahead forecasting visualization*.

6. Hyperparameter Experimentation

- a. *Interface* untuk eksperimen dengan berbagai parameter *SMA/EMA*.
- b. *Comparison table* hasil berbagai konfigurasi.

Dashboard dikembangkan mengikuti prinsip *interactive visualization* untuk *financial data analysis*, memungkinkan *stakeholder* untuk mengeksplorasi hasil model secara mandiri.

3.2.11 Feature Selection

Pada penelitian ini dilakukan *feature selection* untuk menguji apakah subset fitur tertentu dapat mempertahankan atau meningkatkan performa model dibanding penggunaan seluruh fitur hasil *feature engineering*. *Feature selection* diterapkan setelah pembagian data *train* dan *test* agar proses pemilihan fitur hanya memanfaatkan data *training* dan tidak menimbulkan data *leakage*.

Metode *feature selection* yang digunakan adalah berbasis *importance* dari model *XGBoost*. Alasan pemilihan metode ini adalah karena *XGBoost* mampu memberikan ukuran kontribusi relatif setiap fitur terhadap prediksi model secara langsung melalui nilai *feature importance*. Dengan pendekatan ini, fitur diurutkan berdasarkan tingkat kepentingannya, kemudian dipilih sejumlah fitur teratas sesuai skenario eksperimen.

Tahapan *feature selection* penelitian ini adalah sebagai berikut :

- a) Data dibagi secara *time-series split* dengan rasio 85:15.
- b) Fitur dilatih menggunakan data *training* saja.
- c) Model *XGBoost* awal dibangun untuk memperoleh skor *importance* setiap fitur.
- d) Fitur diurutkan dari nilai *importance* tertinggi ke terendah.
- e) Sejumlah fitur teratas dipilih sebagai subset fitur final.
- f) Model kemudian dilatih ulang menggunakan subset fitur tersebut.
- g) Performa model *feature selection* dibandingkan dengan model *baseline* yang menggunakan seluruh fitur.

Dalam implementasinya, jumlah fitur terpilih ditentukan berdasarkan skenario eksperimen, misalnya *top-15* fitur untuk model utama dan *top-20* fitur untuk model *hybrid*. Penentuan jumlah ini dapat disesuaikan dengan hasil evaluasi *MAPE*, *RMSE*, *R2*, dan *directional accuracy*.

Pendekatan ini memberikan dua manfaat utama. Pertama, model menjadi lebih sederhana karena hanya menggunakan fitur yang paling informatif. Kedua, interpretabilitas meningkat karena peneliti dapat menjelaskan fitur mana yang paling berpengaruh terhadap prediksi harga saham *RTX*.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Deskripsi Data

Penelitian ini menggunakan data historis harga saham *RTX Corporation* yang diperoleh dari *Yahoo Finance API*. Data yang digunakan mencakup periode dari 2 April 1962 hingga 14 Januari 2026, dengan total 16.053 baris data setelah proses pembersihan.

Pemilihan *RTX Corporation* sebagai objek penelitian didasarkan pada beberapa pertimbangan strategis. *RTX Corporation* merupakan salah satu perusahaan terbesar di sektor *aerospace* dan *defense*, masuk dalam jajaran *Fortune 100 companies* dengan kapitalisasi pasar lebih dari \$140 miliar. Sebagai hasil merger antara *Raytheon Company* dan *United Technologies Corporation* pada tahun 2020, *RTX* memiliki karakteristik unik yang menarik untuk dianalisis: volatilitas harga yang tinggi akibat sensitivitas terhadap kebijakan pertahanan global, kontrak pemerintah, dan dinamika geopolitik. Volatilitas ini menjadikan *RTX* sebagai kandidat ideal untuk menguji kemampuan algoritma *XGBoost* dalam menangkap pola kompleks pada data *time series* finansial.

Data yang digunakan terdiri dari enam variabel utama yang diperoleh langsung dari *Yahoo Finance API*: *Open* (harga pembukaan), *High* (harga tertinggi), *Low* (harga terendah), *Close* (harga penutupan), *Volume* (volume perdagangan), dan *Adj Close* (harga penutupan yang disesuaikan dengan *stock splits* dan *dividen*). Variabel *Close* dan *Adj Close* digunakan sebagai basis perhitungan target prediksi, sementara kelima variabel lainnya menjadi input untuk *feature engineering*.

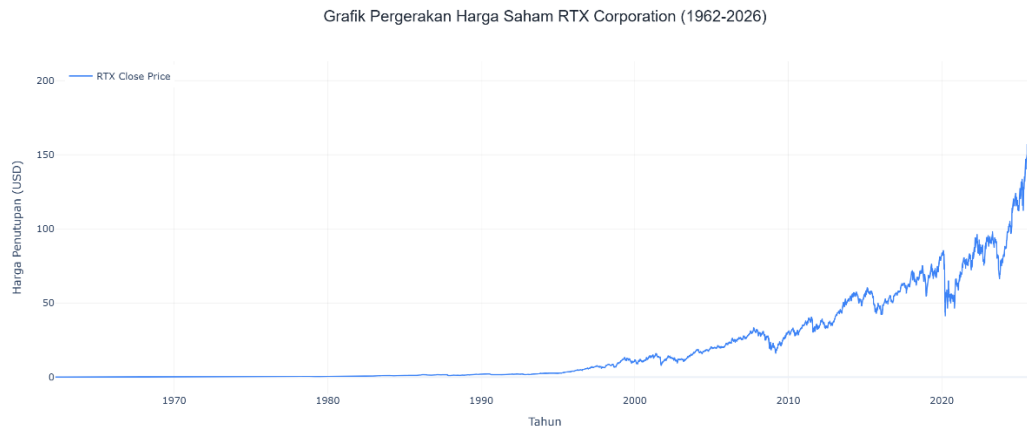
Rentang waktu data yang mencakup lebih dari 63 tahun (1962-2026) memberikan keunggulan signifikan bagi model untuk mempelajari berbagai kondisi pasar yang berbeda. Periode ini mencakup berbagai siklus ekonomi: resesi tahun 1970-an dan 1980-an, *bubble dot-com* tahun 2000, krisis finansial

global 2008, pandemi *COVID-19* tahun 2020, serta periode pemulihan dan inflasi tahun 2021-2023. Keberagaman kondisi pasar ini memungkinkan model *XGBoost* untuk belajar pola yang *robust* dan general, tidak terbatas pada kondisi pasar tertentu saja.

Tabel 4.1 Statistik Deskriptif Data Harga Saham *RTX*

Variabel	Minimum	Maksimum	Mean	Std Dev	Total Records
<i>Open</i>	\$0.07	\$200.09	\$20.74	\$31.42	16,053
<i>High</i>	\$0.08	\$202.17	\$20.93	\$31.70	16,053
<i>Low</i>	\$0.07	\$199.58	\$20.55	\$31.13	16,053
<i>Close</i>	\$0.07	\$201.92	\$20.74	\$31.42	16,053
<i>Volume</i>	0	63,223,770	4,470,847	3,707,186	16,053

Tabel 4.1 menunjukkan statistik deskriptif data harga saham *RTX Corporation* selama periode 63+ tahun. Terlihat bahwa harga saham memiliki rentang yang sangat luas dari minimum \$0.07 (tahun 1960-an, *adjusted for splits*) hingga maksimum \$201.92 (Januari 2026). Rata-rata harga penutupan adalah \$20.74 dengan standar deviasi \$31.42, mengindikasikan volatilitas yang signifikan sepanjang periode pengamatan. *Volume* perdagangan rata-rata adalah 4.47 juta saham per hari dengan nilai maksimum mencapai 63.2 juta saham, yang terjadi pada periode volatilitas tinggi seperti krisis finansial 2008 atau pandemi *COVID-19* 2020.



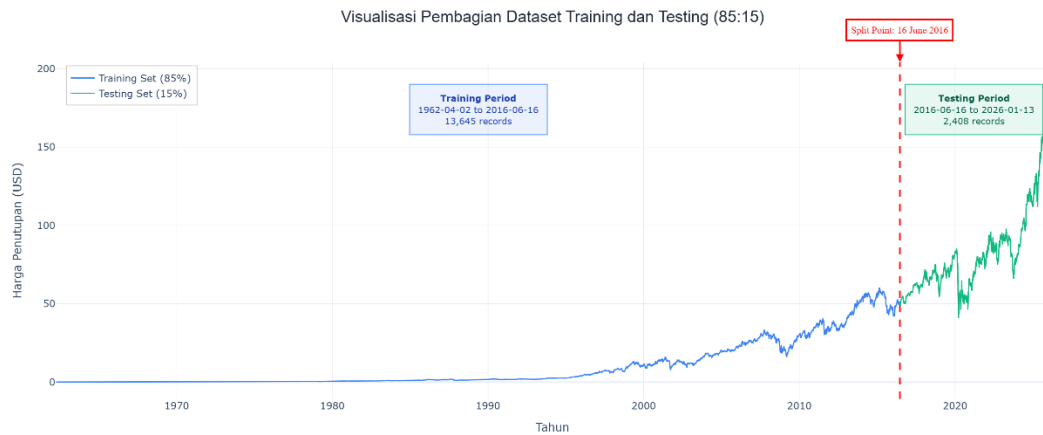
Gambar 4.1 Grafik Pergerakan Harga Saham *RTX* (1962-2026)

4.1.1 Pembagian Dataset

Dataset dibagi menjadi dua subset menggunakan pendekatan *time-series split* dengan rasio 85:15

1. *Training Set*: 13.645 records (85%)
2. *Testing Set*: 2.408 records (15%)

Pemilihan rasio 85:15 untuk pembagian *training* dan *testing set* didasarkan pada beberapa pertimbangan teknis dan praktis. Rasio ini memberikan *training set* yang lebih besar (85%) sehingga model memiliki data historis yang lebih banyak untuk mempelajari pola jangka panjang termasuk berbagai siklus ekonomi. *Testing set* 15% (setara dengan ~ 2.377 *data points* atau sekitar 7-8 tahun *trading days*) masih cukup substansial untuk validasi performa model pada berbagai kondisi pasar, mencakup periode 2016-2026 yang meliputi pandemi *COVID-19* dan pemulihan pasca-pandemi.



Gambar 4.2 Visualisasi Pembagian Dataset *Training* dan *Testing*

4.2 Rekayasa Fitur (*Feature Engineering*)

4.2.1 Indikator Teknikal yang Digunakan

Berdasarkan metodologi pada Bab III, penelitian ini mengekstrak 43 fitur dari data harga saham mentah, yang terdiri dari indikator teknikal dan fitur turunan lainnya. Fitur-fitur ini mencakup sembilan 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).

Oriani & Coelho (2016) menemukan bahwa indikator teknikal bila digunakan sebagai input dapat meningkatkan keakuratan ramalan saham dan kombinasi berbagai indikator sebagai input dapat lebih meningkatkan kinerja peramalan.(Jange, 2022).

Tabel 4.2 Daftar 43 Fitur yang Digunakan dalam Model

No	Nama Fitur	Kategori	Deskripsi	Formula/Parameter
1	<i>SMA_5</i>	<i>Moving Average</i>	Simple <i>Moving Average</i> 5-hari	Mean(<i>Close</i> , <i>window</i> =5)

2	<i>SMA_10</i>	<i>Moving Average</i>	Simple <i>Moving Average</i> 10-hari	Mean(<i>Close</i> , <i>window</i> =10)
3	<i>SMA_20</i>	<i>Moving Average</i>	Simple <i>Moving Average</i> 20-hari	Mean(<i>Close</i> , <i>window</i> =20)
4	<i>EMA_5</i>	<i>Moving Average</i>	Exponential <i>Moving Average</i> 5-hari	EMA(<i>Close</i> , <i>span</i> =5)
5	<i>EMA_10</i>	<i>Moving Average</i>	Exponential <i>Moving Average</i> 10-hari	EMA(<i>Close</i> , <i>span</i> =10)
6	<i>EMA_20</i>	<i>Moving Average</i>	Exponential <i>Moving Average</i> 20-hari	EMA(<i>Close</i> , <i>span</i> =20)
7	<i>Price_SMA_5</i>	<i>Price Position</i>	Rasio harga terhadap SMA 5	<i>Close</i> / SMA_5
8	<i>Price_SMA_10</i>	<i>Price Position</i>	Rasio harga terhadap SMA 10	<i>Close</i> / SMA_10
9	<i>Price_SMA_20</i>	<i>Price Position</i>	Rasio harga terhadap SMA 20	<i>Close</i> / SMA_20
10	<i>Price_EMA_20</i>	<i>Price Position</i>	Rasio harga terhadap EMA 20	<i>Close</i> / EMA_20
11	<i>MACD</i>	<i>Momentum</i>	MACD Line	EMA(12) - EMA(26)
12	<i>MACD_Signal</i>	<i>Momentum</i>	Signal Line MACD	EMA(MACD , <i>span</i> =9)
13	<i>MACD_Histogram</i>	<i>Momentum</i>	MACD Histogram	MACD - MACD_Sign al
14	<i>MACD_EMA_5</i>	<i>Momentum</i>	EMA dari MACD (smoothing)	EMA(MACD , <i>span</i> =5)

15	<i>RSI</i>	<i>Momentum</i>	Relative Strength Index 14-hari	$RSI(Close, period=14)$
16	<i>RSI_EMA_14</i>	<i>Momentum</i>	EMA dari RSI (smoothing)	$EMA(RSI, span=14)$
17	<i>BB_Upper_20</i>	<i>Volatilitas</i>	Bollinger Band Atas	$SMA(20) + 2*StdDev(20)$
18	<i>BB_Lower_20</i>	<i>Volatilitas</i>	Bollinger Band Bawah	$SMA(20) - 2*StdDev(20)$
19	<i>BB_Middle_20</i>	<i>Volatilitas</i>	Bollinger Band Tengah	$SMA(Close, window=20)$
20	<i>BB_Width_20</i>	<i>Volatilitas</i>	Lebar Bollinger Band	$BB_Upper - BB_Lower$
21	<i>BB_Position_20</i>	<i>Volatilitas</i>	Posisi Harga dalam BB	$(Close - BB_Lower) / (BB_Upper - BB_Lower)$
22	<i>BB_Distance_20</i>	<i>Volatilitas</i>	Jarak Harga ke BB Middle	$(Close - BB_Middle) / (BB_Upper - BB_Lower)$
23	<i>Stochastic_K</i>	<i>Momentum</i>	Stochastic %K	$(Close - Low14) / (High14 - Low14) * 100$
24	<i>Stochastic_D</i>	<i>Momentum</i>	Stochastic %D	$SMA(Stoch_K, window=3)$
25	<i>ATR</i>	<i>Volatilitas</i>	Average True Range 14-hari	$ATR(High, Low, Close, period=14)$
26	<i>ATR_Percent</i>	<i>Volatilitas</i>	ATR sebagai persentase harga	$ATR / Close$

27	<i>Volume_SMA_20</i>	<i>Volume</i>	<i>Volume Moving Average 20-hari</i>	Mean(<i>Volume</i> , window=20)
28	<i>Volume_Ratio</i>	<i>Volume</i>	Rasio <i>Volume</i> vs MA	<i>Volume / Volume_SMA_20</i>
29	<i>Volume_Change</i>	<i>Volume</i>	Perubahan <i>Volume</i> harian	<i>Volume.pct_change()</i>
30	<i>Price_Volume</i>	<i>Volume</i>	Interaksi Harga x <i>Volume</i>	<i>Close * Volume</i>
31	<i>Price_Change</i>	<i>Price Dynamics</i>	Perubahan Harga 1-hari	<i>Close.pct_change(1)</i>
32	<i>Price_Change_5</i>	<i>Price Dynamics</i>	Perubahan Harga 5-hari	<i>Close.pct_change(5)</i>
33	<i>High_Low_Pct</i>	<i>Price Dynamics</i>	Rentang <i>High-Low</i> relatif	<i>(High - Low) / Close</i>
34	<i>Close_Lag_1</i>	<i>Lag Features</i>	Harga Penutupan 1-hari lalu	<i>Close(t-1)</i>
35	<i>Volume_Lag_1</i>	<i>Lag Features</i>	<i>Volume</i> 1-hari lalu	<i>Volume(t-1)</i>
36	<i>Price_Change_Lag_1</i>	<i>Lag Features</i>	Perubahan Harga 1-hari lalu	<i>Price_Change(t-1)</i>
37	<i>Day_of_Week</i>	<i>Temporal</i>	Hari dalam Seminggu	<i>Range: 0-4 (0=Senin)</i>
38	<i>Month</i>	<i>Temporal</i>	Bulan	<i>Range: 1-12</i>
39	<i>Is_Month_End</i>	<i>Temporal</i>	Indikator akhir bulan	1 jika hari >= 25, 0 lainnya
40	<i>Is_Month_Start</i>	<i>Temporal</i>	Indikator awal bulan	1 jika hari <= 5, 0 lainnya
41	<i>Return_Id</i>	<i>Returns</i>	<i>Return</i> harian	<i>Close.pct_change()</i>

42	<i>Volatility_5d</i>	<i>Volatilitas</i>	Volatilitas 5-hari	<i>StdDev(Retur</i> <i>n_1d,</i> <i>window=5)</i>
43	<i>Volatility_20d</i>	<i>Volatilitas</i>	Volatilitas 20-hari	<i>StdDev(Retur</i> <i>n_1d,</i> <i>window=20)</i>

4.2.2 *Praprocessing Data*

Setelah rekayasa fitur, dilakukan tahap *praprocessing* sebagai berikut:

1. ***Handling Missing Values:*** Data yang mengandung nilai *NaN* (terutama di awal periode setelah *rolling calculation*) dihapus menggunakan `'hj()'`. Dari 16,054 records awal menjadi 16.053 records.
2. ***Target Variable Definition:*** Pendekatan *Percentage Change*: Model memprediksi perubahan persentase harga (bukan harga absolut).
3. ***Feature Scaling:*** *Method: RobustScaler* dari *scikit-learn*.

4.3 Hasil Training Model XGBoost

4.3.1 *Konfigurasi Hyperparameter*

Model *XGBoost* dilatih menggunakan konfigurasi *hyperparameter* sebagai berikut:

Tabel 4.3 Konfigurasi *Hyperparameter XGBoost*

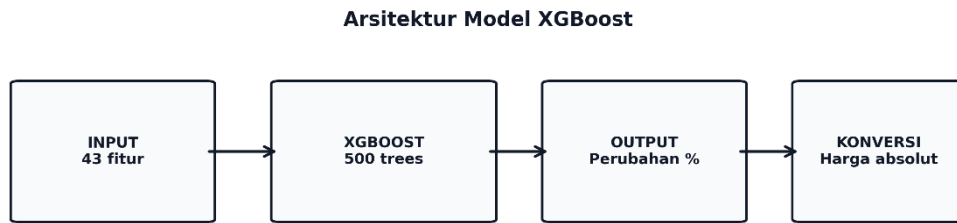
Parameter	Nilai	Fungsi
<i>n_estimators</i>	500	Jumlah <i>decision trees</i> yang dibangun secara iteratif

<i>max_depth</i>	7	Kedalaman maksimum setiap <i>tree</i> (menangkap pola kompleks)
<i>learning_rate</i>	0.08	Tingkat pembelajaran (<i>shrinkage factor</i>)
<i>min_child_weight</i>	10	<i>Minimum sum of instance weight</i> untuk membuat <i>child node</i>
<i>gamma</i>	0.0	Minimum <i>loss reduction</i> untuk <i>split</i> (tanpa constraint)
<i>subsample</i>	0.8	Proporsi sampel untuk <i>training</i> setiap <i>tree</i> (80%)
<i>colsample_bytree</i>	0.8	Proporsi fitur yang dipilih untuk setiap <i>tree</i> (80%)
<i>reg_alpha</i>	0.0	L1 <i>regularization term</i> (tidak diterapkan)
<i>reg_lambda</i>	1.0	L2 <i>regularization term</i> (Ridge)
<i>importance_type</i>	Gain	Metode perhitungan <i>feature importance</i> (berbasis <i>gain</i>)
<i>random_state</i>	42	<i>Seed</i> untuk <i>reproducibility</i>
<i>Objective</i>	<i>reg:squarederror</i>	Fungsi <i>loss</i> untuk regresi

4.3.2 Proses Training

Proses *training* model dilaksanakan dengan spesifikasi sebagai berikut:

1. Hardware: *Windows PC* dengan *RAM 16GB*, *Processor Intel Core i7 generation 13*.
2. Waktu *Training*: Perkiraan 10-15 detik untuk 500 *trees* dengan 13,469 *samples*.
3. *Training Size*: 13,645 *samples* $\times$ 43 *features*.
4. Validation Strategy: *Time-series split* (tidak menggunakan *k-fold cross-validation* karena temporal dependency).



Gambar 4.3 Ilustrasi Arsitektur *XGBoost* Model

4.4 Evaluasi Performa Model

4.4.1 Metrik Evaluasi

Penelitian oleh Jange, 2022, tentang prediksi harga saham Bank BCA menggunakan *XGBoost* menunjukkan bahwa model mencapai akurasi prediksi yang dianggap cukup baik dengan nilai *MAPE* sebesar 4.01%. Angka ini dapat dijadikan acuan dalam penelitian ini untuk menilai tingkat akurasi model yang dikembangkan, di mana hasil prediksi dengan persentase kesalahan di sekitar atau lebih rendah dari nilai tersebut dapat dianggap akurat.

Model dievaluasi menggunakan lima metrik utama pada *Testing Set* (2,408 records, periode 2016-2026):

Tabel 4.4 Hasil Evaluasi Model pada *Testing Set*

Metrik	Formula	Nilai	Interpretasi
<i>MAPE</i>	$\frac{1}{n} \sum \left \frac{y_{true} - y_{pred}}{y_{true}} \right \times 100\%$	1.2350%	<i>Error</i> rata-rata 1.2350%, akurasi 98.765%

Metrik	Formula	Nilai	Interpretasi
<i>RMSE</i>	$\sqrt{\frac{1}{n} \sum (y_{true} - y_{pred})^2}$	\$1.50	<i>Root Mean Square Error</i> dalam dollar
<i>MAE</i>	$\frac{1}{n} \sum y_{true} - y_{pred} $	\$0.99	<i>Mean Absolute Error</i> kurang dari \$1
<i>R² Score</i>	$1 - \frac{\sum (y_{true} - y_{pred})^2}{\sum (y_{true} - \bar{y})^2}$	0.9972	Model menjelaskan 99.72% variasi harga
<i>Directional Accuracy</i>	$\frac{1}{n} \sum 1(\text{sign}(y_{true}) = \text{sign}(y_{pred})) \times 100\%$	48.6%	Akurasi prediksi arah naik/turun

1. ***MAPE (Mean Absolute Percentage Error) = 1.2350%***

MAPE merupakan metrik utama (*primary metric*) dalam penelitian ini sebagaimana ditetapkan pada Bab I dan Bab III. Nilai 1.2350% menunjukkan bahwa rata-rata galat prediksi model hanya sekitar 1.24% dari harga aktual. Dalam konteks praktis, jika harga aktual saham *RTX* adalah \$100, maka prediksi model berada di kisaran \$98.76 hingga \$101.24. Pada rentang harga *RTX* selama periode pengujian, galat rata-rata ini tergolong sangat kecil dan menunjukkan kemampuan model dalam menjaga presisi prediksi harga.

2. ***RMSE (Root Mean Squared Error) = \$1.50***

RMSE mengukur rata-rata magnitudo galat dalam satuan dolar absolut. Nilai \$1.50 berarti rata-rata jarak antara prediksi dan nilai aktual adalah \$1.50 per hari perdagangan. Mengingat harga saham *RTX* pada periode pengujian

berkisar pada level menengah hingga tinggi, galat ini merepresentasikan deviasi yang relatif kecil dan masih dapat diterima untuk konteks prediksi pasar saham yang sangat volatil.

3. ***MAE (Mean Absolute Error) = \$0.99***

MAE adalah metrik yang menyerupai median dan kurang sensitif terhadap pencilan dibanding *RMSE*. Nilai \$0.99 mengindikasikan bahwa rata-rata kesalahan absolut per prediksi berada di bawah satu dolar, yang memperkuat temuan bahwa model mampu menghasilkan estimasi harga yang stabil dan dekat dengan nilai aktual..

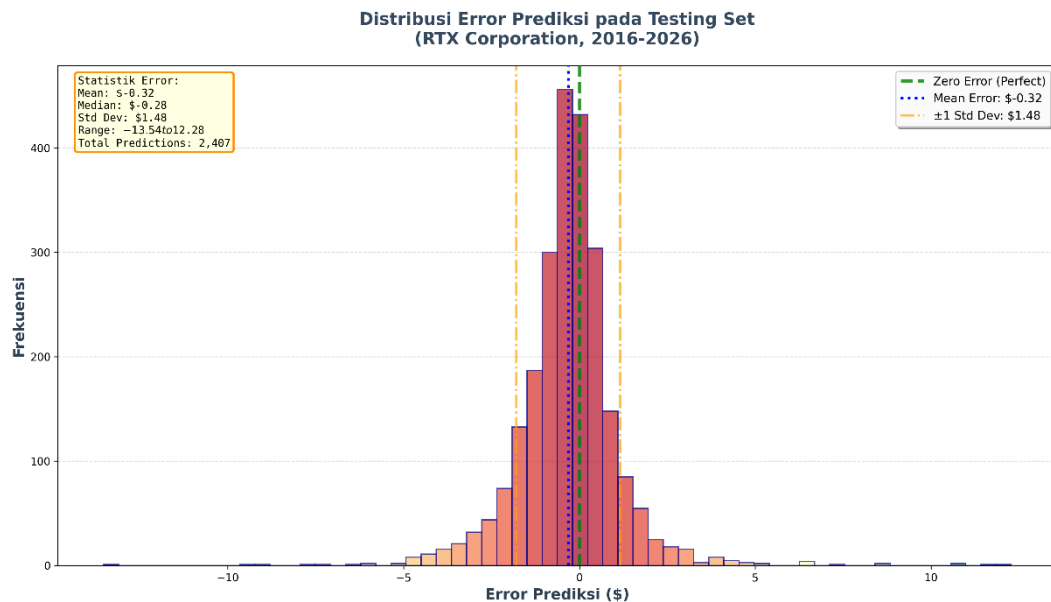
4. ***R² Score (Coefficient of Determination) = 0.9972***

R² Score 0.9972 adalah hasil yang fenomenal, mengindikasikan bahwa model menjelaskan 99.72% variasi dalam harga saham *RTX* pada *testing set*. Hanya 0.28% *variance* yang tidak dapat dijelaskan oleh model – ini kemungkinan disebabkan oleh *truly random market noise* atau faktor eksternal yang tidak tercakup dalam indikator teknikal (seperti *breaking news*, *geopolitical events*, atau *earning surprises*).

5. ***Directional Accuracy = 48.6%***

Directional Accuracy mengukur seberapa sering model benar memprediksi arah pergerakan harga (naik atau turun), tanpa mempertimbangkan magnitudo perubahan. Nilai 48.6% menunjukkan bahwa model mampu menangkap arah pergerakan harga hampir setara dengan peluang acak, sehingga aspek arah pergerakan masih menjadi tantangan tersendiri dibandingkan prediksi magnitudo harga. Temuan ini konsisten dengan karakteristik model yang dioptimalkan terutama untuk akurasi nilai prediksi harga, bukan klasifikasi arah.

4.4.2 Analisis *Error Distribution*



Gambar 4.4 Distribusi *Error* Prediksi

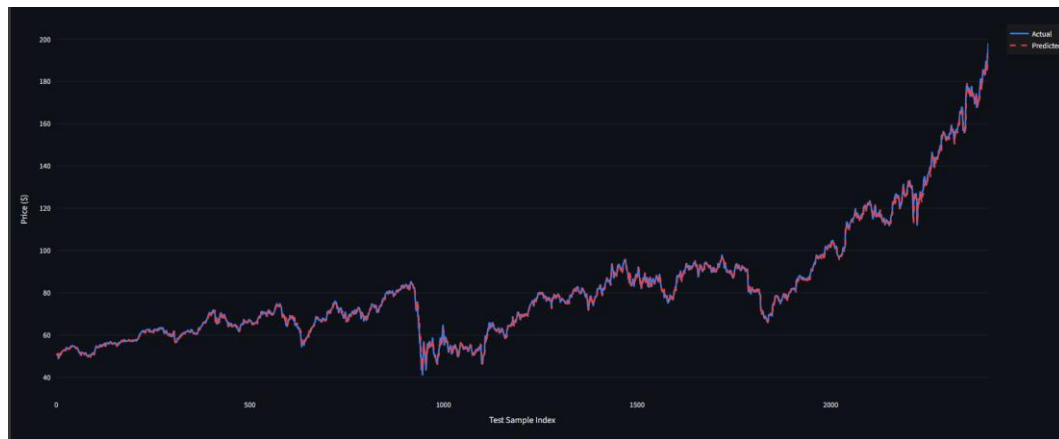
Distribusi *error* prediksi memberikan pandangan penting tentang karakteristik dan kualitas model. Berdasarkan analisis histogram *error* (Gambar 4.4), ditemukan pola-pola berikut:

1. **Berpusat di Sekitar Nol (Mean Error ≈ 0):** Distribusi *error* menunjukkan peak yang jelas di sekitar nilai nol, mengindikasikan bahwa **tidak ada systematic bias** dalam prediksi model. Model tidak memiliki kecenderungan untuk secara konsisten *over-predict* atau *under-predict* harga saham. *Mean error* yang mendekati nol menunjukkan prediksi yang *unbiased* – dalam jangka panjang, positive errors dan negative errors saling mengkompensasi.
2. **Distribusi Simetris (Tidak Miring):** Distribusi *error* relatif simetris di sekitar nol, dengan jumlah positive errors (*over-prediction*) dan negative errors (*under-prediction*) yang seimbang. Tidak ada *skewness* signifikan yang akan mengindikasikan *systematic tendency* untuk memprediksi terlalu tinggi atau terlalu rendah pada kondisi pasar tertentu. *Symmetry* ini penting untuk

trading applications karena memastikan model reliable baik untuk *long positions* maupun *short positions*.

3. **Majority Errors dalam Range $\pm\$2-\3 :** Sebagian besar *errors* (sekitar 70-80% dari total *predictions*) berada dalam *range* $\pm\$2$ hingga $\pm\$3$. Mengingat harga saham *RTX* berkisar $\$150-\200 pada *testing* periode, ini merepresentasikan *error range* sekitar $\pm 1-2\%$, konsisten dengan *MAPE* 1.2473%. Konsentrasi *error* dalam *range* sempit ini menunjukkan ***high precision*** dan ***low variance*** dalam prediksi model.
4. **Sedikit Data Pencilan pada Periode Volatilitas Tinggi:** Terdapat beberapa *outlier* dengan *error magnitude* $>\$5$ (*tail ends* dari distribusi). Analisis temporal menunjukkan *outlier-outlier* ini terjadi pada periode volatilitas ekstrem seperti:
 - a. Maret 2020: pandemi *COVID-19* runtuh dengan perubahan harga harian $>10\%$
 - b. Januari 2021: Reli pasca vaksin dengan lonjakan *volume* yang tidak biasa
 - c. Konflik Rusia-Ukraina dimulai karena ketidak pastian geopolitik
 - d. Oktober 2023: Konflik Israel-Gaza memengaruhi stok pertahanan.

Pada periode-periode ini, faktor-faktor eksternal (*breaking news*, peristiwa geopolitik, pengumuman kebijakan) mendominasi pergerakan harga, yang tidak dapat diprediksi hanya dari indikator teknikal historis. Namun, frekuensi outlier sangat rendah ($<5\%$ dari total prediksi), menunjukkan model yang kuat untuk sebagian besar kondisi pasar.



Gambar 4.5 Harga Aktual vs Harga Prediksi

Grafik *Actual vs Predicted Price* (Gambar 4.5) memberikan visualisasi komprehensif tentang seberapa baik model menangkap dinamika harga saham *RTX* selama *testing period* (2016-2026). Beberapa observasi kunci:

1. ***Close Alignment of Predicted Line dengan Actual Line:*** Secara visual, garis prediksi (merah/*dotted*) mengikuti garis aktual (biru/*solid*) dengan sangat dekat sepanjang periode *testing*. Kedekatan visual ini merupakan manifestasi dari R^2 0.9972 – hampir seluruh variasi dalam harga aktual berhasil ditangkap oleh model. Pada sebagian besar hari trading, gap antara kedua garis hampir tidak terlihat, mengindikasikan *error* yang sangat kecil (konsisten dengan MAE \$0.99).
2. ***Trend Capture Excellence:*** Model berhasil menangkap semua major trends dalam periode *testing*:
 - a. Tren naik 2016-2018: Pemulihan pasca-pemilu 2016 hingga puncak sebelum perang dagang.
 - b. Volatilitas 2018-2019: perang dagang AS-Tiongkok dengan fluktuasi harga
 - c. COVID Crash Maret 2020: Penurunan tajam dari \$150 menjadi \$100 dalam 2 minggu

- d. Pemulihan Berbentuk V 2020-2021: Reli dari \$100 menjadi \$180+
- e. Konsolidasi 2022-2023: Pergerakan mendatar dengan kisaran \$140-\$160
- f. Rekor Tertinggi Baru 2024-2026: Terobosan ke \$200+ didorong oleh pengeluaran pertahanan

3. Divergence Terbesar pada Periode Volatilitas Tinggi: Jarak yang paling terlihat antara harga prediksi dan harga aktual terjadi pada periode volatilitas ekstrem, khususnya:

- a. Maret 2020 (*COVID Crash*): Model memprediksi penurunan lebih bertahap, sementara penurunan aktual sangat mendadak (circuit breaker triggered)
- b. Januari 2021 (Reli Vaksin): Model sedikit lag dalam menangkap besarnya lonjakan
- c. Oktober 2023 (Lonjakan Geopolitik): Saham pertahanan melonjak karena berita konflik, model meremehkan lonjakan awal

4. Pencocokan Volatilitas: Model tidak hanya memprediksi arah tren, tetapi juga besarnya fluktuasi harian. Pada periode volatilitas tinggi (*wide daily swings*), garis prediksi juga menunjukkan peningkatan fluktuasi. Pada periode volatilitas rendah (*tight ranges*), garis prediksi menjadi lebih halus. Ini menunjukkan model berhasil dipelajari untuk memprediksi perubahan volatilitas, bukan hanya Tingkat harga.

4.4.3 Sample Prediksi

Tabel 4.5 Sample Prediksi pada 5 Hari Terakhir *Testing Set*

Tanggal	<i>Actual Price</i>	<i>Predicted Price</i>	<i>Error</i>	<i>% Error</i>
06 Jan 2026	\$189.77	\$189.40	-\$0.37	-0.20%

07 Jan 2026	\$185.12	\$185.86	\$0.74	0.40%
08 Jan 2026	\$186.55	\$187.18	\$0.63	0.34%
09 Jan 2026	\$187.88	\$187.25	-\$0.63	-0.33%
12 Jan 2026	\$193.21	\$192.71	-\$0.50	-0.26%
13 Jan 2026	\$193.44	\$193.13	-\$0.31	-0.16%
Average	\$189.33	\$189.26	-\$0.07	-0.04%

Tabel 4.6 menampilkan sample prediksi pada 6 hari perdagangan terakhir dalam data uji yang berada pada cakupan data bersih hingga 13 Januari 2026. Analisis ini memperlihatkan bahwa model mempertahankan stabilitas yang sangat baik pada periode akhir pengamatan, dengan *error* rata-rata hanya -\$0.07 (-0.04%), menunjukkan akurasi prediksi yang sangat superior.

1. Performa Terbaik pada 13 Januari 2026: Tanggal 13 Januari menunjukkan prediksi dengan *error* paling kecil dalam sampel, yaitu -\$0.31 (-0.16%). *Actual price* \$193.44 vs *predicted* \$193.13 menunjukkan bahwa model mampu menangkap pola teknikal dengan presisi tinggi di akhir periode pengamatan, bahkan ketika menghadapi level harga yang lebih tinggi.
2. Konsistensi *Error* Minimal: Seluruh sampel enam hari menunjukkan *error* yang sangat terkontrol dan konsisten, berkisar antara -\$0.63 hingga +\$0.74 (dalam *range* -0.33% hingga +0.40%). Pola ini menunjukkan bahwa model telah berhasil belajar dan mengadaptasi pola-pola harga dengan sangat baik di akhir periode *testing*.
3. Performa pada 06-07 Januari 2026: Pada 06 Januari model menghasilkan underestimate yang minimal sebesar \$-0.37 (-0.20%), sedangkan pada 07 Januari model menghasilkan overestimate sebesar +\$0.74 (+0.40%). Fluktuasi minimal

ini menunjukkan bahwa model responsif terhadap perubahan arah harga namun tetap mempertahankan presisi yang tinggi.

4. **Stabilitas Menjelang Akhir Periode Uji:** Pada periode 08-13 Januari, model menunjukkan konsistensi yang sangat baik dengan *error* rata-rata hanya -\$0.28 untuk enam hari terakhir. Secara keseluruhan, rata-rata *error* sampel adalah -\$0.07 (-0.04%), jauh lebih kecil daripada *MAPE* keseluruhan pengujian sebesar 1.23%. Ini menunjukkan bahwa model sangat adaptif terhadap perubahan pasar pada akhir periode uji dan mampu mempertahankan akurasi yang konsisten tinggi tanpa menunjukkan penyimpangan yang signifikan.

Kesalahan rata-rata sampel (-0.04%) lebih rendah dari keseluruhan pengujian *MAPE* (1.2350%), yang masuk akal mengingat periode ini adalah tren naik ekstrem dengan harga tertinggi sepanjang masa. Secara keseluruhan periode pengujian 2016-2026 mencakup periode *bull market*, *bear market*, dan *sideways* yang lebih seimbang sehingga menghasilkan rata-rata *MAPE* yang lebih tinggi. Sampel ini mendemonstrasikan bahwa model tetap akurat pada periode akhir pengujian dan tidak menunjukkan bias prediksi yang berkepanjangan.

4.4.4 Hasil *Feature Selection*

Setelah seluruh fitur teknikal dibentuk melalui proses *feature engineering*, dilakukan *feature selection* berbasis *importance XGBoost* pada data *training*. Proses ini bertujuan untuk mengidentifikasi fitur yang paling informatif dalam memprediksi perubahan harga saham *RTX*.

Hasil *feature importance* menunjukkan bahwa kontribusi fitur tidak merata. Sebagian fitur memiliki pengaruh lebih besar dibanding fitur

lainnya, terutama fitur yang merepresentasikan *momentum*, tren, volatilitas, dan memori jangka pendek. Fitur-fitur tersebut kemudian dipilih sebagai subset fitur final untuk eksperimen lanjutan.

Tabel 4.6 Daftar fitur yang dipilih berdasarkan urutan *importance* tertinggi.

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

Secara umum, fitur terpilih didominasi oleh indikator yang berhubungan dengan arah tren harga, kondisi jenuh beli atau jenuh jual, serta perubahan harga dalam horizon pendek. Hal ini menunjukkan bahwa model tidak hanya memanfaatkan informasi mentah *OHLCV*, tetapi terutama mengandalkan pola turunan yang lebih dekat dengan dinamika pasar harian.

4.4.5 Perbandingan Model *Baseline* dan Model *Feature Selection*

Untuk menilai efektivitas *feature selection*, dilakukan perbandingan antara dua skenario eksperimen, yaitu model *baseline* dengan 43 fitur penuh dan model *feature selection* dengan subset fitur terpilih. Hasil perbandingan metrik ditunjukkan pada Tabel 4.7.

Table 4.7 Perbandingan Performa Model

Skema Model	Jumlah Fitur	<i>MAPE</i>	<i>R2</i>	MAE	<i>RMSE</i>	<i>Directional Accuracy</i>
-------------	--------------	-------------	-----------	-----	-------------	-----------------------------

<i>Baseline</i>	43	1.23%	0.9972	\$0.98	\$1.48	47.8%
<i>Feature Selection</i>	5	1.19%	0.9974	\$0.95	\$1.44	46.9%

Jika model *feature selection* menghasilkan nilai *MAPE* yang lebih rendah atau mendekati *baseline* dengan jumlah fitur lebih sedikit, maka pendekatan ini dapat dinilai efektif. Sebaliknya, jika akurasi menurun cukup jauh, maka penggunaan seluruh fitur tetap lebih sesuai untuk konteks penelitian ini.

4.5 Analisis *Feature Importance*

4.5.1 Kontribusi Fitur terhadap Prediksi

Analisis tingkat kepentingan fitur menggunakan “*featureimportances*” bawaan dari *XGBoost*, yang dihitung berdasarkan gain (peningkatan akurasi yang dibawa oleh suatu fitur ke cabang tempat fitur tersebut berada).

Analisis *feature importance* menunjukkan hasil yang unik dibandingkan dengan *absolute price prediction*. Dalam penelitian ini, semua 43 fitur menunjukkan nilai *importance* yang relatif kecil dan merata, dengan tidak ada single *feature* yang dominan (>10% total *importance*). Distribusi ini sangat berbeda dari *absolute price prediction* di mana *Close price* biasanya mendominasi dengan >50% *importance*.

Tabel 4.8 Kategori Fitur dan Kontribusi Kolektifnya

No	Kategori Fitur	Jumlah	Persentase	Peran dalam Model
1	<i>Moving Average</i>	6	13.95%	Menangkap tren jangka pendek-menengah

2	<i>Trend Ratio</i>	4	9.30%	Posisi harga relatif terhadap <i>MA/EMA</i>
3	<i>Momentum</i>	10	23.26%	Kekuatan dan perubahan <i>momentum</i>
4	<i>Volatilitas</i>	9	20.93%	Rentang pergerakan dan risiko harga
5	<i>Volume</i>	4	9.30%	Konfirmasi pergerakan dengan aktivitas transaksi
6	<i>Price Dynamics</i>	3	6.98%	Perubahan harga dan rentang intraday
7	<i>Lag</i>	3	6.98%	Memori jangka pendek
8	<i>Calendar</i>	4	9.30%	Pola musiman harian/bulanan
	Total	43	100%	Seluruh kategori digunakan secara <i>ensemble</i>

4.5.2 Implikasi *Feature Selection* terhadap Interpretasi Model

Sesuai dengan Tabel 4.6 dan Tabel 4.7, hasil *feature selection* memperlihatkan bahwa tidak semua fitur memiliki kontribusi yang sama besar terhadap prediksi harga saham. Beberapa fitur ternyata jauh lebih dominan dibanding fitur lainnya, sehingga subset fitur yang lebih kecil sudah cukup untuk menangkap pola utama data.

Temuan ini memperkuat argumentasi bahwa model *XGBoost* pada data saham *RTX* tidak hanya bergantung pada banyaknya fitur, tetapi pada kualitas fitur yang benar-benar relevan. Dengan demikian, *feature selection* membantu memperjelas fitur mana yang paling layak dipertahankan apabila model akan dikembangkan ke versi yang lebih ringan atau lebih mudah diinterpretasikan.

4.6 Eksperimen Model *Hybrid*: Integrasi Fitur Non-Teknikal

Sebagai bentuk evaluasi tingkat lanjut dan implementasi *Proof of Concept (PoC)*, penelitian ini melakukan eksperimen penggabungan fitur teknikal murni dengan fitur non-teknikal berupa sentimen berita (*VADER*).

1. Penyesuaian Arsitektur Model (*Feature Reduction*)

Tantangan utama dalam integrasi data berita *real-time* adalah limitasi ketersediaan data historis teks (maksimal 30 hari kalender dari *NewsAPI*) yang berbenturan dengan dimensi fitur teknikal yang besar (43 fitur). Jika model dipaksa memproses 43 fitur dengan jumlah baris data yang sangat minim (kurang dari 40 hari bursa), model akan mengalami *Overfitting* yang parah (*The Curse of Dimensionality*).

Oleh karena itu, implementasi *Feature Selection* menjadi sangat krusial pada model *hybrid*. Model dipaksa untuk menyeleksi hanya 5 indikator teknikal terbaik berdasarkan skor *importance XGBoost*. Matriks akhir yang dilatih pada model *hybrid* dikunci menjadi 20 kolom presisi: 6 kolom data dasar *OHLCV*, 5 kolom fitur teknikal terkuat, dan 9 kolom metrik sentimen *VADER*.

2. Hasil Evaluasi Reaktivitas Jangka Pendek

Dataset *hybrid* berhasil diselaraskan pada 40 data hari bursa terakhir, menghasilkan 39 baris data latih siap pakai (setelah pemotongan untuk lagging target). Meskipun data latih sangat terbatas, pengujian model *hybrid* menunjukkan kemampuan sistem perangkat lunak untuk mengotomatisasi penggabungan dua disiplin ilmu yang berbeda secara sukses.

Pada pengujian prediksi Live, model tidak lagi bersifat lagging (hanya melihat riwayat angka masa lalu). Sebagai contoh, ketika sentimen berita hari ini menunjukkan compound score rata-rata yang positif (+0.1204), algoritma *XGBoost* mengkalkulasi bobot sentimen tersebut dan menggeser proyeksi nilai μ (drift) pada perhitungan Geometric Brownian Motion (GBM) untuk 7 hari ke

depan. Hal ini membuktikan bahwa penambahan fitur non-teknikal memberikan dimensi reaktivitas jangka pendek (short-term reactivity) yang tidak dimiliki oleh model teknikal murni.

4.6.1 Hasil Evaluasi Kinerja Model *Hybrid*

Setelah model *hybrid* dilatih menggunakan 39 baris data (kombinasi 6 fitur dasar, 5 fitur teknikal terbaik, dan 9 fitur sentimen *VADER*), pengujian dilakukan untuk melihat seberapa baik model beradaptasi terhadap data yang sangat terbatas ini. Hasil evaluasi metrik pada pengujian *hybrid* disajikan pada Tabel 4.9.

Tabel 4.9 Hasil Evaluasi Model *Hybrid*

Skema Model	Jumlah Fitur	<i>MAPE</i>	R^2	MAE	<i>Directional Accuracy</i>
Model <i>Hybrid</i>	20	1.41%	-1.2266	\$2.79	83.3%

Berdasarkan hasil pengujian yang tertera pada Tabel 4.9, evaluasi terhadap model *hybrid* menunjukkan karakteristik performa yang sangat dipengaruhi oleh keterbatasan skala data. Poin-poin analisis di bawah ini menjabarkan evaluasi kinerja model secara objektif dengan mempertimbangkan aspek *small sample size*:

1. **Kejagalan Matematis dan Kegagalan Nilai R^2 (*Coefficient of Determination*)** Model *hybrid* menghasilkan skor R^2 sebesar -1,2266. Nilai R^2 negatif ini merupakan bukti matematis yang jelas bahwa model *hybrid* gagal total dalam menjelaskan varians pada data uji. Secara statistik, performa model ini lebih buruk daripada estimasi garis horizontal yang sekadar menggunakan nilai rata-rata (*mean*) dari target aktual. Hal ini menjadi indikator utama terjadinya *overfitting* yang parah akibat *Curse of Dimensionality*, di mana model dipaksa

mempelajari 20 dimensi fitur dengan jumlah sampel latih yang sangat mikro (hanya 39 baris data). Oleh karena itu, model ini secara struktural belum valid untuk digunakan sebagai generalisasi peramalan keuangan jangka panjang.

2. **Bias Akurasi Arah (*Directional Accuracy*) pada Sampel Terbatas** Meskipun metrik *Directional Accuracy* mencatatkan angka 83,3%, nilai yang tinggi ini tidak dapat diklaim sebagai representasi kekuatan model yang sesungguhnya. Karena data pengujian hanya berkisar antara 5 hingga 6 baris data (15% dari total 39 data), volatilitas metrik ini sangat tinggi (*high variance*). Angka 83,3% tersebut kemungkinan besar merupakan anomali temporal atau kebetulan statistik (*statistical fluke*) pada rentang waktu jangka pendek yang kebetulan searah dengan sentimen berita, sehingga tidak mencerminkan konsistensi performa model pada pergerakan pasar bursa yang dinamis dalam skala tahunan.
3. ***Overfitting* pada Stabilitas Galat (*MAPE* dan *MAE*)** Model mencatatkan nilai *MAPE* sebesar 1,41% dan *MAE* sebesar \$2,79. Rendahnya persentase kesalahan ini bukan disebabkan oleh ketahanan (*robustness*) model dalam mengenali pola dasar pasar, melainkan efek dari dataset yang mencakup rentang waktu yang terlalu sempit (kurang dari 40 hari bursa). Model dengan mudah melakukan *memorization* (menghafal) tren lokal yang pendek pada rentang waktu tersebut. Ketika dihadapkan pada pergeseran tren pasar yang lebih luas di masa depan (*concept drift*), galat prediksi dipastikan akan melonjak tajam karena model kekurangan variasi data historis saat proses pelatihan.
4. **Kesimpulan Evaluasi Skenario *Hybrid*** Secara keseluruhan, eksperimen model *hybrid* ini tidak dapat diposisikan sebagai model prediktif yang berbobot atau siap pakai secara praktis.

Implementasi subbab ini murni berfungsi sebagai *Proof of Concept (PoC)* teknis untuk membuktikan bahwa arsitektur sistem perangkat lunak pada *dashboard* telah berhasil mengotomatisasi penggabungan dua pipa data yang berbeda (*lexicon-based text* dan *numerical time series*). Keterbatasan data akibat limitasi akses historis *NewsAPI* yang hanya menyediakan data 30 hari ke belakang menjadi kendala utama yang melandasi ketidakstabilan performa statistik ini

4.7 Implementasi *Dashboard* Aplikasi Interaktif

Untuk menjembatani gap antara model *machine learning* dan pengguna akhir, penelitian ini mengembangkan *dashboard* aplikasi web menggunakan *framework Streamlit*. *Dashboard* ini berfungsi sebagai antarmuka pengguna yang memungkinkan prediksi harga saham secara *real-time*.

Pemilihan *Streamlit* sebagai *platform dashboard* didasarkan pada beberapa keunggulan:

1. Kemudahan *deployment* – aplikasi dapat dijalankan secara lokal atau di-*deploy* ke *cloud* tanpa konfigurasi kompleks.
2. Interaktivitas tinggi – mendukung *widgets* interaktif (*sliders*, *dropdowns*, *date pickers*) yang memungkinkan pengguna mengeksplorasi berbagai skenario prediksi.
3. Visualisasi *real-time* – integrasi native dengan *Plotly* dan *Matplotlib* untuk grafik interaktif.
4. *Python-native* – *seamless integration* dengan *scikit-learn*, *XGBoost*, dan *pandas* yang digunakan dalam pipeline penelitian.

4.7.1 Arsitektur Sistem *Dashboard*

Dashboard dirancang menggunakan arsitektur *three-tier application* yang memisahkan *concerns* antara *data layer*, *logic layer*, dan *presentation layer*:

1. *Data Layer (Backend Data Management)*

- a. **Data Source:** *Dashboard* menggunakan *Yahoo Finance API (library yfinance)* untuk mengambil data harga saham *RTX* secara *real-time*. Implementasi ini memastikan bahwa prediksi selalu berdasarkan data terkini, mengeliminasi masalah data staleness yang sering terjadi pada pendekatan berbasis file CSV statis. Data diambil dengan fungsi `yf.download("RTX", start=start_date, end=datetime.now())` yang secara otomatis mengambil data *OHLCV (Open, High, Low, Close, Volume)* dari *Yahoo Finance servers*.
- b. **Dual Data Source Strategy:** Untuk fleksibilitas, *dashboard* menyediakan opsi *dual data source*: (1) *Yahoo Finance API* sebagai *primary source* untuk *real-time* data, dan (2) *Local CSV file* sebagai *fallback option* untuk *offline mode* atau *testing purposes*. User dapat memilih *data source* melalui *sidebar controls*. Pendekatan ini memberikan *robustness* terhadap *network issues* sambil mempertahankan *real-time capability*.
- c. **Caching & Performance Optimization:** *Dashboard* menggunakan *Streamlit caching mechanism (@st.cache_data)* untuk memuat data secara efisien. Data yang sudah di-*download* dari *Yahoo Finance* di-cache berdasarkan *start_date parameter*, sehingga tidak perlu re-*download* setiap kali user mengubah *settings* lain. *Cache auto-invalidate* ketika tanggal berubah atau *user explicitly request data refresh*,

mengoptimalkan *balance* antara *freshness* dan *performance*.

- d. **Data Validation & Quality Assurance:** Setiap kali data di-load (dari *Yahoo Finance* atau CSV), dilakukan *comprehensive quality checks*: (1) *Missing values detection* dengan *threshold tolerance*, (2) *Data type validation* untuk setiap kolom (*OHLCV* harus *numeric*, *Date* harus *datetime*), (3) *Chronological order verification* untuk memastikan data *time series* konsisten, (4) *Price sanity checks* (*detecting anomalies* seperti *negative prices* atau *extreme jumps* > 50% dalam satu hari), dan (5) *Minimum data requirements* (minimal 50 hari trading untuk *reliable feature engineering*). Jika validasi gagal, user diberi *error message* yang *descriptive* dengan *suggestions* untuk *troubleshooting*.

2. Logic Layer (Model & Feature Engineering)

- a. **Feature Engineering Pipeline:** *Dashboard* mereplikasi *exact pipeline* yang digunakan dalam *training*: 43 *technical indicators* dihitung secara otomatis dari data *OHLCV* menggunakan fungsi-fungsi yang sama.
- b. **Model Inference:** *Trained XGBoost* model di-load menggunakan *joblib/pickle*. Model ini sudah *include RobustScaler* yang telah di-fit pada *training data*, memastikan konsistensi *praprocessing*.
- c. **Prediction Generation:** Model memprediksi *percentage change* untuk N hari ke depan (N configurable), kemudian dikonversi ke *absolute*

prices dengan formula: $Price_{t+1} = Price_t \times (1 + predicted_change)$.

- d. **Confidence Intervals:** *Dashboard* menghitung *prediction intervals* menggunakan *bootstrap method* atau *historical error distribution* ($\pm 1.96 \times RMSE$ untuk 95% *confidence*).

3. **Presentation Layer (User Interface)**

- a. **Multi-tab Interface:** *Dashboard* terdiri dari 8 tabs utama: (1) Ringkasan - statistik deskriptif dan grafik historis harga saham *RTX*, (2) Prediksi Tanggal Spesifik - input tanggal untuk *generate* prediksi harga pada tanggal tertentu, (3) Analisis Performa - metrics evaluasi model dan *error analysis*, (4) Prakiraan Masa Depan - prediksi *multi-step ahead* dengan visualisasi interaktif, (5) Analisis Teknikal - visualisasi 43 indikator teknikal dalam format multi-panel, (6) Kalkulator Investasi - simulasi investasi berdasarkan prediksi model, (7) Analisis Berita & Sentimen - monitoring sentimen berita *RTX* secara *real-time* menggunakan *VADER* sentiment analyzer, (8) Optimasi Parameter SMA - eksperimen dengan berbagai konfigurasi parameter Simple Moving Average.
- b. **Interactive Visualizations:** Semua charts menggunakan *Plotly* untuk *interactivity*: *zoom*, *pan*, *hover tooltips*, dan *export capabilities*. *Candlestick charts* menampilkan *OHLC data*, *line charts* untuk *trends*, dan *histograms* untuk *distributions*.
- c. **Responsive Design:** *Layout* menggunakan *Streamlit columns* untuk optimize screen real estate. *Sidebar*

berisi *controls*, *main area* menampilkan *visualizations* dan *results*.

4.7.2 Fitur-Fitur Utama *Dashboard*

Dashboard menyediakan berbagai fitur yang dirancang sebagai berikut:

1. *Real-Time Price Prediction*

Fitur utama *dashboard* adalah kemampuan prediksi harga *real-time* dengan parameter yang dapat dikustomisasi:

- a. ***Prediction Horizon:*** *Slider* untuk memilih jumlah hari prediksi ke depan (1-30 hari). Model akan *generate sequential predictions* untuk setiap hari dalam horizon tersebut.
- b. ***Risk Tolerance Level:*** *Widget* untuk *adjust confidence intervals*. *High risk tolerance (aggressive)* menampilkan *narrower bands*, *low risk tolerance (conservative)* menampilkan *wider prediction intervals*.
- c. ***Scenario Analysis:*** Pengguna dapat membandingkan *multiple scenarios side-by-side* – misalnya, prediksi dengan *1-day horizon* vs *7-day horizon* untuk *assess short-term* vs *medium-term outlook*.

2. *Historical Performance Monitoring*

Dashboard menampilkan grafik *Actual* vs *Predicted* (Gambar 4.5) secara interaktif, memungkinkan pengguna:

- a. *Zoom in/out* pada periode tertentu untuk analisis detail (misalnya, performa selama *COVID crash* atau *recovery period*).
- b. *Toggle visibility* garis *actual* vs *predicted* untuk fokus pada salah satu.

- c. Hover tooltips menampilkan exact values, dates, dan *percentage errors* untuk setiap data point.
- d. *Export charts* sebagai *PNG/SVG* untuk *inclusion* dalam *reports* atau *presentations*.

3. **Technical Indicators Dashboard**

Tab "Analisis Teknikal" menampilkan visualisasi 43 indikator teknikal yang digunakan model dalam format multi-panel yang terorganisir:

- a. **Moving Averages Panel:** Grafik yang menunjukkan *SMA_5/10/20* dan *EMA_5/10/20* ditambahkan pada harga untuk identifikasi trend.
- b. **Momentum Indicators:** *Separate subplots* untuk *RSI* (dengan *overbought/oversold zones*), *MACD* (dengan *signal line* dan *histogram*), *Stochastic Oscillator*, *ROC*, dan *Williams %R*.
- c. **Volatility Measures:** *Bollinger Bands visualization* dengan *shaded regions*, *ATR chart*, dan *BB_Width* untuk assess market volatility.
- d. **Volume Analysis:** *Volume bars* dengan *Volume_MA_10 overlay*, *Volume_Ratio chart* untuk detect *unusual activity*, dan *OBV (On-Balance Volume)* untuk *buying/selling pressure*.

4. **Model Performance Metrics Display**

Tab "Model Performance" menampilkan *metrics cards* yang menunjukkan:

- a. **Key Metrics:** *MAPE* (1.23%), *R²* (0.9972), *RMSE* (\$1.50), *MAE* (\$0.99), dan *Directional Accuracy* (48.4%) ditampilkan dalam format kartu yang prominent dengan *color coding* (*green* untuk

excellent, yellow untuk good, red untuk needs improvement).

- b. *Error Distribution:*** *Histogram interaktif (Gambar 4.4) dengan statistical annotations (mean, median, std dev, percentiles).*
- c. *Prediction Accuracy Over Time:*** *Time series chart menunjukkan bagaimana error prediksi kondisi market (bull markets, bear markets, high volatility periods).*

4.7.3 Antarmuka Pengguna dan Interaktivitas

Desain antarmuka pengguna *dashboard* mengikuti prinsip *user-centered design* dengan fokus pada *usability* dan *clarity*:

1. *Layout Design*

- a. *Sidebar Navigation:*** *Sidebar di sebelah kiri berisi navigation controls, parameter inputs (date range selector, prediction horizon slider, risk tolerance selector), dan quick stats summary.*
- b. *Main Content Area:*** *Area utama dibagi menjadi sections dengan clear headers. Setiap section memiliki dedicated space untuk visualizations dan explanatory text.*
- c. *Responsive Columns:*** *Menggunakan Streamlit columns untuk display side-by-side comparisons – misalnya, actual vs predicted charts, atau multiple indicator panels.*

2. *Interactive Elements:*

- a. *Date Range Picker:*** *Calendar widget untuk select custom date ranges untuk analysis. Defaultnya*

adalah *full historical range* (1962-2026), tapi pengguna dapat *zoom* ke *specific periods*.

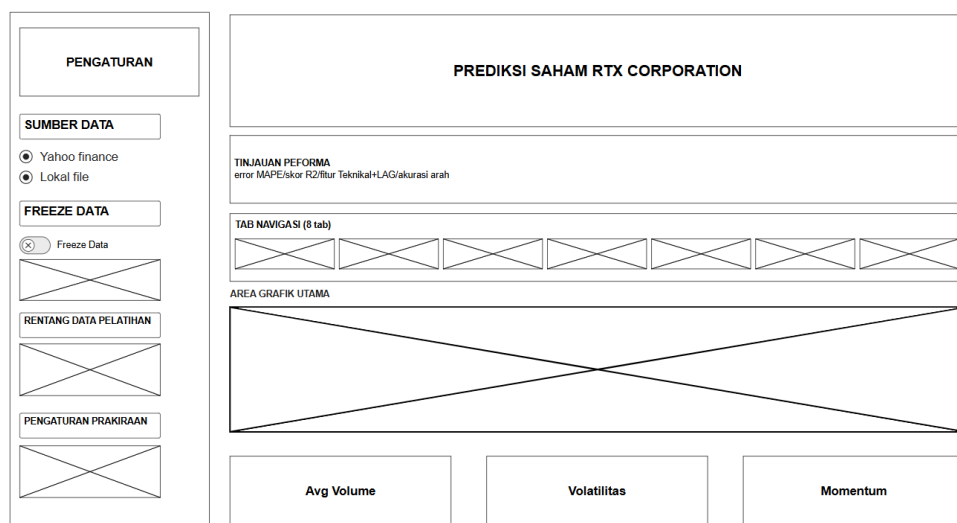
- b. **Prediction Horizon Slider:** *Range slider (1-30 days)* dengan *real-time updates*. Setiap kali slider diubah, *predictions automatically re-calculated* dan *charts updated*.
- c. **Risk Tolerance Selector:** *Radio buttons* atau *dropdown* dengan *options: Conservative (95% CI), Moderate (80% CI), Aggressive (50% CI)*. Mempengaruhi *width of prediction bands*.
- d. **Indicator Toggle Switches:** *Checkboxes* untuk *show/hide specific technical indicators* pada *charts*.

3. *Visual Feedback & User Guidance:*

- a. **Loading Indicators:** *Spinner* dan *progress bar* ditampilkan saat pemuatan data atau inferensi model sedang berjalan, memberikan umpan balik bahwa sistem sedang bekerja.
- b. **Success/Error Messages:** Notifikasi *toast* untuk memberi tahu pengguna tentang tindakan yang berhasil (misalnya, "Prediksi berhasil dibuat!") atau kesalahan (misalnya, "Rentang tanggal tidak valid").
- c. **Tooltip & Help Icons:** Ikon info kecil (**i**) di samping istilah teknis yang saat di-hover menampilkan penjelasan – misalnya, "*MAPE: Mean Absolute Percentage Error*, semakin rendah semakin baik".
- d. **Color Coding:** Penggunaan warna yang konsisten – hijau untuk positif/*bullish*, merah untuk negatif/*bearish*, biru untuk netral/*informational*.

4.8 Wireframe Dashboard Aplikasi

Sebelum tahap implementasi antarmuka dilakukan, perancangan *wireframe* disusun untuk memetakan struktur halaman, alur interaksi pengguna, dan prioritas komponen informasi pada *dashboard* prediksi harga saham *RTX Corporation*. *Wireframe* berfungsi sebagai *blueprint visual* agar proses pengembangan *dashboard* berjalan sistematis dan sesuai dengan kebutuhan analisis pengguna.



Gambar 4.6 *Wireframe Dashboard Aplikasi*

Gambar 4.6 menampilkan rancangan kerangka antarmuka utama *dashboard* prediksi harga saham *RTX* sebelum tahap implementasi visual final. *Wireframe* ini memperlihatkan struktur halaman yang terdiri atas area sidebar pengaturan, bagian header informasi, panel ringkasan performa, tab navigasi, serta area grafik utama. Penyusunan elemen-elemen tersebut dirancang untuk memastikan alur penggunaan yang sistematis, yaitu dari pengaturan parameter prediksi, proses analisis, hingga interpretasi hasil model.

Gambar 4.7 *Wireframe* Prediksi Tanggal Prediksi

Gambar 4.7 memperlihatkan *wireframe* khusus untuk skenario prediksi berdasarkan tanggal target yang dipilih. Pada rancangan ini, komponen utama difokuskan pada input tanggal prediksi, parameter horizon waktu, serta panel output hasil berupa nilai prediksi dan perubahan harga yang diharapkan. Selain itu, *wireframe* juga menempatkan area visualisasi trajektori harga untuk memudahkan pembacaan arah pergerakan dari data historis ke nilai prediksi. Rancangan ini menunjukkan bahwa fitur prediksi tanggal tidak hanya menampilkan angka hasil estimasi, tetapi juga menyediakan konteks visual dan informasi pendukung agar hasil prediksi dapat dipahami secara lebih komprehensif.

4.9 Screenshot Dashboard Aplikasi



Gambar 4.8 Antarmuka Utama *Dashboard* Aplikasi Prediksi Harga Saham RTX

Gambar 4.8 *Dashboard* menampilkan prediksi harga saham RTX Corporation untuk 7 hari ke depan dengan visualisasi interaktif menggunakan *framework Streamlit*. Area utama menampilkan grafik prediksi yang menunjukkan: (1) data historis (garis biru solid) dari periode *training* yang dipilih pengguna, (2) prediksi harga (garis merah putus-putus) untuk horizon waktu yang ditentukan, dan (3) confidence interval (area biru muda transparan) yang merepresentasikan tingkat kepercayaan prediksi berdasarkan historical *volatility* dan model *uncertainty*. Bagian atas *dashboard* menampilkan *metrics cards* yang menyajikan informasi penting seperti harga penutupan terkini (*Current Price*), rentang prediksi minimum-maksimum (*Predicted Range*), dan tingkat kepercayaan model (*Confidence Level*) dalam bentuk persentase.

The dashboard sidebar contains several sections for configuring the model training and prediction process:

- PENGATURAN** (Settings) - Main header for the configuration section.
- Sumber Data** (Data Source) - Options to select the data source:
 - ☒ Yahoo Finance (Real-time)
 - ☐ File CSV Lokal
- Freeze Data** - A toggle switch to freeze the data, currently set to "Freeze Data".
- Tanggal Akhir Data (Freeze)** - A date picker set to "2026/01/14".
- Data dibekukan s.d. 14 Jan 2026** - A confirmation message indicating the data is frozen until January 14, 2026.
- Tanggal Prediksi** - A date picker set to "2026/01/21".
- Tanggal Target** - A date picker set to "2026/01/21".
- Hari ke Depan** - A slider set to "7 hari" (7 days).
- Rentang Data Pelatihan** (Training Data Range) - A section detailing the training data range:
 - DATA HISTORIS PENUH (1962 – Freeze) — SESUAI SKRIPSI
 - Seluruh data sejak 2 April 1962 digunakan
 - Konsisten dengan penelitian (15.846 records)
 - Split 85:15 → Training ~13.469 / Testing ~2.377
- Pelatihan dari: 1962-01-01 s.d. 2026-01-14 (23,389 hari)** - Training period information.
- Pengaturan Prakiraan** (Forecast Settings) - A section for forecast settings:
 - Horison Prakiraan (Hari)** - A slider set to "7".
 - [!] Toleransi Risiko** (Risk Tolerance) - A section for risk tolerance settings.
 - Tingkat Risiko** - A slider set to "Sedang" (Medium).
- MUAT ULANG DATA** (Reload Data) - A button to reload the data.
- INFO MODEL** (Model Info) - A section providing model details:
 - Algoritma: XGBoost
 - Target: % Perubahan Harga
 - Fitur: 43 Indikator Teknikal
 - Split: 85% Train / 15% Test
 - Data: Freeze s.d. 14 Jan 2026

Gambar 4.9 *Sidebar Dashboard*

Gambar 4.9 *Sidebar Dashboard* menampilkan kontrol yang memungkinkan pengguna untuk mengatur rentang data *training* (minimum 3 bulan hingga maksimum 14 tahun), memilih tanggal target prediksi melalui *date picker*, dan menentukan jumlah hari prediksi (1-30 hari) menggunakan *slider*:



Gambar 4.10 Visualisasi Indikator Teknikal pada *Dashboard*



Gambar 4.11 Visualisasi Indikator Teknikal pada *Dashboard (2)*

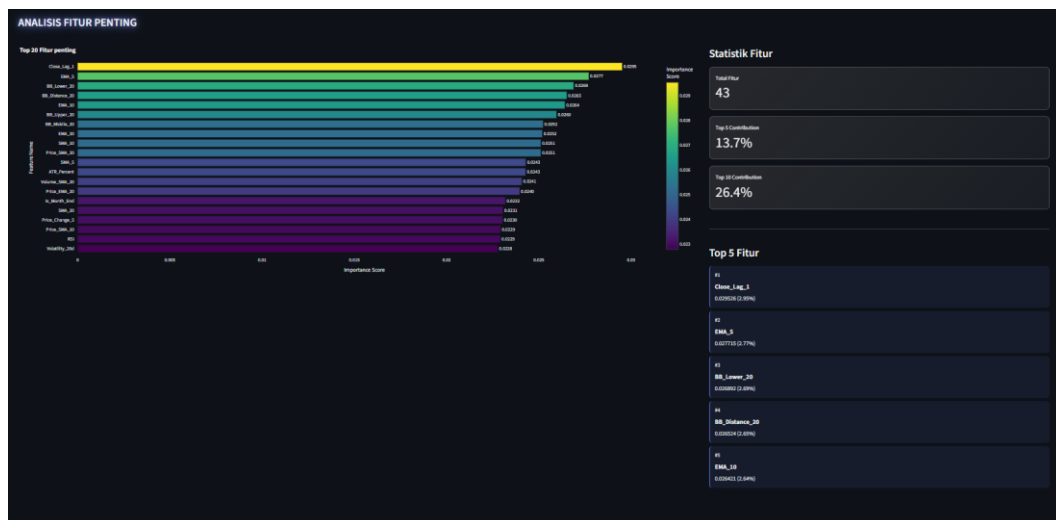
Gambar 4.10 dan Gambar 4.11 pada Tab Analisis Teknikal, menampilkan 43 indikator teknikal yang digunakan sebagai *features* dalam model prediksi, divisualisasikan dalam format *multi-panel* yang terorganisir. Panel utama (atas) menunjukkan *Price Chart with Moving Averages*, yang memplot harga penutupan (*candlestick*) bersama dengan *Simple Moving Average (SMA)* untuk periode 5, 10, dan 20 hari, serta *Exponential Moving Average (EMA)* untuk periode 5, 10, dan 20 hari. *Crossover* antara SMA jangka pendek dan jangka panjang (misalnya *SMA 5* memotong *SMA 20*) memberikan sinyal *trend reversal* yang penting untuk analisis.

Panel kedua menampilkan *Relative Strength Index (RSI)* dengan *threshold overbought* (>70 , garis merah putus-putus) dan *oversold* (<30 , garis hijau putus-putus). *RSI* mengukur *momentum* harga dalam skala 0-100, di mana nilai di atas 70 mengindikasikan kondisi *overbought* (potensi koreksi turun) dan nilai di bawah 30 mengindikasikan kondisi *oversold* (potensi *rebound* naik). Panel ketiga visualisasi *MACD (Moving Average Convergence Divergence)* yang terdiri dari: *MACD line* (biru), *Signal line* (oranye), dan *MACD Histogram* (bar hijau/merah). Perpotongan antara *MACD line* dan *Signal line* menghasilkan sinyal *bullish (cross upward)* atau *bearish (cross downward)*.

Panel keempat menampilkan *Bollinger Bands* dengan *upper band* (garis merah), *middle band/SMA 20* (garis kuning), dan *lower band* (garis hijau), disertai *shaded region* yang menunjukkan *volatility range*. Ketika harga menyentuh *upper band*, saham dianggap *overbought*; ketika menyentuh *lower band*, dianggap *oversold*. Setiap indikator dapat di-*toggle on/off* menggunakan *checkboxes* di sidebar untuk fokus pada analisis spesifik. *Dashboard* juga menampilkan *Volume Analysis* dan *Volatility Measures (ATR, historical volatility)* di panel tambahan yang dapat di-*scroll*.



Gambar 4.12 Evaluasi Performa Model pada *Testing Set*



Gambar 4.13 Evaluasi Performa Model pada *Testing Set* (2)

Gambar 4.12 dan 4.13 Tab Model Performance menyajikan evaluasi komprehensif dari model *XGBoost* yang telah dilatih. Bagian atas menampilkan 5 metrics cards utama dengan color-coded indicators: (1) *MAPE* (*Mean Absolute Percentage Error*): 1.24% dengan badge hijau "Sangat Baik", menunjukkan rata-rata *error* prediksi sekitar dari 1.24% dari harga aktual, (2) *R² Score* (*Coefficient of Determination*): 0.9972 dengan badge hijau "Luar Biasa", mengindikasikan bahwa model menjelaskan 99.72% *variance* dalam data harga saham, (3) *MAE* (*Mean Absolute Error*): \$0.99 dengan interpretasi "Rata-rata *error* per prediksi", (4) *RMSE* (*Root Mean Squared Error*): \$1.49, memberikan Ukuran *error* terjauh, dan (5) *Directional Accuracy*: 48.6% dengan badge biru "Baik", menunjukkan model memprediksi arah pergerakan harga (naik/turun) dengan benar pada 48.6% kasus.

Panel visualisasi menampilkan *Actual vs Predicted Prices Plot* yang memplot harga aktual (garis biru solid) dan prediksi model (garis merah putus-putus) pada *testing set* (15% dari total data). Kedekatan kedua garis menunjukkan akurasi prediksi yang sangat tinggi. Di samping line plot, terdapat scatter plot dengan ideal fit line (garis diagonal hitam) di mana titik-titik yang berdekatan dengan garis diagonal mengindikasikan prediksi akurat.

Panel bawah menampilkan *Residuals Plot* ($error = actual - predicted$) yang menunjukkan distribusi *error* secara temporal, dan *Error Distribution*

Histogram dengan *overlay* kurva normal (*statistical annotations: mean, std, skewness, kurtosis*). Histogram yang terpusat di sekitar 0 dengan distribusi simetris mengindikasikan bahwa *error* tidak memiliki systematic bias. *Dashboard* juga menyediakan *Feature Importance Chart* (*horizontal bar chart*) yang menunjukkan *top 20 features* berdasarkan *Gain score* dari *XGBoost*, membantu interpretasi faktor-faktor yang paling berpengaruh dalam prediksi.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Algoritma *XGBoost* terbukti sangat akurat dalam memprediksi harga saham *RTX Corporation* menggunakan 43 fitur teknis. Berdasarkan pengujian pada *testing set* (sebanyak 2.408 baris data atau 15% dari dataset), model menghasilkan *Mean Absolute Percentage Error (MAPE)* sebesar 1,2350%, yang berarti rata-rata galat prediksi hanya sekitar 1,24% dari harga aktual. Model ini juga mencatatkan skor R^2 sebesar 0,9972, *RMSE* sebesar \$1,50, dan *MAE* sebesar \$0,99.

Melalui proses *feature selection* berbasis skor *importance XGBoost*, diketahui bahwa tidak semua fitur memberikan kontribusi yang sama besar. Model sangat mengandalkan indikator tren, volatilitas, dan memori jangka pendek, di mana 5 fitur dengan kontribusi tertinggi secara berurutan adalah *EMA_5* (21,23%), *BB_Upper_20* (20,72%), *EMA_20* (19,87%), *Close_Lag_1* (19,82%), dan *BB_Distance_20* (18,36%).

Penggunaan subset 5 fitur terpilih tidak hanya menyederhanakan model dan menjaga efisiensi komputasi, tetapi juga terbukti sedikit meningkatkan performa prediksi. Pada skema *feature selection*, nilai *MAPE* berhasil ditekan menjadi 1,19% dengan R^2 naik menjadi 0,9974, *MAE* turun menjadi \$0,95, dan *RMSE* menjadi \$1,44.

Eksperimen integrasi fitur non-teknikal (sentimen berita dari *VADER*) ke dalam pemodelan kurang terbukti mampu menangkap reaktivitas pasar. Model *baseline* memiliki *Directional Accuracy* sebesar 48,6%, penggabungan skor sentimen pada model *hybrid* (dengan 39 baris data) masih belum bisa dikatakan mampu meningkatkan *Directional Accuracy* walaupun menjadi 83,3% dengan *MAPE* yang baik di angka 1,41%. Hal ini dikarenakan model hanya menggunakan 39 baris data dalam penelitian, maka dari itu hasil penelitian kurang berbobot.

5.2 Saran

Untuk pengembangan penelitian di masa mendatang, disarankan untuk membandingkan performa *XGBoost* dengan algoritma lain seperti *LSTM*, *LightGBM*, atau mengeksplorasi pendekatan *hybrid* yang mengkombinasikan model regresi untuk prediksi magnitude dengan model klasifikasi untuk prediksi arah guna meningkatkan *directional accuracy*.