

IMPLEMENTASI XGBOOST UNTUK PREDIKSI SAHAM

Muhamad Taufiq Kurniawan, Irfan Pratama

Sistem Informasi, Universitas Mercu Buana Yogyakarta
Jl. Jembatan Merah No. 84C, Gejayan, Yogyakarta, Indonesia
191210066@student.mercubuana-yogya.ac.id

ABSTRAK

Penerapan algoritma Machine Learning dalam prediksi saham telah berkembang pesat, namun literatur yang ada umumnya berfokus pada saham berkapitalisasi besar (Blue Chip) dengan tren yang relatif stabil. Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma XGBoost pada objek penelitian dengan karakteristik berbeda, yaitu saham teknologi (GOTO) yang memiliki volatilitas tinggi dan data historis terbatas pasca-IPO. Penelitian ini berupaya menjawab tantangan overfitting yang sering terjadi pada data fluktuatif melalui pendekatan optimasi model. Metodologi penelitian melibatkan pra-pemrosesan data dengan penambahan indikator teknikal (RSI, MACD, MA) serta dekomposisi fitur waktu. Proses pemodelan dilakukan secara komparatif antara parameter default dan parameter hasil optimasi menggunakan Grid Search CV. Hasil pengujian menunjukkan perbedaan performa yang signifikan, di mana model dengan parameter standar menghasilkan tingkat kesalahan (MAPE) sebesar 36,419%. Namun, setelah dilakukan penyesuaian hyperparameter (Learning rate: 0.2, Max Depth: 3, n_Estimators: 200), akurasi model meningkat drastis dengan penurunan MAPE menjadi 1,957%, MAE 1,215, dan RMSE 1,505. Berdasarkan hasil tersebut, disimpulkan bahwa XGBoost memiliki kapabilitas yang tangguh (robust) dalam memprediksi saham bervolatilitas tinggi, dengan catatan bahwa penerapan Hyperparameter Tuning merupakan langkah krusial untuk memastikan model dapat menangkap pola tren secara akurat di tengah fluktuasi pasar.

Kata kunci : GOTO, hyperparameter tuning, prediksi saham, volatilitas tinggi, XGBoost.

1. PENDAHULUAN

Kegiatan perdagangan saham adalah bagian integral dalam dunia keuangan global, dan untuk investor, pedagang, dan perusahaan, pemahaman yang mendalam tentang pergerakan harga saham memiliki nilai strategis yang besar. Pada dasarnya, pergerakan saham dipengaruhi oleh berbagai faktor, termasuk data ekonomi, politik, dan sosial. Saham sendiri merupakan salah satu bukti kepemilikan saham sebagai investor juga sebagai bagian kepemilikan perusahaan [1]. Analisis pergerakan saham yang memadai memerlukan pemahaman mendalam tentang interaksi variabel-variabelnya. Seperti kinerja perusahaan, sentiment pasar, ekonomi makro, supply, demand, dan lainlain yang seringkali rumit dan tidak linear. Salah satu contohnya adalah saham GOTO. Kondisi pergerakan saham dari GOTO saat melantai di pasar modal terjadi lonjakan yang signifikan. Tingginya minat pasar membuat pesanan saham GOTO mengalami kelebihan permintaan. Namun, dua pekan setelah perdagangan, terjadi penurunan pada harga saham GOTO bahkan berada jauh di bawah harga Initial Public Offering. Investor dilanda kepanikan yang menjadi pemicu investor melakukan aksi jual secara besar-besaran sehingga menjadi pemicu terjadinya penurunan harga saham [2].

Dengan meningkatnya pemanfaatan machine learning dalam analisis keuangan, metode seperti XGBoost telah menjadi populer untuk memprediksi. XGBoost merupakan model dari algoritma tree gradient boosting yang memiliki kecepatan optimasi sepuluh kali lipat dibandingkan dengan metode gradient boosting lainnya [3]. Beberapa penelitian

sebelumnya telah membuktikan bahwa algoritma machine learning dapat menghasilkan prediksi yang lebih akurat dibandingkan metode tradisional. Penelitian lain menunjukkan bahwa kombinasi model machine learning dengan analisis teknikal dapat meningkatkan keakuratan prediksi saham secara signifikan [4]. Selain itu, menurut penelitian pendekatan ensemble berbasis gradient boosting yang mampu meningkatkan prediksi saham dengan mempertimbangkan faktor makroekonomi [5].

Pada penelitian ini, peneliti akan menggunakan jenis algoritma time series, yaitu XGBoost (extreme Gradient Boosting). XGBoost merupakan algoritma yang ditingkatkan dari gradient boosting, gradient boosting adalah salah satu algoritme yang digunakan untuk membangun model prediktif dalam pembelajaran mesin. Algoritme ini bekerja dengan cara menggabungkan beberapa model prediktif yang lemah menjadi satu model prediktif yang kuat. Setiap model prediktif lemah digunakan untuk memperbaiki kesalahan prediksi model sebelumnya. Saat ditingkatkan ke XGBoost ada beberapa proses yang ditambahkan yaitu pemangkasan, newton boosting, dan parameter pengacakan ekstra. Proses pemangkasan atau penyusutan proporsional simpul daun digunakan untuk meningkatkan generalisasi model. proses newton boosting adalah proses untuk menyediakan rute langsung sehingga tidak memerlukan penurunan gradient. Proses pengacakan parameter bertujuan untuk mengurangi korelasi antar tree sehingga dapat meningkatkan kekuatan algoritma ensemble. Penggunaan metode ensemble seperti XGBoost memiliki performa lebih baik dalam

menangani data saham yang memiliki fluktuasi tinggi dibandingkan dengan model regresi linier biasa. Implementasi XGBoost dalam analisis pasar saham memberikan peningkatan signifikan dibandingkan dengan model tradisional dalam mengatasi noise pada data keuangan [6]. Penelitian lain menegaskan bahwa penggunaan teknik boosting dalam machine learning dapat meningkatkan stabilitas prediksi harga saham pada berbagai sektor industri [7].

Meskipun algoritma Machine Learning seperti XGBoost telah terbukti handal dalam prediksi finansial, tinjauan literatur menunjukkan adanya kesenjangan (research gap) pada objek penelitian di Indonesia. Mayoritas studi terdahulu cenderung berfokus pada saham sektor perbankan dan Blue Chip yang memiliki karakteristik volatilitas rendah serta tren historis yang panjang dan stabil, sebagaimana terlihat pada penelitian prediksi saham BCA [3], BBRI [9], maupun indeks LQ45 [6]. Hingga saat ini, belum banyak literatur yang secara spesifik menguji ketahanan (robustness) algoritma XGBoost ketika dihadapkan pada saham sektor teknologi baru (seperti GOTO). Saham ini memiliki karakteristik unik berupa volatilitas ekstrem (high volatility) dan data historis yang terbatas (short-term data) pasca-IPO. Oleh karena itu, penelitian ini menawarkan kebaruan (novelty) dengan melakukan evaluasi empiris terhadap efektivitas Hyperparameter Tuning pada XGBoost dalam menangani data yang mengandung noise tinggi. Penelitian ini berkontribusi dalam membuktikan apakah mekanisme regularisasi pada XGBoost mampu meredam fluktuasi drastis pada saham teknologi, yang mana aspek ini belum terakomodasi dalam penelitian-penelitian terdahulu yang berfokus pada saham stabil.

2. TINJAUAN PUSTAKA

Landasan Prediksi harga saham merupakan proses untuk memperkirakan nilai suatu saham pada periode mendatang dengan memanfaatkan data historis yang tersedia. Proses ini tergolong sulit karena pergerakan harga saham bersifat fluktuatif dan tidak dapat dipastikan arahnya. Perubahan harga saham dipengaruhi oleh berbagai faktor, mulai dari kondisi ekonomi makro, kinerja perusahaan, hingga sentimen pasar yang berkembang di masyarakat. Selain itu, data saham umumnya memiliki karakteristik nonlinier serta mengandung noise, sehingga jika hanya dianalisis menggunakan metode konvensional, hasil prediksi yang diperoleh sering kali kurang optimal [8], [9]. Oleh karena itu, diperlukan pendekatan yang mampu menangkap pola-pola kompleks yang terdapat dalam data saham.

Perkembangan teknologi komputasi mendorong penggunaan machine learning dalam bidang prediksi harga saham. Machine learning memungkinkan sistem untuk mempelajari pola dari data secara otomatis tanpa harus menetapkan aturan secara eksplisit. Dalam penelitian prediksi saham, pendekatan supervised learning banyak digunakan karena data historis harga saham tersedia secara lengkap dan memiliki label yang

jas. Sejumlah penelitian menunjukkan bahwa metode machine learning mampu memberikan hasil prediksi yang lebih baik dibandingkan metode statistik tradisional, terutama dalam menghadapi data yang dinamis serta berpola tidak linier [10], [11].

Salah satu algoritma machine learning yang cukup banyak digunakan dalam prediksi harga saham adalah XGBoost (Extreme Gradient Boosting). XGBoost merupakan algoritma yang termasuk ke dalam metode ensemble learning, di mana beberapa decision tree dibangun secara bertahap untuk memperbaiki kesalahan prediksi sebelumnya. Algoritma ini dirancang agar memiliki efisiensi komputasi yang baik serta dilengkapi dengan mekanisme regularisasi guna mengurangi risiko terjadinya overfitting [8]. Beberapa penelitian menunjukkan bahwa XGBoost mampu menghasilkan performa prediksi yang lebih unggul dibandingkan algoritma lain seperti Random Forest dan Support Vector Machine dalam kasus prediksi harga saham [12],[10].

Data saham yang digunakan dalam penelitian umumnya berbentuk data deret waktu atau time series. Data ini mencakup harga pembukaan, harga tertinggi, harga terendah, harga penutupan, serta volume transaksi. Salah satu ciri utama data time series adalah adanya keterkaitan antar waktu, di mana nilai pada suatu periode dipengaruhi oleh nilai pada periode sebelumnya. Kondisi tersebut menuntut penggunaan metode analisis yang mampu memahami pola berdasarkan urutan waktu agar hasil prediksi yang diperoleh menjadi lebih akurat [13].

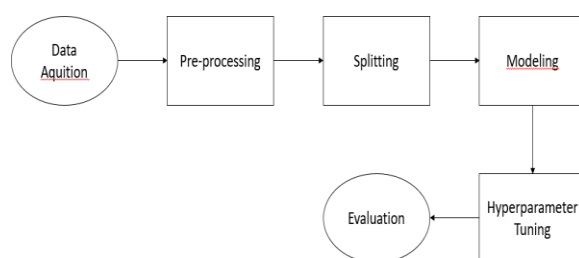
Untuk meningkatkan kinerja model prediksi, tahap feature engineering menjadi bagian yang penting dalam proses penelitian. Feature engineering merupakan proses mengolah data mentah menjadi fitur yang lebih informatif dan relevan dengan tujuan prediksi. Dalam prediksi harga saham, indikator teknikal seperti Moving Average (MA), Relative Strength Index (RSI), dan Moving Average Convergence Divergence (MACD) sering digunakan karena indikator-indikator tersebut dapat membantu model dalam mengenali pola pergerakan harga saham [14]. Beberapa penelitian juga menyebutkan bahwa pemilihan fitur yang tepat dapat memberikan peningkatan yang signifikan terhadap akurasi model prediksi [12].

Berdasarkan penelitian-penelitian terdahulu, penerapan algoritma XGBoost dalam prediksi harga saham menunjukkan hasil yang cukup menjanjikan. Penelitian yang dilakukan oleh Zhang et al. menunjukkan bahwa penggunaan XGBoost yang dikombinasikan dengan indikator teknikal mampu meningkatkan akurasi prediksi harga saham [12]. Penelitian lain juga menyatakan bahwa metode ensemble learning cenderung menghasilkan performa yang lebih stabil dibandingkan metode tunggal dalam memprediksi pergerakan pasar saham [15], [10], [11]. Temuan-

temuan tersebut menjadi dasar dalam pemilihan algoritma XGBoost sebagai metode yang digunakan dalam penelitian ini..

3. METODE PENELITIAN

Penelitian ini menerapkan metodologi Waterfall guna menjamin alur pengembangan model yang sistematis dan terukur, di mana setiap tahapan dari pra-pemrosesan hingga evaluasi diselesaikan secara linear untuk menjaga konsistensi data time series. Pendekatan ini dipilih karena lebih relevan untuk studi eksperimental komparatif dibandingkan siklus iteratif CRISP-DM, mengingat fokus utama penelitian adalah pembuktian validitas teknis algoritma XGBoost secara bertahap. Metodologi penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Metodologi penelitian

3.1. Data Acquisition

Peneliti menggunakan sumber data untuk mendukung penyelesaian penelitian ini. Data sekunder didapatkan melalui website finance.yahoo.com. Dari website tersebut didapatkan harga *Date*, *open*, *high*, *low*, *close*, *close(adj)* dan jumlah volume pembelannya. Adapun data yang diambil dari hari terdapatnya saham pada bursa saham yaitu 11 April 2022 sampai 29 Januari 2026. Mengingat fokus penelitian adalah harga penutupan (*Close price*).

3.2. Preprocessing

Sebelum pemodelan, dilakukan seleksi fitur (feature selection) untuk mereduksi dimensi data. Atribut *close (adj)* dieliminasi karena dianggap redundan atau tidak digunakan sebagai variabel prediktor dalam penelitian ini. Selanjutnya dilakukan proses data cleaning atau menghapus data kosong maupun data ganda. Penelitian ini memperkaya dataset dengan indikator teknikal yang dihitung secara

komputasi, meliputi Moving Averages (MA), Relative Strength Index (RSI), dan Moving Average Convergence Divergence (MACD). Penambahan fitur ini bertujuan agar algoritma XGBoost dapat mempelajari pola tidak hanya dari harga, tetapi juga dari momentum dan tren pergerakan.

Selain itu, dilakukan standarisasi format waktu pada kolom date untuk memastikan data terbaca sebagai deret waktu (time series). Variabel target ditetapkan sebagai harga penutupan (*Close*), sedangkan variabel prediktor adalah fitur yang tersisa setelah penghapusan kolom *close (adj)*. Terakhir, dilakukan ekstraksi fitur temporal (temporal

decomposition). Variabel tanggal didekomposisi menjadi atribut numerik terpisah yang meliputi Tahun, Bulan, Hari, dan Hari dalam Seminggu (*Day of Week*). Langkah ini bertujuan agar model Machine Learning dapat menangkap pola musiman (seasonality) dan siklus mingguan.

3.3. Splitting Data

Untuk memastikan keabsahan model, langkah selanjutnya adalah membagi dataset menjadi dua bagian terpisah data latih (*training data*) dan data uji (*testing data*). Tahap ini merupakan prosedur fundamental dalam pengembangan model *machine learning* untuk mengevaluasi kinerja model secara objektif. Tujuan utama dari pembagian data adalah untuk melatih model pada satu set data (data latih) dan kemudian menguji seberapa baik model tersebut dapat melakukan generalisasi pada data baru yang belum pernah dilihat sebelumnya (data uji). Untuk data deret waktu (*time series*) seperti harga Ethereum, pembagian data tidak boleh dilakukan secara acak. Pembagian harus dilakukan secara kronologis untuk menjaga dependensi temporal data. Hal ini mensimulasikan skenario dunia nyata di mana kita menggunakan data masa lalu untuk memprediksi data di masa depan. Dalam penelitian ini, rasio pembagian yang digunakan adalah 80:20, dengan rincian yaitu 80% sebagai data latih (*training data*) dan 20% sebagai data uji (*testing data*).

3.4. Modelling

Model prediksi dibangun menggunakan algoritma eXtreme Gradient Boosting (XGBoost). XGBoost bekerja dengan prinsip boosting, yaitu membangun ansambel pohon keputusan secara berurutan, di mana setiap pohon baru berusaha memperbaiki kesalahan (residual) dari pohon sebelumnya. Secara matematis, XGBoost berusaha meminimalkan Fungsi Objektif reguler berikut pada setiap iterasi

$$Obj^{(t)} = \sum_{i=1}^n \left(y_i, y_i^{t-1} + f_t(x_i) \right) + \Omega(f_t)$$

Di mana:

- L adalah fungsi kerugian (loss function), dalam penelitian ini digunakan Squared Error (reg:squarederror) untuk mengukur selisih antara nilai aktual y_i dan prediksi.
- $\Omega(f_t)$ adalah suku regularisasi (regularization term) yang mengontrol kompleksitas model untuk mencegah overfitting. Suku ini mencakup parameter seperti α (L1 regularization) yang digunakan dalam model awal].

3.5. Hyperparameter Tuning

Untuk meningkatkan kinerja model melampaui baseline, dilakukan proses optimasi sistematis menggunakan Grid Search Cross-Validation (GridSearchCV) dengan skema validasi silang 3-lipatan (3-fold cv). Ruang pencarian parameter

didefinisikan untuk mengeksplorasi kombinasi optimal dapat dilihat pada tabel 1.

Tabel 1. Parameter GridSearchCV

Parameter	Nilai
Learning rate	0.05, 0.1, 0.2
Max depth	3, 5, 7
N_estimators	50, 100, 200

3.6. Evaluation

Pada tahap akhir, model dengan konfigurasi parameter terbaik (best estimator) yang diperoleh dari GridSearchCV dievaluasi kembali menggunakan data uji sekuensial ($X_{(test)}$) yang belum pernah tersentuh selama proses optimasi. Kinerja model final diukur menggunakan metrik Optimized MAE,

Optimized RMSE, dan Optimized MAPE. Hasil ini kemudian dibandingkan dengan nilai evaluasi model awal untuk mengkuantifikasi peningkatan akurasi yang berhasil dicapai melalui proses tuning. Selain analisis kuantitatif, dilakukan inspeksi visual terhadap tabel komparasi antara nilai aktual dan nilai prediksi pada sampel data uji untuk memastikan model menangkap tren dengan baik. Secara matematis, MAE, RMSE, dan MAPE menilai tingkat error dari model berikut iterasinya:

$$\text{MAE: } \frac{1}{n} \sum_{i=1}^n |y_i - y'_i|$$

$$\text{RMSE: } \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}}$$

$$\text{MAPE: } \frac{1}{n} \sum_{i=1}^n \frac{|y_i - y'_i|}{y'_i} \times 100\%$$

Sebagai langkah terakhir menuju implementasi praktis (deployment), model terbaik diekspor dan disimpan dalam format biner (.xgb), memungkinkan model tersebut digunakan kembali untuk prediksi di masa depan tanpa memerlukan pelatihan ulang.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan dan pra-pemrosesan Data

Setelah melakukan observasi terhadap pasar saham, peneliti memutuskan saham GOTO sebagai sampel penelitian. Saham GOTO dipilih karena memiliki fluktuasi harga yang cukup kontras, hal tersebut dapat memberikan lebih banyak variasi untuk model belajar pola dari data, terutama jika perubahan harga disebabkan oleh faktor yang dapat diprediksi (misalnya berita ekonomi, perubahan kebijakan, dll). Berikut contoh sampel data mentah dapat dilihat pada tabel 2.

Tabel 2. Sampel data

No	Date	Close	High	Low	Open	Volume
1	11/04/2022	382	416	372	400	9410897000
2	12/04/2022	370	442	360	422	3887331000
3	13/04/2022	374	380	360	370	3262811400
4	14/04/2022	376	382	374	374	3675981900
5	18/04/2022	378	380	370	376	2660312700
6	19/04/2022	358	380	358	378	2252971800
7	20/04/2022	338	364	338	358	5804281200
8	21/04/2022	340	340	336	340	1670584600
9	22/04/2022	340	340	330	340	6075753400
10	25/04/2022	328	338	318	318	1960587900
...						
...						
997	26/01/2026	60	62	59	60	4896210800
998	27/01/2026	65	65	58	60	12307760600
999	28/01/2026	60	64	58	62	10963550000
910	29/01/2026	63	63	55	60	15220468500

Tabel 3. Hasil Pra-pemrosesan data

No	Close	High	Low	Volume	Open	MA5	MA10	MA20	RSI	MACD	Signal Line	Year	Month	Day	Day of week
1	382	416	372	9410897000	400	-	-	-	-	-	-	2022	4	11	0
2	370	442	360	3887331000	422	-	-	-	0,0	-0,9	-0,19	2022	4	12	1
3	374	380	360	3262811400	370	-	-	-	25,0	-1,4	-0,43	2022	4	13	2
4	376	382	374	3675981900	374	-	-	-	33,3	-1,5	-0,65	2022	4	17	3
5	378	380	370	2660312700	376	376,0	-	-	40,0	-1,5	-0,81	2022	4	18	0
6	358	380	358	2252971800	378	371,2	-	-	20,0	-3,0	-1,25	2022	4	19	1
7	338	364	338	5804281200	358	364,8	-	-	13,3	-5,7	-2,16	2022	4	20	2
8	340	340	336	1670584600	340	358,0	-	-	16,1	-7,7	-3,27	2022	4	23	3
9	340	340	330	6075753400	340	350,8	-	-	16,1	-9,1	-4,44	2022	4	24	4
10	328	338	318	1960587900	318	340,8	358,4	-	13,5	-11,1	-5,78	2022	4	25	0
...															
...															

No	Close	High	Low	Volume	Open	MA5	MA10	MA20	RSI	MACD	Signal Line	Year	Month	Day	Day of week
997	60	62	59	4896210800	60	62,4	64,8	65,45	21,4	-0,86	0,04	2026	1	26	0
998	65	65	58	12307760600	60	62,2	64,7	65,5	42,1	-0,71	-0,11	2026	1	27	1
999	60	64	58	10963550000	62	61,4	64,0	65,25	34,8	-0,99	-0,28	2026	1	28	2
910	63	63	55	15220468500	60	61,6	63,4	65,20	44,0	-0,96	-0,42	2026	1	29	3

Setelah melalui tahap pra-pemrosesan seperti feature selection, penambahan indikator teknikal, standarisasi format waktu, serta ekstraksi fitur (*Date*). Data yang akan digunakan seperti yang dapat dilihat pada Tabel 3. Dengan total 910 baris data selanjutnya, data dibagi menggunakan skema hold-out dengan rasio 80:20. Sebanyak 80% atau sebanyak 728 baris data dialokasikan untuk proses pelatihan (training), sedangkan 20% sisanya atau 182 baris data digunakan sebagai data uji (testing), baik data training maupun testing dibagi secara sekuensial untuk mensimulasikan skenario dunia nyata di mana kita menggunakan data masa lalu untuk memprediksi data di masa depan.

4.2. Pemodelan dan optimasi model

Parameter default digunakan pada tahap awal pemodelan, dengan nilai Learning rate :0.1 , Max Depth :5 , dan n_Estimators :10. Untuk menjawab permasalahan terkait akurasi prediksi, penelitian ini tidak hanya menerapkan model XGBoost dengan parameter *default*, melainkan melakukan optimasi menggunakan metode *Grid Search Cross-Validation* (GridSearchCV) dengan skema 3-fold cross validation. Proses ini bertujuan mencari kombinasi parameter terbaik (best estimator) yang menghasilkan nilai error terendah. Hasilnya, parameter terbaik berisi nilai Learning rate: 0.2, Max Depth: 3, dan n_Estimators: 200.

4.3. Evaluasi

Evaluasi performa model dilakukan dengan membandingkan hasil prediksi (y_{pred}) terhadap data aktual (y_{actual}) pada data uji yang belum pernah dilihat model sebelumnya. Pengukuran menggunakan tiga metrik standar evaluasi regresi, yaitu MAE, RMSE, dan (MAPE).

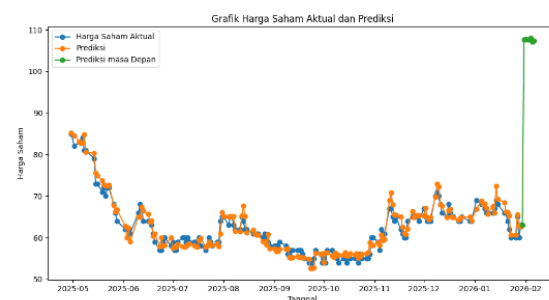
Tabel 4. Parameter XGBoost

Evaluasi model	Nilai Error	
	Model awal	Model dioptimasi
MAE	22.349	1.215
RMSE	22.467	1.505
MAPE	36.419	1.957

Tabel 4 menunjukkan nilai evaluasi model menggunakan parameter default. Setelah dilakukan GridSearchCV dan parameter terbaik diimplementasikan pada model.

Berdasarkan hasil pengujian pada Tabel 4, terdapat temuan ilmiah yang signifikan terkait perilaku algoritma XGBoost pada data volatil. Penggunaan parameter default menghasilkan tingkat kesalahan

yang sangat tinggi dengan MAPE mencapai 36,419%. Hal ini mengindikasikan bahwa model standar gagal menangkap pola tren pada saham GOTO dan cenderung mengalami underfitting akibat ketidakmampuan algoritma menyesuaikan diri dengan fluktuasi harga yang ekstrem. Sebaliknya, setelah dilakukan optimasi hyperparameter (dengan Learning rate: 0.2, Max Depth: 3, dan n_Estimators: 200) , performa model meningkat drastis dengan penurunan MAPE menjadi 1,957%. Selisih performa yang signifikan antara model default dan optimized ini menjadi bukti empiris bahwa pada kasus saham teknologi dengan volatilitas tinggi, penerapan XGBoost tidak dapat dilakukan secara plug-and-play. Temuan ini menjawab celah penelitian (gap) bahwa karakteristik data GOTO yang fluktuatif (non-linier) memerlukan penanganan khusus melalui penyetelan parameter kedalaman pohon (max depth) untuk mencegah model menangkap noise sebagai pola (overfitting). Dengan demikian, penelitian ini berhasil membuktikan bahwa XGBoost yang dioptimasi memiliki ketahanan (robustness) yang baik bahkan pada instrumen investasi yang berisiko tinggi, melengkapi studi terdahulu yang mayoritas hanya menguji pada saham stabil.



Gambar 2. Grafik harga actual vs prediksi

Visualisasi perbandingan antara harga aktual dan hasil prediksi dapat dilihat pada Gambar 2. Grafik menunjukkan bahwa garis prediksi (warna oranye) mampu mengikuti tren garis aktual (warna biru) dengan sangat rapat, bahkan pada saat terjadi lonjakan harga (spikes) maupun penurunan tren. Serta garis Hijau yang menunjukkan hasil prediksi dimasa depan (diluar dataset)

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil membuktikan bahwa algoritma XGBoost memiliki kapabilitas untuk menangani karakteristik data saham GOTO yang memiliki volatilitas tinggi (High Volatility). Hal ini dibuktikan secara empiris melalui penurunan tingkat

kesalahan yang drastis, di mana model yang dioptimasi mampu mencapai akurasi tinggi (MAPE 1,957%) dibandingkan model standar yang mengalami kegagalan prediksi (MAPE 36,419%). Penerapan teknik Feature Engineering (penambahan indikator teknikal dan dekomposisi waktu) serta Hyperparameter Tuning terbukti menjadi langkah krusial. Penelitian menyimpulkan bahwa pada kasus saham teknologi baru, algoritma XGBoost tidak dapat diterapkan secara instan (plug-and-play), melainkan memerlukan pembatasan kompleksitas model (terutama pada parameter Max Depth) untuk menghindari noise pasar. Hasil evaluasi menunjukkan bahwa model yang dikembangkan memiliki performa yang cukup baik dalam menangani data saham dengan fluktuasi tinggi. Namun, ada batasan dalam sejauh apa memprediksi saham di masa depan sesuai dengan berapa banyak data yang ada di dataset. Diharapkan penelitian selanjutnya dapat membandingkan metode XGBoost dengan algoritma time series lainnya seperti Prophet atau Transformer-based model untuk meningkatkan akurasi prediksi serta mengatasi batasan prediksi di masa depan. Selain itu, penggabungan data makroekonomi dan analisis sentimen dapat menjadi pendekatan yang menarik untuk penelitian mendatang.

DAFTAR PUSTAKA

- [1] A. W. Mulyawan, "Perdagangan Saham di Bursa Efek : Konsep , Mekanisme , dan Perspektif Islam Stock Trading on the Stock Exchange : Concepts , Mechanisms and Islamic Perspectives," vol. 2, no. 2, pp. 68–90, 2025.
- [2] W. S. Wudjud, M. Butar-butur, S. A. Sitohang, and A. Sinaga, "Gambaran Kinerja Keuangan dan Pergerakan Harga Saham PT Goto Gojek Tokopedia Tbk," vol. 4, pp. 7581–7595, 2024.
- [3] S. T. Nurmaya, B. Soesilo, and D. A. Fatah, "prediksi harga saham pada bank bca tbk dengan teknik xgboost (extreme gradient boosting)," vol. 12, no. 2 2023.
- [4] Q. Chen, W. Zhang, and Y. U. Lou, "Forecasting Stock Prices Using a Hybrid Deep Learning Model Integrating Attention Mechanism , Multi-Layer Perceptron , and Bidirectional Long-Short Term Memory Neural Network," pp. 117365–117376, 2020, doi: 10.1109/ACCESS.2020.3004284.
- [5] N. T. Nurlela, F. R. Umbara, and F. Kasyidi, "Prediksi Return Saham Index LQ45 Menggunakan Metode Extreme Gradient Boosting (XGBOOST)," pp. 272–277, 2024
- [6] A. F. Alfa, D. A. Prasetya, P. Korespondensi, S. V. Regression, and B. Optimization, "implementasi xgboost dan support vector regression dalam implementation of xgboost and support vector regression in," vol. 12, no. 6, pp. 1443–1450, 2025.
- [7] A. Baradja and T. I. Tjendrowasono, "Penerapan Machine Learning untuk Meningkatkan Prediksi Mata Uang Forex dengan Indikator Teknikal," vol. 2, no. 1, pp. 38–41, 2023.
- [8] V. Y. Bkv and K. P. Shailaja, "Prediction of Stock Price Using XG Boost : A," vol. 12, no. 7 July, 2024.
- [9] N. Selayanti, D. A. Putri, and M. Idhom, "prediksi harga penutupan saham bbri dengan model hybrid lstm-xgboost," vol. 5, no. 1, pp. 52–64, 2025.
- [10] O. Guennioui, D. Chiadmi, and M. Amghar, "Improving Global Stock Market Prediction with XGBoost and LightGBM Machine Learning Models," no. 212, pp. 2603–2610, 2023.
- [11] A. F. Alkayes and T. Sugihartono, "Perbandingan Algoritma XGBoost dan LSTM dalam Prediksi Harga Saham Tesla Menggunakan Data Tahun 2025 Abstrak Comparison of XGBoost and LSTM Algorithms in Tesla Stock Price Prediction Using 2025 Data Abstract," vol. 5, no. 6, pp. 1563–1573, 2025.
- [12] A. Performa and A. Xgboost, "Performance Analysis of XGBoost Algorithm to Determine the Most Optimal Parameters and Features in Predicting Stock Price Movement," vol. 20, no. 1, pp. 91–102, 2023, doi: 10.31515/telematika.v20i1.9329.
- [13] P. H. Vuong, T. T. Dat, T. K. Mai, P. H. Uyen, and P. T. Bao, "Stock-Price Forecasting Based on XGBoost and LSTM," vol. 40, no. 1 2022, doi: 10.32604/csse.2022.017685.
- [14] T. Astutiningsih, D. Retno, and S. Saputro, "SPECTA Journal of Technology," vol. 7, no. 3, pp. 632–641, 2023, doi: 10.35718/specta.v7i3.1031.
- [15] J. Yao, "Stock Prediction of Google based on ARIMA , XGBoost and," vol. 44, pp. 414–421, 2023.