

Perbandingan Prediksi Harga Closing Saham BCA Menggunakan ADABOOST, XGBoost, CATBoost

Jason Wiedardi Limtara ¹⁾ Wilson Alfando ²⁾ Ricky Cangniago ³⁾ Aurellius Jeremiah Thomas ⁴⁾ Teny Handhayani ⁵⁾

¹⁾ Teknik Informatika Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta 11440 Indonesia
email : jason.535220212@stu.untar.ac.id

²⁾ Teknik Informatika Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta 11440 Indonesia
email : wilson.535220219@stu.untar.ac.id

³⁾ Teknik Informatika Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta 11440 Indonesia
email : ricky.535220210@stu.untar.ac.id

⁴⁾ Teknik Informatika Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta 11440 Indonesia
email : aurellius.535220213@stu.untar.ac.id

⁵⁾ Teknik Informatika Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta 11440 Indonesia
email : tenyh@fti.untar.ac.id

ABSTRAK

Penelitian ini membandingkan kinerja tiga algoritma boosting, yaitu AdaBoost, XGBoost, dan CatBoost, dalam memprediksi harga penutupan saham Bank Central Asia (BBCA). Tujuan dari penelitian ini adalah untuk menentukan algoritma mana yang memberikan model paling akurat dan efisien dalam hal akurasi prediksi, waktu komputasi, dan pemanfaatan sumber daya. Dataset yang digunakan terdiri dari data historis saham BCA, termasuk fitur seperti opening price, high, low, volume, dan previous closing prices. Metodologi penelitian meliputi data preprocessing, feature selection, pembagian data menjadi training set dan testing set, serta pelatihan model menggunakan masing-masing algoritma dengan hyperparameter tuning. Hasil eksperimen menunjukkan bahwa AdaBoost mencapai kinerja keseluruhan terbaik, memberikan akurasi prediksi tertinggi dan hasil yang stabil dengan waktu komputasi yang relatif lebih rendah dibandingkan dengan XGBoost dan CatBoost.

Keywords

AdaBoost, BBKA, CatBoost, Price Prediction, XGBoost.

1. Pendahuluan

Prediksi harga saham merupakan salah satu area yang paling menantang dan diminati dalam analisis keuangan [1]. Hal ini disebabkan oleh sifat pasar saham yang sangat non-linier, dinamis, dan rentan terhadap berbagai faktor eksternal yang sulit diprediksi, seperti kondisi ekonomi makro, kebijakan pemerintah, sentimen pasar, dan berita global [2]. Keakuratan dalam memprediksi pergerakan

harga saham dapat memberikan keuntungan kompetitif yang signifikan, membantu investor untuk mengambil keputusan strategis, mengoptimalkan keuntungan, dan yang terpenting, memitigasi risiko investasi [3], [4].

Seiring dengan kemajuan pesat di bidang teknologi dan komputasi, pemanfaatan metode *machine learning* untuk peramalan pasar finansial telah menarik minat yang signifikan dari para peneliti dan praktisi [5], [6]. Algoritma *machine learning* terbukti mampu menangani pola data yang sangat kompleks dan memberikan hasil prediksi yang lebih akurat dibandingkan metode statistik tradisional seperti ARIMA atau GARCH [2], [3]. Kemampuannya dalam mengenali pola-pola tersembunyi dan hubungan non-linier dari data historis menjadikannya alat yang ampuh untuk pemodelan finansial [2], [7].

Maksud dan tujuan dari penelitian ini adalah untuk melakukan analisis komparatif yang mendalam terhadap performa tiga algoritma *ensemble learning* dari keluarga *boosting*, yaitu AdaBoost, XGBoost, dan CatBoost [4], [8]. Studi kasus yang diangkat adalah prediksi harga penutupan (*closing price*) saham PT Bank Central Asia Tbk (BBKA), yang merupakan salah satu saham dengan kapitalisasi pasar terbesar dan paling likuid di Bursa Efek Indonesia [9]. Tujuan utamanya adalah untuk mengidentifikasi algoritma mana yang memberikan tingkat akurasi prediksi tertinggi, efisiensi komputasi terbaik, dan stabilitas model yang paling andal untuk data saham BBKA.

Beberapa penelitian sebelumnya telah menunjukkan keunggulan algoritma berbasis *boosting*. XGBoost, misalnya, telah terbukti sangat efektif untuk prediksi

harga saham dan seringkali menunjukkan performa yang kuat dan konsisten di berbagai pasar [8], [10]. Sebuah studi tentang prediksi harga saham pasca-*stock split* juga menunjukkan bahwa XGBoost mampu memberikan hasil yang sangat kompetitif dibandingkan dengan metode lain seperti Regresi Linier dan Support Vector Regression (SVR) [11]. Penelitian lain juga mengonfirmasi bahwa model *boosting* seperti XGBoost seringkali lebih unggul dari model tunggal dalam tugas prediksi finansial [12]. Sementara itu, CatBoost, yang dirancang khusus untuk menangani fitur kategorikal secara efisien, juga menunjukkan potensi besar dalam berbagai tugas prediksi, termasuk di sektor keuangan [13].

Meskipun algoritma boosting seperti AdaBoost, XGBoost, dan CatBoost telah digunakan secara luas dalam berbagai aplikasi prediktif, namun literatur menunjukkan bahwa masih terbatas penelitian yang secara langsung membandingkan ketiganya dalam konteks prediksi harga saham di pasar modal Indonesia. Hal ini membuka peluang riset untuk mengeksplorasi performa relatif masing-masing model dalam kondisi pasar domestik dengan karakteristik data lokal yang khas [14].

Oleh karena itu, kontribusi utama dari makalah ini adalah menyajikan bukti empiris mengenai perbandingan kinerja langsung antara ketiga algoritma boosting AdaBoost, XGBoost, dan CatBoost dalam konteks pasar saham Indonesia yang memiliki karakteristik unik, seperti volatilitas tinggi, keterbatasan data historis, dan dominasi investor ritel. Penelitian ini diharapkan dapat mengisi kesenjangan literatur yang masih terbatas dalam komparasi metode ensemble learning pada pasar domestik [15].

2. Metode Penelitian

Penelitian ini merupakan penelitian dengan pendekatan kuantitatif dengan metode analisis sekunder. Informasi yang dipakai adalah nilai saham PT. Bank BCA Tbk (BBCA) dalam rentang waktu tertentu, mulai dari 01 Januari 2020 hingga 25 September 2025 yang diperoleh dari Yahoo Finance.

2.1 Metode Penelitian

1. **Pengumpulan Data**
Data yang digunakan berupa data historis saham BCA yang diperoleh dari sumber publik seperti Yahoo Finance. Data meliputi atribut tanggal, harga pembukaan (*open*), harga tertinggi (*high*), harga terendah (*low*), harga penutupan (*close*), *adjusted close* (*adj close*) dan volume perdagangan (*volume*).
2. **Data Preprocessing**
Pada tahap ini, dilakukan pengecekan data serta pembersihan data untuk menghilangkan *missing*

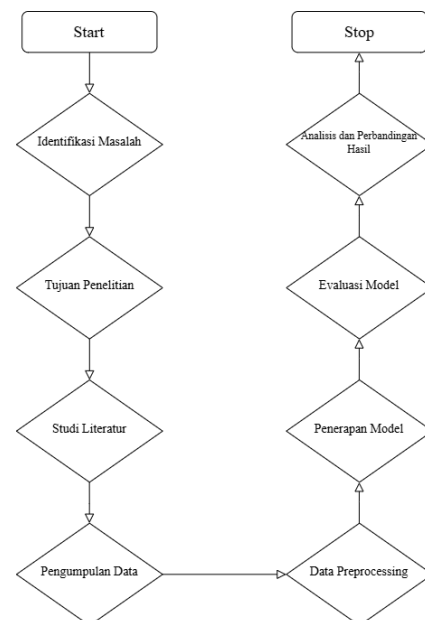
values dan data anomali guna menghindari kesalahan dalam proses analisis dan pelatihan algoritma. Pada tahap selanjutnya, dataset dibagi menjadi dua bagian, yaitu *data training* sebesar 80% dan *data testing* sebesar 20%.

3. Pelatihan Model

Pada tahap ini, dilakukan pemodelan pada dataset menggunakan 3 metode yaitu, AdaBoost, CatBoost, dan XGBoost. Metode ini digunakan untuk memprediksi harga saham serta mengevaluasi metode mana yang paling akurat dalam memprediksi harga *closing* saham BBCA.

2.2 Alur Penelitian

Dalam pelaksanaan penelitian, alur penelitian yang diterapkan dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

1. Identifikasi Masalah

Tahap awal penelitian adalah identifikasi masalah fundamental yang menjadi dasar penelitian. Pasar saham memiliki karakteristik yang kompleks, non-linier, dan mengandung unsur ketidakpastian (*noise*) yang signifikan, sehingga prediksi pergerakan harganya menjadi sebuah tantangan ilmiah. Masalah utama yang dirumuskan adalah kebutuhan untuk memperoleh model peramalan harga penutupan saham yang tidak hanya memiliki tingkat akurasi tinggi (meminimalkan galat prediksi), tetapi juga efisien secara komputasi (membutuhkan waktu pelatihan yang wajar) agar dapat diaplikasikan secara praktis oleh para pemangku kepentingan, seperti investor dan analis keuangan.

2. Tujuan Penelitian

Berdasarkan identifikasi masalah, tujuan penelitian dirumuskan secara spesifik dan terukur. Tujuan utama dari penelitian ini adalah untuk melaksanakan analisis komparatif secara kuantitatif terhadap performa tiga algoritma *boosting*, yaitu AdaBoost, XGBoost, dan CatBoost, dalam konteks prediksi harga saham BBKA. Tujuan ini dipecah menjadi beberapa sasaran yang lebih detail:

- Mengevaluasi akurasi prediktif dari setiap model dengan menggunakan metrik evaluasi standar untuk regresi.
- Mengukur efisiensi komputasi dari setiap algoritma melalui pencatatan waktu yang diperlukan untuk proses pelatihan model.
- Mengidentifikasi model yang paling optimal dengan menganalisis trade-off antara akurasi prediksi dan efisiensi komputasi.

3. Studi Literatur

Pada tahap ini, dilakukan kajian pustaka secara mendalam terhadap penelitian-penelitian terdahulu yang relevan. Fokus utama adalah pada literatur di bidang keuangan komputasional, khususnya yang membahas aplikasi *machine learning* untuk prediksi pasar saham. Tujuan dari studi pustaka ini adalah untuk membangun landasan teoretis yang kokoh, memahami posisi penelitian ini dalam konteks *state-of-the-art*, serta memvalidasi pemilihan algoritma AdaBoost, XGBoost, dan CatBoost sebagai model kandidat yang relevan dan berkinerja tinggi.

4. Pengumpulan Data

Penelitian ini menggunakan data sekunder yang bersifat kuantitatif dan runtut waktu (*time series*). Data yang dikumpulkan adalah data historis harga saham harian PT Bank Central Asia Tbk (BBCA) dalam rentang waktu 1 Januari 2020 hingga 25 September 2025. Data ini diperoleh dari sumber publik yang kredibel, yaitu Yahoo Finance. Atribut data yang digunakan dalam penelitian ini mencakup:

- Open: Harga pembukaan
- High: Harga tertinggi harian
- Low: Harga terendah harian
- Close: Harga penutupan, yang ditetapkan sebagai variabel target (*dependen*)
- Volume: Volume saham yang diperdagangkan

5. Data Preprocessing

Data mentah yang telah dikumpulkan kemudian melalui tahap persiapan yang krusial sebelum digunakan untuk pemodelan. Tahap ini meliputi:

- Pembersihan Data: Verifikasi integritas data untuk memastikan tidak terdapat nilai yang hilang (*missing values*) atau anomali yang dapat mengganggu kinerja model.
- Rekayasa Fitur (*Feature Engineering*): Membuat fitur-fitur prediktor (variabel independen) dari data historis. Dalam penelitian ini, teknik *lag features* diterapkan, di mana harga penutupan dari N hari sebelumnya ($t-1, t-2, \dots, t-N$) digunakan untuk memprediksi harga penutupan pada hari- t .
- Pembagian Dataset: Dataset dibagi menjadi dua bagian dengan proporsi 80% untuk data latih (*training set*) dan 20% untuk data uji (*testing set*). Pembagian ini dilakukan secara kronologis (berurutan berdasarkan waktu) untuk mencegah kebocoran data (*data leakage*) dan untuk mensimulasikan kondisi peramalan yang realistis.

6. Penerapan Model

Pada tahap ini, ketiga algoritma (AdaBoost, XGBoost, dan CatBoost) dilatih secara terpisah menggunakan *training set*. Proses pelatihan merupakan fase di mana setiap model secara iteratif menyesuaikan parameter internalnya untuk mempelajari pola, tren, dan hubungan statistik yang ada antara fitur-fitur prediktor (harga historis) dengan variabel target (harga penutupan).

7. Evaluasi Model

Setelah proses pelatihan selesai, performa setiap model dievaluasi menggunakan *testing set*, yaitu data yang belum pernah dilihat oleh model sebelumnya. Tahap ini bertujuan untuk mengukur kemampuan generalisasi model. Kinerja prediksi dibandingkan dengan nilai aktual menggunakan metrik-metrik berikut:

- Mean Absolute Error (MAE): Rata-rata dari nilai absolut galat.
- Root Mean Square Error (RMSE): Akar kuadrat dari rata-rata galat kuadrat, yang memberikan penalti lebih besar pada galat yang besar.
- Mean Absolute Percentage Error (MAPE): Rata-rata persentase galat, yang mudah diinterpretasikan.
- R^2 Score (Koefisien Determinasi): Ukuran proporsi varians dalam variabel dependen yang dapat diprediksi dari variabel independen.

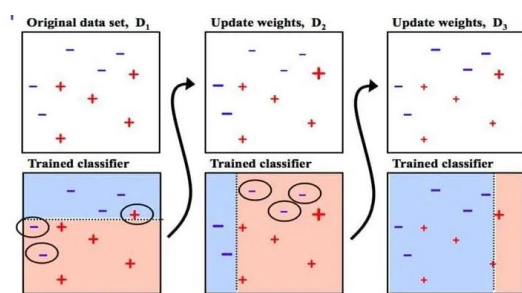
8. Analisis dan Perbandingan Hasil

Tahap akhir adalah melakukan analisis sintesis terhadap hasil evaluasi. Kinerja kuantitatif dari ketiga model dibandingkan secara langsung. Analisis ini tidak hanya terfokus pada metrik

akurasi, tetapi juga mempertimbangkan efisiensi komputasi (waktu pelatihan). Berdasarkan bukti empiris dari analisis komparatif ini, ditarik kesimpulan untuk menjawab pertanyaan penelitian, yaitu menentukan algoritma mana yang menunjukkan kinerja paling unggul dan optimal untuk studi kasus yang diangkat.

2.3 Algoritma AdaBoost

Algoritma ini disebut Adaptive Boosting karena bobot diberikan ulang pada setiap *instance*, dengan bobot yang lebih tinggi diberikan pada *instance* yang salah diklasifikasikan. Algoritma AdaBoost bersifat iteratif atau berulang. Cara kerja algoritma ini dimulai dengan melatih sebuah *weak classifier* pada data pelatihan [16].

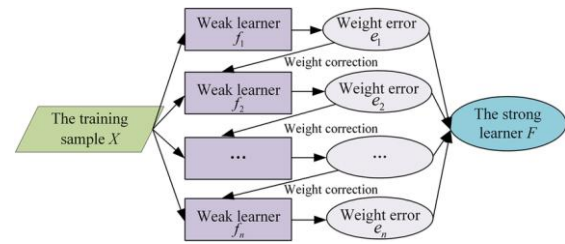


Gambar 2. Ilustrasi Cara Kerja AdaBoost

Weak classifier kemudian diberi bobot berdasarkan performanya. Selanjutnya, algoritma melatih *weak classifier* kedua menggunakan data yang telah diberi bobot. *Weak classifier* kedua kemudian diberi bobot berdasarkan performanya. Proses ini diulang sejumlah iterasi tertentu atau hingga tingkat kesalahan berada di bawah ambang batas yang ditentukan. Classifier akhir adalah rata-rata terbobot dari semua *weak classifiers*. Bobot ditentukan berdasarkan tingkat kesalahan dari masing-masing *weak classifier*. Semakin rendah tingkat kesalahan, semakin tinggi bobotnya.

2.4 Algoritma CatBoost

Algoritma CatBoost dikembangkan untuk meningkatkan performa model *Machine Learning* dengan fokus pada kecepatan, akurasi, dan kemampuan penanganan fitur kategorikal.

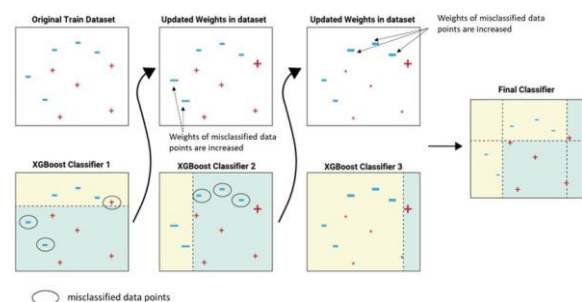


Gambar 3. Ilustrasi Cara Kerja CatBoost

CatBoost didasarkan pada teknik Gradient Boosting, yang merupakan salah satu metode populer dalam algoritma Boosting [17]. Gradient Boosting bekerja dengan menggabungkan sejumlah model sederhana (*weak learner*) secara sekuensial. Setiap model dihasilkan dengan meminimalkan gradien dari fungsi kerugian (*loss function*) berdasarkan *residual error* dari model sebelumnya. Dengan melakukan iterasi ini, Gradient Boosting secara adaptif meningkatkan performa model secara keseluruhan.

2.5 Algoritma XGBoost

XGBoost merupakan salah satu implementasi dari Gradient Boosting yang dirancang untuk membentuk model prediksi yang kuat dengan menggunakan sejumlah model lemah, seperti pohon keputusan (*decision tree*). Algoritma Gradient Boosting menggabungkan pembelajar lemah (*weak learner*) menjadi satu pembelajar kuat (*strong learner*) tunggal secara iteratif [18].



Gambar 4. Ilustrasi Cara Kerja XGBoost

Teknik ini menggunakan *weak classifier* untuk secara iteratif membangun sebuah *strong classifier*. *Weak learner* atau *weak classifier* adalah model yang kinerjanya hanya sedikit lebih baik dari tebakan acak. Bobot awal didistribusikan secara merata pada seluruh data pelatihan. Setelah prediksi dilakukan oleh *weak classifier* pertama (pada gambar *XGBoost Classifier 1*), bobot yang lebih besar diberikan pada data yang salah diklasifikasikan. Hal ini dilakukan agar *weak classifier* berikutnya, yaitu *XGBoost Classifier 2*, dapat lebih fokus pada data yang sulit tersebut. Kemudian setelah prediksi oleh *weak classifier* kedua (pada gambar *XGBoost Classifier 2*), bobot yang lebih besar kembali diberikan pada data yang salah diklasifikasikan, dan proses ini berlanjut seterusnya. Classifier akhir merupakan kombinasi berbobot dari seluruh *weak classifier* yang telah dibentuk.

2.6 Metode Evaluasi

Untuk mengukur dan membandingkan kinerja dari ketiga model yang diuji, penelitian ini menggunakan beberapa metrik evaluasi standar [19]. Hasil prediksi dari setiap model pada data uji dibandingkan dengan nilai aktualnya, dan tingkat kesalahan (*error*) dihitung menggunakan metrik-metrik berikut:

1. Mean Absolute Error (MAE)

Mengukur rata-rata dari selisih absolut antara nilai prediksi dan nilai aktual. MAE memberikan gambaran mengenai besaran rata-rata kesalahan tanpa memperhatikan arahnya [20].

$$\frac{1}{m} \sum_{i=1}^m |X_i - Y_i|$$

2. Root Mean Square Error (RMSE)

Merupakan akar kuadrat dari rata-rata selisih kuadrat antara nilai prediksi dan aktual. RMSE memberikan bobot lebih besar pada kesalahan yang besar, sehingga metrik ini sangat sensitif terhadap *outlier* atau prediksi yang sangat melenceng [20].

$$\sqrt{\frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2}$$

3. Mean Absolute Percentage Error (MAPE)

Menghitung rata-rata persentase kesalahan absolut. Metrik ini sangat intuitif karena menyajikan kesalahan dalam bentuk persentase, sehingga mudah untuk diinterpretasikan [21].

$$\frac{1}{m} \sum_{i=1}^m \left| \frac{Y_i - X_i}{Y_i} \right|$$

4. R^2 Score (Koefisien Determinasi)

Mengukur seberapa baik model dapat menjelaskan variasi dari data. Nilai R^2 Score berkisar antara 0 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan bahwa model semakin baik dalam memprediksi data [22].

$$1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (Y_i - \bar{Y})^2}$$

3. Hasil dan Pembahasan

3.1 Dataset

Sumber data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari website resmi Yahoo Finance. Dataset ini berisi harga saham harian PT Bank Central Asia (BCA) yang mencakup periode 1 Januari 2020 hingga 25 September 2025. Data tersebut terdiri dari beberapa atribut, antara lain 'Date' (tanggal transaksi), 'Open' (harga buka saham), 'High' (harga tertinggi saham), 'Low' (harga terendah saham), 'Close' (harga tutup saham), dan 'Volume' (jumlah saham yang diperdagangkan). Pada penelitian ini, atribut 'Close' yang akan digunakan sebagai variabel target prediksi.

Date	Open	High	Low	Close	Volume
Jan 20, 2020	7,000.00	7,000.00	7,000.00	7,000.00	100,000,000
Jan 21, 2020	7,000.00	7,000.00	7,000.00	7,000.00	100,000,000
Jan 22, 2020	7,000.00	7,000.00	7,000.00	7,000.00	100,000,000
Jan 23, 2020	7,000.00	7,000.00	7,000.00	7,000.00	100,000,000
Jan 24, 2020	7,000.00	7,000.00	7,000.00	7,000.00	100,000,000
Jan 25, 2020	7,000.00	7,000.00	7,000.00	7,000.00	100,000,000
Jan 26, 2020	7,000.00	7,000.00	7,000.00	7,000.00	100,000,000
Jan 27, 2020	7,000.00	7,000.00	7,000.00	7,000.00	100,000,000
Jan 28, 2020	7,000.00	7,000.00	7,000.00	7,000.00	100,000,000
Jan 29, 2020	7,000.00	7,000.00	7,000.00	7,000.00	100,000,000
Jan 30, 2020	7,000.00	7,000.00	7,000.00	7,000.00	100,000,000

Gambar 5. Tampilan Website Yahoo Finance

Pada penelitian ini, peneliti menggunakan library *yfinance* pada bahasa pemrograman Python. Perintah untuk memperoleh data harga saham PT Bank Central Asia (BBCA.JK) ditampilkan pada Gambar 6.

```
import yfinance as yf
import pandas as pd

# Ticker Yahoo Finance
ticker = "BBCA.JK"

# Unduh data harga saham dari 1 Jan 2020 sampai 25 Sep 2025
df = yf.download(ticker, start="2020-01-01", end="2025-09-25", progress=False)

# Lihat 5 baris pertama
print(df.head())

# Simpan ke csv
df.to_csv("BBCA_JK_history.csv", index=True)

# Simpan ke Excel
df.to_excel("BBCA_JK_history.xlsx", index=True)

print("Data berhasil disimpan ke BBCA_JK_history.csv dan BBCA_JK_history.xlsx")
```

Gambar 6. Pengambilan Data

Data yang telah ditarik dari library kemudian akan dimasukkan ke dalam Excel atau format .xlsx yang kemudian file tersebut akan digunakan dalam melakukan prediksi harga *closing* saham BBCA.

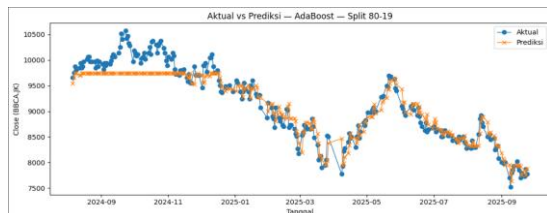
Data yang telah disimpan dalam format Excel selanjutnya dilakukan pembersihan dan *preprocessing* data. Proses pembersihan dilakukan guna memastikan bahwa data yang digunakan dalam penelitian bersifat valid dan konsisten. Pada tahap ini, dilakukan pemeriksaan data yang kosong atau *missing values* serta data duplikat yang berpotensi mempengaruhi akurasi model.

Setelah data dinyatakan bersih, dilakukan tahap pembagian data (*data split*) menjadi dua bagian yaitu data latih dan data uji dengan proporsi 80% data latih dan 20% data uji. *Data split* dilakukan secara berurutan agar tidak mengganggu sifat kronologis dari data saham yang bersifat *time series*.

3.2 Hasil Eksperimen

Pada tahap eksperimen, dilakukan dengan menerapkan tiga algoritma ensemble learning meliputi AdaBoost, CatBoost, dan XGBoost dalam melakukan prediksi harga *closing* saham BBCA. Pada fitur prediktor, penelitian ini menggunakan 30 fitur yang merupakan data harga *closing* saham dari 30 hari sebelumnya (*lag-30*). Kemudian model dilatih untuk memprediksi harga *closing* di hari berikutnya menggunakan tiga algoritma *ensemble learning* yang sudah ditentukan, yaitu AdaBoost, CatBoost, XGBoost.

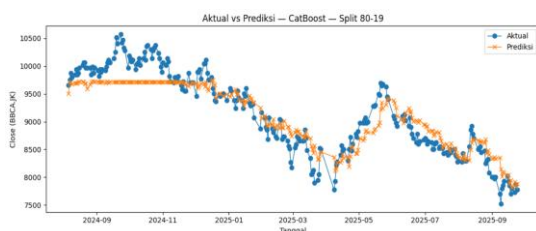
1. AdaBoost



Gambar 7. Grafik Hasil Algoritma AdaBoost

Hasil grafik pada prediksi harga *closing* saham BBCA menggunakan AdaBoost menunjukkan bahwa AdaBoost mampu mengikuti pola pergerakan harga saham dengan cukup baik terutama pada bagian pertengahan hingga akhir data uji. Pada grafik tersebut, terlihat kurva prediksi cenderung sejajar dengan kurva aktual, yang menandakan model berhasil mengikuti tren utama, baik pada saat harga saham mengalami penurunan maupun peningkatan. Namun pada awal periode pengujian, hasil prediksi cenderung datar dan tidak mengikuti kurva aktual. Dengan demikian, AdaBoost memiliki kemampuan untuk memprediksi pola umum harga saham yang cukup baik, namun AdaBoost kurang responsif terhadap perubahan harga yang bersifat cepat dan tidak terduga.

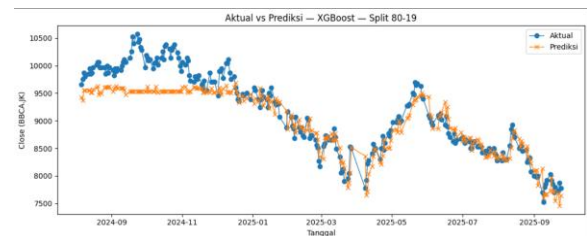
2. CatBoost



Gambar 8. Grafik Hasil Algoritma CatBoost

Hasil grafik pada prediksi harga *closing* saham BBCA menggunakan CatBoost menunjukkan bahwa CatBoost dapat mengikuti pola harga saham dengan cukup baik. CatBoost mampu mengikuti tren utama pergerakan harga, baik saat harga saham mengalami penurunan maupun kenaikan, meskipun pada beberapa titik tinggi prediksi sedikit tertinggal dari kurva aktual. Hal ini menunjukkan bahwa CatBoost memiliki kemampuan generalisasi yang baik serta menghasilkan prediksi yang stabil dan halus. Dengan demikian, CatBoost memiliki kemampuan untuk mempresentasikan pola harga saham dengan akurat dan konsisten, serta CatBoost memiliki kemampuan yang lebih baik dibandingkan AdaBoost dalam menyesuaikan fluktuasi harga yang kompleks.

3. XGBoost



Gambar 9. Grafik Hasil Algoritma XGBoost

Hasil grafik pada prediksi harga *closing* saham BBCA menggunakan XGBoost menunjukkan bahwa XGBoost dapat mengikuti tren harga dengan baik, termasuk saat terjadi penurunan tajam maupun kenaikan harga. Hasil prediksi terlihat lebih stabil dan akurat dibandingkan dua model lainnya. Perbedaan antara kurva aktual dan kurva prediksi relatif kecil, menandakan bahwa XGBoost mampu menangkap pola non-linear dan kompleks dari data historis dengan lebih efektif. Dengan demikian, XGBoost menghasilkan performa terbaik dalam memprediksi harga penutupan saham BCA. Model ini mampu menyeimbangkan akurasi dan stabilitas prediksi, sehingga lebih andal digunakan untuk memodelkan data saham yang dinamis dan fluktuatif.

Tabel 1. Perbandingan Hasil Evaluasi Setiap Algoritma

Algoritma	AdaBoost	CatBoost	XGBoost
MAE	184.5846	241.5416	248.1175
RMSE	243.8408	296.3128	329.0475
MAPE (%)	1.97	2.62	2.61
R^2 Score	0.8935	0.8428	0.8061
Train Time (s)	0.6297	23.5195	1.4458

Berdasarkan hasil eksperimen yang telah dilakukan, dapat dilihat pada Tabel 1, bisa disimpulkan bahwa algoritma AdaBoost menunjukkan kinerja paling unggul di antara dua model lainnya yang diuji (CatBoost dan XGBoost) baik dari segi akurasi maupun waktu *training*.

Dari sisi akurasi, AdaBoost secara konsisten menghasilkan tingkat kesalahan prediksi yang paling rendah dibandingkan dua algoritma lainnya. Hal ini dibuktikan dengan nilai *Root Mean Square Error* (RMSE) sebesar 243.8408 dan *Mean Absolute Error* (MAE) sebesar 184.5846, yang keduanya merupakan angka terendah dibandingkan XGBoost dan CatBoost. Selain itu, nilai R^2 Score yang mencapai 0.8935 adalah yang tertinggi, yang mengindikasikan bahwa model AdaBoost mampu menjelaskan sekitar 89.35% variasi data harga saham BBKA, menunjukkan kecocokan model yang sangat baik.

Dari segi efisiensi secara waktu, AdaBoost juga menjadi model yang paling cepat dalam proses pelatihan (*training time*), dengan hanya membutuhkan waktu 0.6297 detik. Waktu ini jauh lebih singkat jika dibandingkan dengan XGBoost yang memerlukan 1.4458 detik dan CatBoost yang memakan proses waktu paling lama, yaitu 23.5195 detik. Meskipun XGBoost dan CatBoost seringkali dianggap sebagai algoritma yang sangat kuat dalam berbagai kasus, untuk konfigurasi dan dataset pada penelitian ini, keduanya menunjukkan tingkat kesalahan yang lebih tinggi dan R^2 Score yang lebih rendah. Dengan demikian, berdasarkan kombinasi akurasi prediksi dan kecepatan komputasi, AdaBoost terbukti menjadi model yang lebih unggul dalam kasus ini dibandingkan dua model algoritma lainnya yang diuji.

4. Kesimpulan

Berdasarkan analisis komparatif yang telah dilakukan terhadap tiga algoritma *boosting* untuk memprediksi harga penutupan saham BBKA, dapat ditarik kesimpulan yang lebih rinci. Penelitian ini menyimpulkan bahwa dari ketiga model yang diuji, AdaBoost merupakan model yang paling unggul dan optimal untuk studi kasus ini. Keunggulan ini tidak hanya terlihat pada satu metrik, tetapi secara konsisten di berbagai aspek evaluasi. Secara kuantitatif, superioritas AdaBoost terbukti melalui pencapaian nilai RMSE terendah sebesar 243.8408 dan R^2 Score tertinggi sebesar 0.8935, yang menandakan bahwa model ini memiliki tingkat kesalahan prediksi paling kecil sekaligus kemampuan terbaik dalam menjelaskan variabilitas data harga saham. Dari sisi efisiensi komputasi, AdaBoost juga menunjukkan kinerja terbaik dengan waktu pelatihan hanya 0.6297 detik, menjadikannya model yang paling efisien untuk dilatih dibandingkan XGBoost dan CatBoost yang jauh lebih lambat.

Namun, penelitian ini memiliki beberapa keterbatasan. Pertama, lingkup penelitian hanya terbatas pada satu jenis saham (*blue-chip*), yaitu BBKA, sehingga hasilnya mungkin tidak dapat digeneralisasi untuk saham dari sektor lain yang memiliki tingkat volatilitas dan likuiditas berbeda. Kedua, fitur yang digunakan sebagai

prediktor hanya terbatas pada data harga historis (*fitur lag*), tanpa mengintegrasikan faktor fundamental perusahaan, data ekonomi makro, atau sentimen pasar dari berita yang dapat memengaruhi pergerakan harga secara signifikan. Ketiga, proses optimasi *hyperparameter* untuk setiap model dilakukan secara standar dan bukan melalui pencarian ekstensif, yang berpotensi memengaruhi performa akhir dari masing-masing algoritma.

Untuk pengembangan penelitian selanjutnya, beberapa saran dapat dipertimbangkan. Pertama, studi serupa dapat diperluas dengan menerapkan perbandingan model pada portofolio saham yang lebih beragam dari berbagai sektor di Bursa Efek Indonesia. Kedua, penelitian di masa depan disarankan untuk memperkaya fitur input dengan data multi-modal, seperti mengintegrasikan analisis sentimen dari berita keuangan menggunakan teknik *Natural Language Processing* (NLP) atau memasukkan variabel ekonomi makro seperti suku bunga dan inflasi. Ketiga, disarankan untuk membandingkan performa model *boosting* ini dengan arsitektur lain yang dirancang khusus untuk data sekuensial, seperti *Long Short-Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU), yang mungkin dapat menangkap dependensi temporal data dengan lebih baik.

REFERENSI

- [1] Putra Lidyansyah Aldy and Kurniawati Ana, "Analisis Prediksi Harga Saham PT. Astra International Tbk Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) dan Support Vector Regression (SVR)," *Jurnal Ilmiah Komputasi*, vol. 20, no. 3, Sep. 2021, doi: 10.32409/jikstik.20.3.2732.
- [2] M. Nabipour, P. Nayyeri, H. Jabani, S. Shahab, and A. Mosavi, "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis," *IEEE Access*, vol. 8, pp. 150199–150212, 2020, doi: 10.1109/ACCESS.2020.3015966.
- [3] I. E. Livieris, N. Kiriakidou, S. Stavroyiannis, and P. Pintelas, "An advanced CNN-LSTM model for cryptocurrency forecasting," *Electronics (Switzerland)*, vol. 10, no. 3, pp. 1–16, Feb. 2021, doi: 10.3390/electronics10030287.
- [4] M. Mishra, M. Patil, G. R. Student, T. Chaudhari, P. Guide, and A. Professor, "A SURVEY ON STOCK PRICE PREDICTION USING MACHINE LEARNING," 2021. [Online]. Available: <http://www.ijeast.com>
- [5] E. Purnama Sari *et al.*, "Studi Literatur Deep Learning dan Machine Learning untuk Analisis dan Prediksi Pasar Saham: Metodologi,

- Representasi Data dan Studi Kasus,” 2024. [Online]. Available: <https://journal.scitechgrup.com/index.php/jtsm>
- [6] D. Setiawan, T. Wijayanti Septiarini, and H. Kurniawan, “SYSTEMATIC LITERATURE REVIEW: METODE DAN TEKNIK PERAMALAN CRYPTOCURRENCY,” *Prosiding Seminar Nasional Sains dan Teknologi Seri III Fakultas Sains dan Teknologi*, vol. 2, no. 1, 2025.
- [7] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, “Financial time series forecasting with deep learning: A systematic literature review: 2005–2019,” *Applied Soft Computing Journal*, vol. 90, May 2020, doi: 10.1016/j.asoc.2020.106181.
- [8] Y. Zhang, “Stock Price Prediction Method Based on XGBoost Algorithm,” 2023, pp. 595–603. doi: 10.2991/978-94-6463-030-5_60.
- [9] F. A. Kurnia, M. Hardianti, M. Sinurat, and L. Cahyadi, “Analisis Prediksi Harga Saham PT. BCA Dengan Menggunakan Metode ARIMA,” *eCo-Fin*, vol. 7, no. 2, pp. 880–896, Jun. 2025, doi: 10.32877/ef.v7i2.2373.
- [10] P. H. Vuong, T. T. Dat, T. K. Mai, P. H. Uyen, and P. T. Bao, “Stock-price forecasting based on XGBoost and LSTM,” *Computer Systems Science and Engineering*, vol. 40, no. 1, pp. 237–246, 2022, doi: 10.32604/CSSE.2022.017685.
- [11] M. Yusuf Andrika and M. Rahardi, “Comparative Study of Linear Regression, SVR, and XGBoost for Stock Price Prediction After a Stock Split,” 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [12] H. N. Bhandari, B. Rimal, N. R. Pokhrel, R. Rimal, K. R. Dahal, and R. K. C. Khatri, “Predicting stock market index using LSTM,” *Machine Learning with Applications*, vol. 9, p. 100320, Sep. 2022, doi: 10.1016/j.mlwa.2022.100320.
- [13] R. Irfannandhy, L. B. Handoko, and N. Ariyanto, “Analisis Performa Model Random Forest dan CatBoost dengan Teknik SMOTE dalam Prediksi Risiko Diabetes,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, pp. 714–723, Dec. 2024, doi: 10.29408/edumatic.v8i2.27990.
- [14] S. Mohapatra, R. Mukherjee, A. Roy, A. Sengupta, and A. Puniyani, “Can Ensemble Machine Learning Methods Predict Stock Returns for Indian Banks Using Technical Indicators?,” *Journal of Risk and Financial Management*, vol. 15, no. 8, Aug. 2022, doi: 10.3390/jrfm15080350.
- [15] I. K. Nti, A. F. Adekoya, and B. A. Weyori, “A comprehensive evaluation of ensemble learning for stock-market prediction,” *J Big Data*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40537-020-00299-5.
- [16] P. Sornsuwit, “ENHANCE WEAK LEARNER MODEL OF ADABOOST (EWDM) FOR DIABETES MELLITUS CLASSIFICATION,” *International Journal of Innovative Computing, Information and Control*, vol. 18, no. 4, pp. 1117–1132, Aug. 2022, doi: 10.24507/ijicic.18.04.1117.
- [17] J. T. Hancock and T. M. Khoshgoftaar, “CatBoost for big data: an interdisciplinary review,” *J Big Data*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40537-020-00369-8.
- [18] P. Zhang, Y. Jia, and Y. Shang, “Research and application of XGBoost in imbalanced data,” *Int J Distrib Sens Netw*, vol. 18, no. 6, Jun. 2022, doi: 10.1177/15501329221106935.
- [19] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” *PeerJ Comput Sci*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.
- [20] T. O. Hodson, “Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not,” Jul. 19, 2022, *Copernicus GmbH*. doi: 10.5194/gmd-15-5481-2022.
- [21] E. Vivas, H. Allende-Cid, and R. Salas, “A systematic review of statistical and machine learning methods for electrical power forecasting with reported mape score,” Dec. 01, 2020, *MDPI AG*. doi: 10.3390/e22121412.
- [22] J. Gao, “R-Squared (R^2) – How much variation is explained?,” *Research Methods in Medicine & Health Sciences*, vol. 5, no. 4, pp. 104–109, Sep. 2024, doi: 10.1177/26320843231186398.