

Performance Analysis of XGBoost Algorithm to Determine the Most Optimal Parameters and Features in Predicting Stock Price Movement

Analisis Performa Algoritma XGBoost untuk Menentukan Parameter dan Fitur yang Paling Optimal dalam Memprediksi Pergerakan Harga Saham

¹ **Affan Ardana**

¹ Informatika, Universitas Amikom Yogyakarta, Indonesia

^{1*} email.affanardana@gmail.com

*: *Penulis korespondensi (corresponding author)*

Informasi Artikel

Received: December 2022

Revised: January 2023

Accepted: January 2023

Published: February 2023

Abstract

Purpose: The research aims to find the best parameters and features for predicting stock price movement using the XGBoost algorithm. The parameters are searched using the RMSE value, and the features are searched using the importance value.

Design/methodology/approach: The research data is the stock data of Amazon.com company (AMZN). The dataset contains the Date, Low, Open, Volume, High, Close, and Adjusted Close features. The dataset is ensured to have no missing data by handling missing values. The input feature is selected using the Pearson Correlation feature selection method. To prevent the difference between the highest and lowest stock price from being too far apart, the data is scaled using the scaling method. To avoid bias that may appear in the prediction result, cross-validation is used with the Min Max Scaling method, which will divide the dataset into training data and testing data within a range of 30 days after the training data. The parameters to be tested include $n_estimator = 500$, $early_stopping_round = 3$, $learning_rate = 0.01, 0.05, 0.1$, and max_depth (tree depth) = 3, 4, 5.

Findings/result: The result of the research that a learning rate of 0.05 and a tree depth of 5 obtained the lowest RMSE result compared to other models, with an RMSE of 0.009437. The Low feature obtained the highest importance value among all the models built.

Originality/value/state of the art: This study used testing data within a range of 30 days after the training data and used a combination of parameters, including $n_estimator = 500$, $early_stopping_round = 3$, $learning_rate = 0.01, 0.05, 0.1$, and max_depth (tree depth) = 3, 4, 5.

Keywords: machine learning;
regression; prediction; stock; xgboost
Kata kunci: machine learning; regresi;
prediksi; saham; xgboost

Abstrak

Tujuan: Penelitian bertujuan mencari parameter dan fitur yang terbaik untuk memprediksi pergerakan harga saham menggunakan algoritma XGBoost. Parameter dicari menggunakan nilai RMSE dan fitur dicari menggunakan nilai *importance*.

Perancangan/metode/pendekatan: Data penelitian adalah data saham perusahaan Amazon.com (AMZN). Dataset berisi fitur Date, Low, Open, Volume, High, Close, dan Adjusted Close. Dataset dipastikan tidak ada data kosong dengan *handling missing values*. Fitur yang dipilih sebagai *input* menggunakan *features selection* metode *Pearson Correlation*. Agar perbedaan harga saham tertinggi dan terendah tidak terlampaui jauh maka menggunakan metode *scaling* data. Untuk menghindari bias yang muncul pada hasil prediksi digunakanlah *cross validation* dengan metode *Min Max Scaling* yang akan membagi dataset menjadi data *training* dan data *testing* dalam rentang 30 hari setelah data *training*. Parameter yang akan diuji meliputi *n_estimator* = 500, *early stopping round* = 3, *learning rate* = 0.01, 0.05, 0.1, dan *max_depth* (kedalaman pohon) = 3, 4, 5.

Hasil: Hasil penelitian yang dilakukan menunjukkan bahwa parameter *learning rate* sebesar 0,05 dan kedalaman pohon sebanyak 5 mendapatkan hasil RMSE terendah dari semua model lainnya dengan RMSE sebesar 0,009437. Fitur Low mendapatkan nilai *importance* tertinggi dari semua model yang telah dibangun.

Keaslian/ *state of the art*: Penelitian ini menggunakan data *testing* dalam rentang 30 hari setelah data *training* dan menggunakan kombinasi parameter *n_estimator* = 500, *early stopping round* = 3, *learning rate* = 0.01, 0.05, 0.1, dan *max_depth* (kedalaman pohon) = 3, 4, 5.

1. Pendahuluan

Saham merupakan salah satu jenis investasi yang paling banyak diminati oleh investor. Hal ini dikarenakan investor mengharapkan suatu *profit* yang besar dari saham. Saham termasuk instrumen investasi yang sangat menguntungkan jika diperhitungkan dengan benar [1]. Saham juga memiliki potensi kerugian yang sama besarnya dengan keuntungannya. Investor harus memahami informasi-informasi yang bisa dijadikan pedoman untuk mengatasi hal tersebut [2]. Risiko yang bisa terjadi dari investasi saham yaitu menurunnya harga saham (*capital loss*). Harga saham selalu mengalami kenaikan dan penurunan dari waktu ke waktu. Pergerakan harga tersebut biasanya membentuk sebuah pola tertentu [3]. Keuntungan dapat dimaksimalkan dengan melakukan analisis agar mendapat hasil investasi yang besar [1].

Pada umumnya pergerakan harga saham sangat fluktuatif. Harga saham yang bergerak cepat setiap saat termasuk ke dalam jenis data *time series* [1], [4]. Hal inilah yang membuat para investor kesulitan untuk melakukan analisis terhadap pergerakan harga saham. Salah satu solusi yang dapat ditawarkan untuk membantu investor melakukan analisis pergerakan harga saham adalah dengan melakukan prediksi [1], [5]. Prediksi membantu investor untuk mengetahui harga saham pada suatu waktu tertentu. Berbagai cara bisa digunakan untuk melakukan prediksi seperti dengan menggunakan statistika dan *machine learning* [1], [3]–[6]. Pada *machine learning* cara untuk menemukan sebuah pola dari data saham bisa dilakukan dengan *data mining* [7].

Data mining merupakan sebuah proses untuk menemukan suatu informasi dalam data [7]. *Data mining* yang merupakan multidisiplin ilmu membutuhkan statistika, kecerdasan buatan, dan basis data untuk menemukan pola yang ada pada data. Pola ditemukan dengan memisahkan dan mengenali data yang bermanfaat dari sebuah data utuh yang besar [8], [9]. Ada banyak metode yang ada pada *data mining* contohnya klasifikasi dan regresi [7]–[10]. Data saham yang termasuk data *time series* menggunakan metode regresi dalam melakukan proses *data mining*. Data saham memiliki banyak fitur yang ada di dalamnya seperti harga pembukaan, harga penutupan, harga tertinggi dan harga terendah [11]. Ada beberapa algoritma regresi yang bisa digunakan untuk melakukan prediksi harga saham seperti *Gradient Boost* dan *Support Vector Machine* (SVM) [1], [12].

Algoritma *Gradient Boost* adalah algoritma yang bisa digunakan untuk melakukan klasifikasi maupun regresi. Konsep dasar dari algoritma ini adalah meminimalkan *loss function*. Hasil akhir prediksi dari algoritma *Gradient Boost* adalah hasil dari prediksi setiap pohon keputusan yang dibuat [7]. Salah satu algoritma hasil pengembang dari *Gradient Boost* yang bisa digunakan untuk melakukan regresi yakni *Extreme Gradient Boost* (XGBoost). Algoritma XGBoost dikembangkan oleh Tianqi Chen pada tahun 2016. XGBoost menggunakan metode hasil pengembangan dari algoritma *Gradient Boost Decision Tree* (GBDT) yang dikemukakan oleh Friedman. Metode ini memiliki banyak parameter yang bisa dioptimalisasi guna mendapatkan hasil yang diinginkan [13].

Optimalisasi parameter atau *hyperparameter tuning* merupakan sebuah proses untuk meningkatkan hasil kinerja dari model yang telah dibangun menggunakan algoritma XGBoost. *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 [7]. Parameter yang optimal dalam kasus regresi adalah parameter yang memberikan hasil evaluasi seperti *Root Mean Squared Error* (RMSE) yang terendah [14].

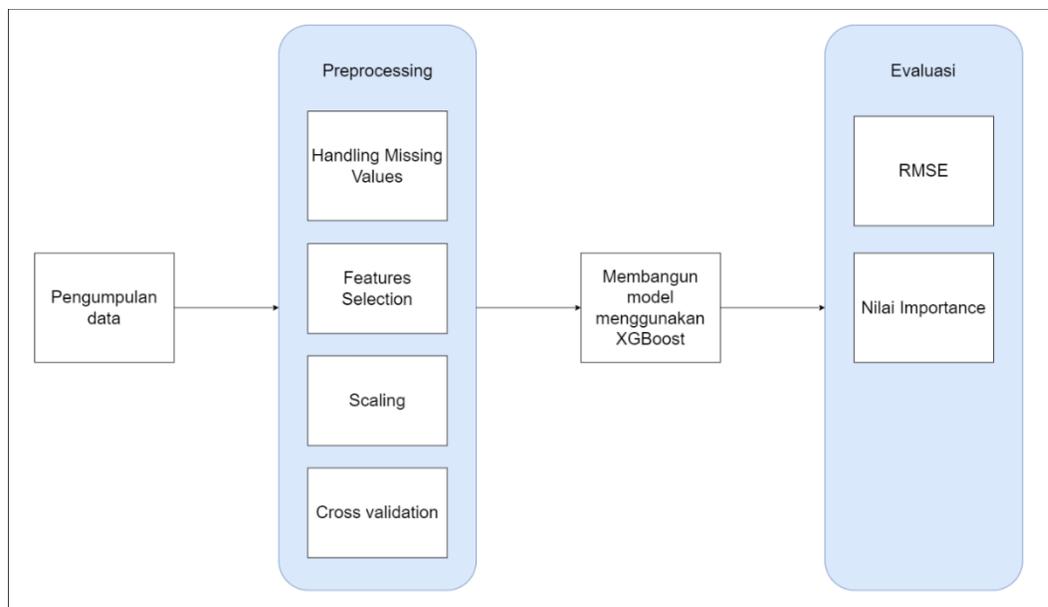
Berbagai penelitian telah dilakukan untuk melakukan regresi terhadap pergerakan harga saham dan mencari parameter yang paling optimal. Penggunaan algoritma XGBoost untuk melakukan klasifikasi kartu kredit menghasilkan nilai akurasi sebesar 80,02%, presisi sebesar 85,32%, dan *recall* sebesar 94,86%. Pada penelitian yang sama dilakukan *hyperparameter tuning* dan mendapat akurasi sebesar 83,42%, presisi sebesar 85,36%, dan *recall* sebesar 95,28%. Hasil penelitian tersebut menunjukkan bahwa *hyperparameter tuning* pada algoritma XGBoost dapat meningkatkan hasil kinerja model [7]. Penelitian lain mengenai prediksi kecelakaan lalu lintas di Bali menggunakan algoritma XGBoost menunjukkan hasil RMSE untuk kejadian kecelakaan sebesar 21,69, jumlah meninggal sebesar 4,92, jumlah luka berat sebesar 4,11, dan jumlah luka ringan sebesar 77,24. Hal ini menunjukkan regresi menggunakan algoritma XGBoost

menghasilkan nilai RMSE yang rendah [15]. Prediksi juga pernah dilakukan menggunakan algoritma SVM untuk memprediksi harga saham PT. Antm.Jk yang menghasilkan nilai RMSE sebesar 22,662 pada model awal sebelum dilakukan optimalisasi parameter dan 10,495 pada model yang telah dilakukan optimalisasi parameter [1]. Tidak hanya menggunakan *machine learning*, prediksi mengenai harga saham juga pernah dilakukan menggunakan statistika seperti pada prediksi harga saham Garuda Indonesia menggunakan metode ARIMA yang menghasilkan nilai RMSE sebesar 38 [3]. *Gradient Boost* lebih menghasilkan akurasi, stabilitas, dan efisiensi lebih tinggi dibandingkan dengan SVM dalam melakukan peramalan radiasi matahari global pada studi kasus China menggunakan data suhu dan curah hujan [12]. Penelitian tentang pilihan fitur yang digunakan dalam melakukan prediksi harga saham juga pernah dilakukan dengan cara menghitung *F-score* fitur tunggal dan melakukan kombinasi dari keseluruhan fitur yang telah ditentukan. Penelitian tersebut mendapatkan hasil akurasi sebesar 71% dan nilai *F-score* sebesar 70% [11].

Berawal dari permasalahan bahwa harga saham sangat fluktuatif yang membuat investor kesulitan melakukan analisis terhadap pergerakan harga saham. Penelitian ini dikembangkan untuk memberikan hasil prediksi yang paling baik menggunakan algoritma XGBoost. Prediksi yang baik berupa hasil prediksi dari model yang mempunyai nilai RMSE terendah. Optimalisasi parameter diperlukan guna mendapatkan nilai RMSE terendah. Penelitian ini juga mencari fitur yang paling dominan dalam memprediksi pergerakan harga saham dengan melihat nilai *importance* terbesar.

2. Metode/Perancangan

Penelitian ini menggunakan metode eksperimen yang digunakan untuk mencari model terbaik dari beberapa parameter yang dimasukkan. Tahapan penelitian yakni pengumpulan data, *preprocessing*, membangun model, dan evaluasi. Tahapan penelitian secara lengkap dapat dilihat pada **Gambar 1**.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Data yang digunakan adalah data saham perusahaan Amazon.com, Inc. yang memiliki nama AMZN. Data memiliki 7 fitur yaitu Date, Low, Open, Volume, High, Close, Adjusted Close. Contoh data dapat dilihat pada **Tabel 1** serta penjelasan Kolom (fitur) dapat dilihat pada **Tabel 2**

Tabel 1. Contoh Data

Date	Low	Open	Volume	High	Close Before	Adjusted Close
18-10-2022	114.790001	119.059998	65607400	119.519997	0.097917	116.360001
19-10-2022	113.220001	114.709999	47198100	116.589996	0.086458	115.070000
20-10-2022	113.510002	113.830002	48795100	118.239998	0.085417	115.250000
21-10-2022	114.500000	114.790001	55611900	119.589996	0.081771	119.320000

Tabel 2. Penjelasan Kolom

Fitur	Deskripsi
Date	Tanggal/sesi perdagangan dari data saham.
Low	Harga terendah saham pada sesi perdagangan.
Open	Harga awal saham pada saat sesi perdagangan dimulai.
Volume	Jumlah total transaksi pada saat sesi perdagangan.
High	Harga tertinggi saham pada sesi perdagangan.
Close Before	Harga akhir saham pada saat sesi perdagangan hari sebelumnya.
Adjusted Close	Harga saham pada akhir sesi yang telah disesuaikan untuk memperhitungkan perubahan dalam struktur modal perusahaan. Perubahan terjadi seperti karena pembayaran dividen atau aksi korporasi maupun pemerintah.

Berdasarkan **Tabel 2**, fitur yang akan diprediksi pada penelitian ini adalah Adjusted Close. Model akan memprediksi Adjusted Close berdasarkan fitur lainnya. Perlu diperhatikan bahwa fitur Close Before digunakan karena harga Close pada hari yang sama belum ada sebab sesi perdagangan belum berakhir.

2.2. Preprocessing

Tahap *preprocessing* akan mengubah dataset menjadi data yang siap untuk dilakukan *training*. Pada ini data akan melalui proses meliputi *handling missing values*, *features selection*, *scaling*, dan *cross validation*. Seluruh tahapan tersebut dilakukan untuk mempersiapkan data menjadi optimal sebelum membangun model.

2.2.1. Handling Missing Values

Tahap *handling missing values* bertujuan untuk melakukan cek dan mengisi maupun menghapus baris yang memiliki nilai kosong. Berdasarkan penjelasan dari fitur di awal, fitur Close Before di dapat dari kolom Close yang digeser ke bawah sebanyak 1 baris. Maka dari itu, baris pertama dataset akan dihapus karena nilai Close Before kosong.

2.2.2. Features Selection

Tahap *features selection* digunakan untuk mengetahui fitur apa saja yang mempunyai hubung satu sama lain. Hal ini bertujuan untuk memilih fitur yang relevan agar menghasilkan performa model yang maksimal serta mengefisienkan komputasi. Metode yang digunakan adalah Pearson

Correlation. Metode ini menghitung korelasi antara setiap fitur dengan target variabel (Adjusted Close) dan juga antara setiap fitur lainnya. Nilai korelasi berkisar dari -1 hingga 1, dengan 0 menunjukkan tidak adanya korelasi dan -1 serta 1 menunjukkan korelasi sempurna negatif dan positif. Fitur dengan korelasi tinggi terhadap target akan dipilih sebagai fitur yang relevan. Fitur yang relevan beserta nilai korelasinya ditunjukkan pada **Tabel 3**.

Tabel 3. Nilai Korelasi

Fitur	Nilai Korelasi
Low	0.999908
Open	0.000800
High	0.999904
Close Before	0.99680

2.2.3. Scaling

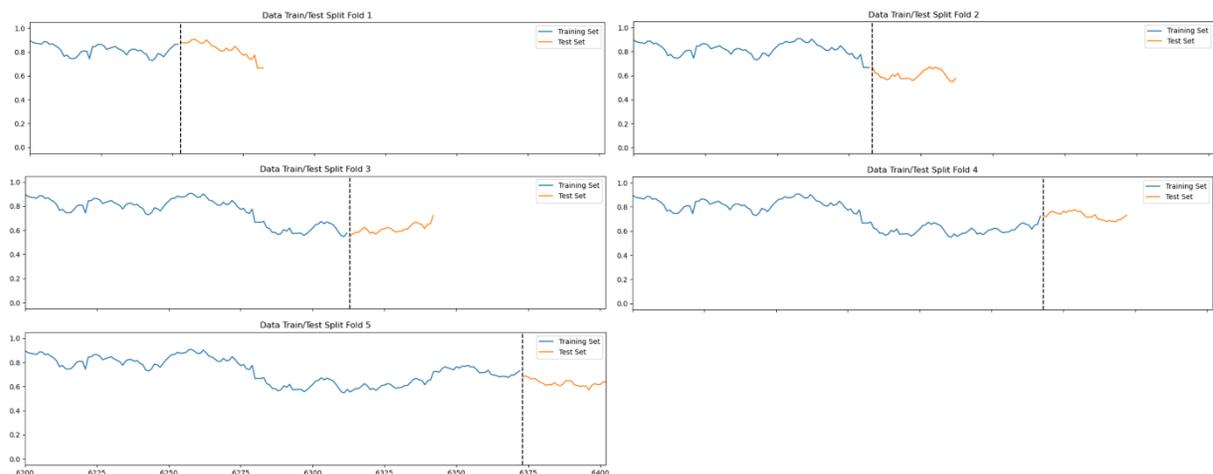
Tahap *scaling* adalah proses pengubahan skala dari nilai-nilai asli dalam data. Hal ini dilakukan untuk menghindari perbedaan skala yang besar antar fitur yang dapat mempengaruhi performa model. Metode yang digunakan pada tahap *scaling* adalah *Min Max Scaling*. Metode ini mengubah nilai fitur dalam rentang 0 hingga 1 dengan mempertahankan perbedaan nilai dalam rentang tersebut. Contoh data yang sudah dilakukan *scaling* ada pada **Tabel 4**.

Tabel 4. Contoh Data Hasil *Scaling*

Date	Low	Open	High	Close Before	Adjusted Close
1997-05-19	0.081250	0.088021	0.088542	0.086458	0.085417
1997-05-20	0.081771	0.086458	0.087500	0.085417	0.081771
1997-05-21	0.068750	0.081771	0.082292	0.081771	0.071354

2.2.4. Cross Validation

Tahap *cross validation* bertujuan untuk memastikan bahwa model mampu melakukan prediksi dengan baik pada baru yang belum pernah dilihat sehingga tidak terjadi bias. *Cross validation* pada data saham menggunakan *Time Series Split* yang membagi data *time series* menjadi beberapa bagian berdasarkan waktu. Data saham akan dibagi menjadi data *training* dan *testing* sebanyak 5 *fold*. Panjang data *testing* adalah 30 hari setelah data *training*. Grafik data *training* dan *testing* dapat dilihat pada **Gambar 2**.



Gambar 2. Data *Training* dan *Testing*

2.3. Membangun Model

Setelah data melewati *preprocessing*, tahap selanjutnya adalah membangun model untuk melakukan prediksi saham menggunakan algoritma XGBoost. Banyak parameter yang digunakan dalam algoritma XGBoost seperti *max_depth*, *learning_rate*, *n_estimator*, *subsample*, *gamma*, dan masih banyak lainnya. Pada penelitian ini parameter yang akan diuji meliputi *n_estimator* = 500, *early stopping round* = 3, *learning rate* = 0.01, 0.05, 0.1, dan *max_depth* (kedalaman pohon) = 3, 4, 5. Berdasarkan parameter yang akan diuji akan dihasilkan kombinasi sembilan model yang berbeda.

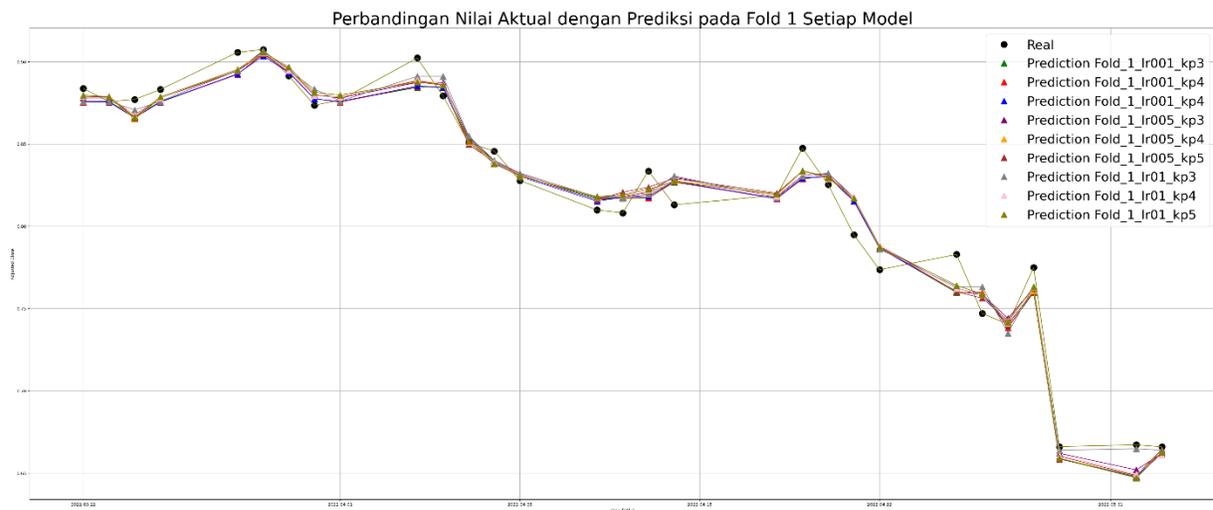
2.4. Evaluasi

Evaluasi dilakukan untuk menentukan parameter dan fitur yang paling optimal untuk melakukan prediksi pergerakan harga saham. Parameter terbaik ditentukan menggunakan metrik *root mean squared error* (RMSE) yang terkecil dari kesembilan model yang dibangun. RMSE yang dimaksud adalah rata-rata RMSE dari kelima *fold* yang telah terbentuk pada tahap *cross validation*. Fitur yang paling optimal ditentukan dengan mencari nilai *importance* tertinggi di antara fitur dari sembilan model.

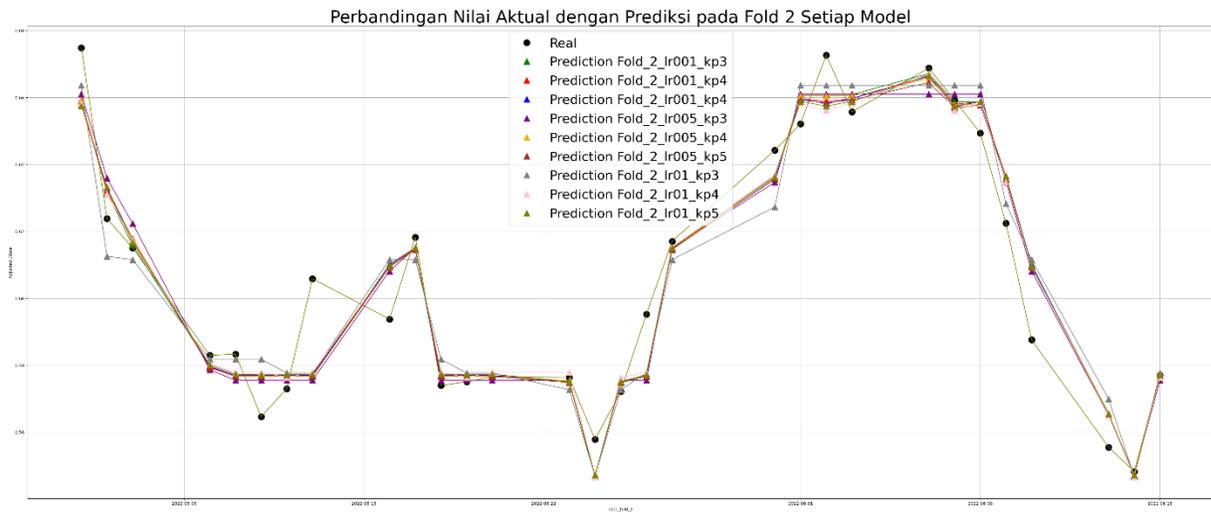
3. Hasil dan Pembahasan

3.1. Parameter Paling Optimal

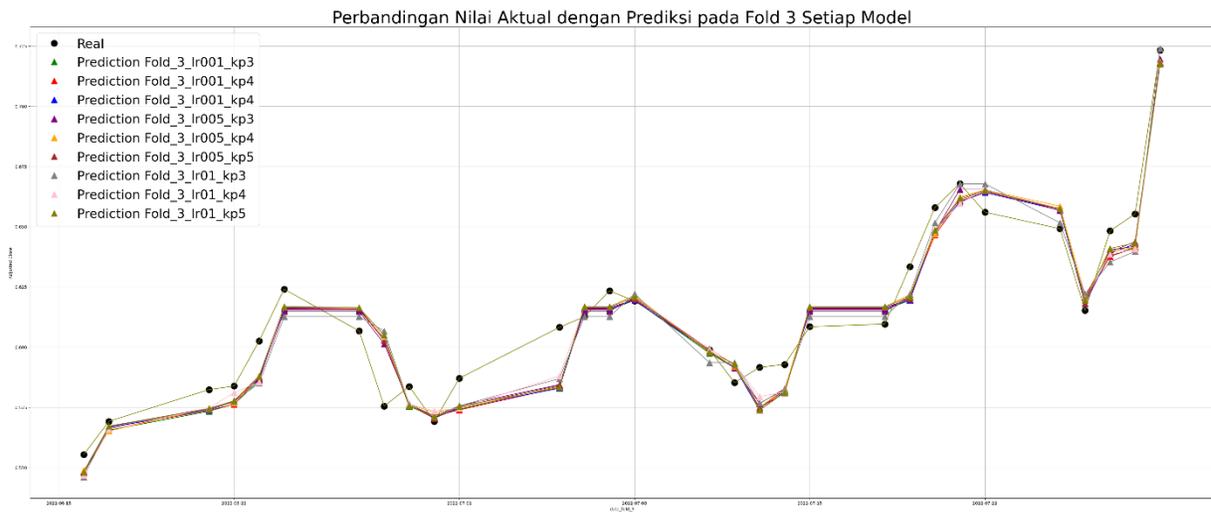
Model dibangun dengan mencoba semua parameter yang telah ditentukan untuk memprediksi semua *fold* yang telah dibentuk pada tahap *cross validation*. Pada setiap *fold* menghasilkan sembilan hasil prediksi dengan parameter yang berbeda-beda, maka total jumlah hasil prediksi yakni 45 prediksi. Hasil prediksi divisualisasikan menggunakan grafik pada gambar di bawah ini beserta penjelasan tentang singkatan parameter pada **Tabel 5**.



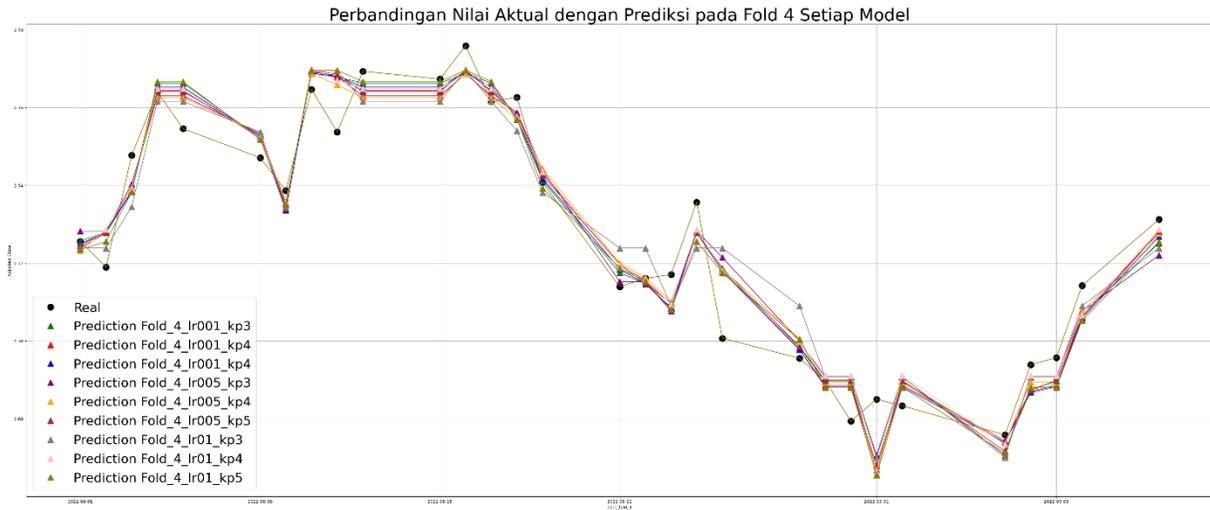
Gambar 3. Hasil Prediksi pada *Fold* 1



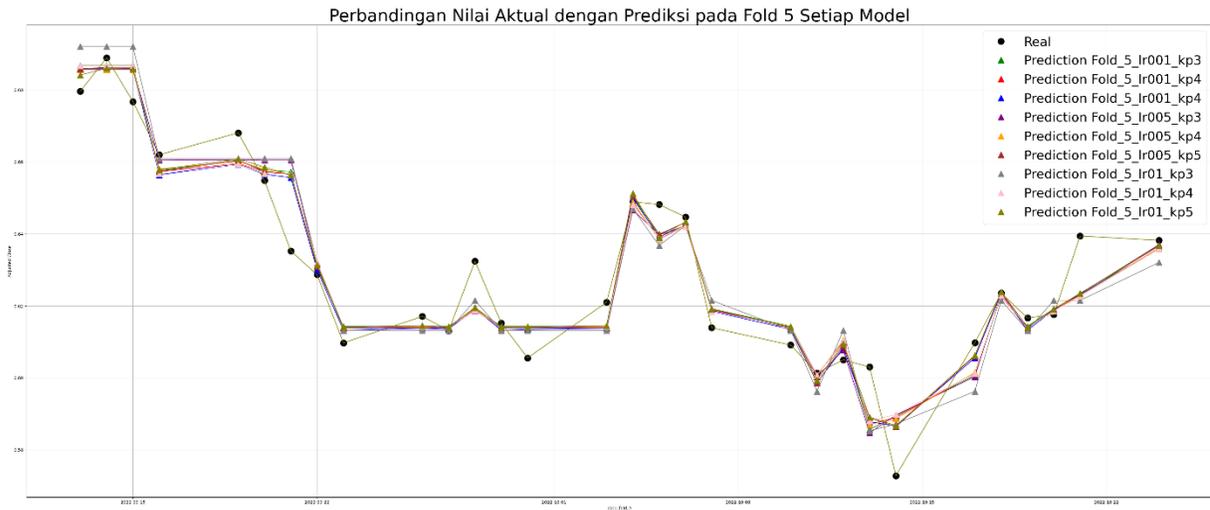
Gambar 4. Hasil Prediksi pada *Fold 2*



Gambar 5. Hasil Prediksi pada *Fold 3*



Gambar 6. Hasil Prediksi pada *Fold 4*



Gambar 7. Hasil Prediksi pada *Fold 5*

Tabel 5. Penjelasan Singkatan Parameter

Singkatan	Penjelasan
lr001	Learning rate sebesar 0,01
lr005	Learning rate sebesar 0,05
lr01	Learning rate sebesar 0,1
kp3	Kedalaman pohon sebanyak 3
kp4	Kedalaman pohon sebanyak 4
kp5	Kedalaman pohon sebanyak 5

Setelah hasil prediksi diketahui, evaluasi untuk menentukan parameter yang paling optimal adalah menggunakan *root mean squared error* (RMSE). Hasil akhir RMSE adalah rata-rata RMSE dari semua *fold* pada parameter yang sama. Penjelasan lebih lengkap ada pada **Tabel 6** di bawah.

Tabel 6. Hasil Evaluasi Menggunakan RMSE

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Rata-rata
RMSE lr001 kp3	0.011270	0.010456	0.011038	0.007673	0.008482	0.009784
RMSE lr001 kp4	0.011146	0.010447	0.011016	0.007420	0.008323	0.009670
RMSE lr001 kp5	0.010932	0.010645	0.010974	0.007675	0.007825	0.009610
RMSE lr005 kp3	0.010530	0.010725	0.010318	0.007761	0.008786	0.009624
RMSE lr005 kp4	0.010689	0.010400	0.010859	0.007512	0.008251	0.009542
RMSE lr005 kp5	0.010498	0.010576	0.010687	0.007684	0.007742	0.009437
RMSE lr01 kp3	0.007742	0.007742	0.011016	0.009040	0.009712	0.010306
RMSE Lr01 kp4	0.010260	0.010413	0.010686	0.007581	0.008402	0.009468
RMSE Lr01 kp5	0.010323	0.010622	0.010942	0.008015	0.007717	0.009524

Berdasarkan tabel hasil evaluasi di atas, rata-rata RMSE setiap model berbeda satu dengan lainnya meskipun perbedaan tersebut sangat kecil. Hal ini berarti perbedaan parameter bisa mengakibatkan perubahan pada performa model. Semakin dalam kedalaman pohon maka nilai RMSE semakin rendah. Pergantian parameter *learning rate* sangat berpengaruh terhadap nilai RMSE. *Learning rate* rendah dengan kedalaman tinggi menghasilkan nilai RMSE lebih besar daripada *learning rate* tinggi dengan kedalaman rendah. Sebagai contoh *learning rate* 0,05 dengan kedalaman 5 menghasilkan nilai RMSE lebih kecil yaitu 0,009437 daripada *learning rate* 0,1 dengan kedalaman pohon 3 yang menghasilkan nilai RMSE 0,010306. Model yang menghasilkan nilai RMSE terendah adalah model dengan parameter *learning rate* 0,05 dan kedalaman pohon 5 yang menghasilkan nilai RMSE sebesar 0,009437.

3.2. Fitur Paling Optimal

Penelitian ini menggunakan empat fitur seperti pada **Tabel 3** yaitu Low, Open, High, Close Before untuk memprediksi Adjusted Close sebagai label. XGBoost yang merupakan algoritma berbasis *ensemble* yang berarti beberapa *decision tree* yang berbeda digabungkan untuk meningkatkan kinerja prediksi. Nilai *importance* digunakan untuk memahami setiap fitur yang digunakan oleh model dan menentukan mana yang paling penting untuk membuat prediksi yang akurat. Salah satu metrik untuk menghitung nilai *importance* adalah menggunakan *F-score*. Berikut merupakan hasil nilai *importance* semua *fold* pada model lr005 kp5.

Tabel 7. Nilai *Importance* pada Model lr005 kp5

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
F-score Low	1437	866	1259	871	1003
F-score High	1229	696	1057	775	888
F-score Open	382	65	266	109	167
F-score Close Before	326	61	233	80	116

Semua *fold* menghasilkan nilai *importance* yang dengan urutan sama yaitu dari tertinggi fitur Low, High, Open, dan Close Before. Fitur Low mendapat nilai *importance* paling tinggi pada semua *fold*. Maka dari itu, fitur Low sangat berperan besar terhadap model untuk melakukan prediksi pergerakan harga saham.

4. Kesimpulan dan Saran

Penelitian ini bertujuan untuk mencari parameter terbaik dan fitur berpengaruh untuk hasil yang optimal algoritma XGBoost dalam memprediksi pergerakan harga saham. Parameter terbaik dicari menggunakan hasil RMSE terkecil setiap model. Pada penelitian ini parameter yang diuji meliputi $n_estimator = 500$, $early\ stopping\ round = 3$, $learning\ rate = 0.01, 0.05, 0.1$, dan max_depth (kedalaman pohon) = 3, 4, 5. Evaluasi memberikan hasil nilai RMSE terendah sebesar 0,009437 pada *learning rate* sebesar 0,05 dan kedalaman pohon sebanyak 5. Fitur yang paling berpengaruh didapat dari nilai *importance* yang dihitung berdasarkan *F-score*. Hasil evaluasi nilai *importance* pada semua *fold* sepakat bahwa fitur Low dalam dataset saham AMZN sangat mempengaruhi *output* prediksi.

Adapun saran untuk penelitian selanjutnya adalah untuk mencoba menggunakan data *testing* yang lebih panjang. Penelitian ini menggunakan data *testing* 30 hari setelah data *training*. Parameter pada penelitian selanjutnya juga bisa lebih menggunakan banyak variasi ataupun menggunakan metode-metode pencarian parameter terbaik seperti *grid search*. Selain itu, data saham juga bisa diganti dengan variasi yang lebih stabil maupun fluktuatif yang nantinya bisa menghasilkan prediksi lebih akurat.

Daftar Pustaka

- [1] Y. Ramdhani and A. Mubarak, "Analisis Time Series Prediksi Penutupan Harga Saham Antm.Jk Dengan Algoritma SVM Model Regresi," *JURNAL RESPONSIF*, vol. 1, no. 1, 2019, [Online]. Available: <http://ejurnal.univbsi.id/index.php/jti>
- [2] A. Mulfita, I. Yusra, S. Tinggi, and I. E. Kbp, "ANALISIS REGRESI DATA PANEL TERHADAP LIKUIDITAS SAHAM DI INDONESIA."
- [3] W. Y. Rusyida and V. Y. Pratama, "Prediksi Harga Saham Garuda Indonesia di Tengah Pandemi Covid-19 Menggunakan Metode ARIMA," *Square : Journal of Mathematics and Mathematics Education*, vol. 2, no. 1, p. 73, Apr. 2020, doi: 10.21580/square.2020.2.1.5626.
- [4] I. Puspitasari and Y. Wilandari, "ANALISIS INDEKS HARGA SAHAM GABUNGAN (IHSG) DENGAN MENGGUNAKAN MODEL REGRESI KERNEL," 2012. [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/gaussian>
- [5] J. Wu, K. Xu, X. Chen, S. Li, and J. Zhao, "Price graphs: Utilizing the structural information of financial time series for stock prediction," Jun. 2021, [Online]. Available: <http://arxiv.org/abs/2106.02522>
- [6] A. FAUZI, "FORECASTING SAHAM SYARIAH DENGAN MENGGUNAKAN LSTM," *Al-Masraf : Jurnal Lembaga Keuangan dan Perbankan*, vol. 4, no. 1, p. 65, Jun. 2019, doi: 10.15548/al-masraf.v4i1.235.
- [7] E. H. Yulianti, O. Soesanto, and Y. Sukmawaty, "Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit," *JOMTA Journal of Mathematics: Theory and Applications*, vol. 4, no. 1, 2022.

- [8] M. Ridwan, “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier”, doi: <https://doi.org/10.21776/jeccis.v7i1.204>.
- [9] P. Meilina, “PENERAPAN DATA MINING DENGAN METODE KLASIFIKASI MENGGUNAKAN DECISION TREE DAN REGRESI,” 2015.
- [10] S. Lorena, B. Ginting, W. Zarman, and I. Hamidah, “DATA MINING UNTUK MEMPREDIKSI MASA STUDI MAHASISWA BERDASARKAN DATA NILAI AKADEMIK,” 2014.
- [11] V.G.Utomo, N.Wakhidah, and A.N.Putri, “Prediksi Harga Saham Dengan Svm(Support Vector Machine) Dan Pemilihan Fitur F-Score,” 2020.
- [12] J. Fan *et al.*, “Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China,” *Energy Convers Manag*, vol. 164, pp. 102–111, May 2018, doi: 10.1016/j.enconman.2018.02.087.
- [13] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, vol. 13-17-August-2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [14] A. Bijaksana Putra Negara, “Penerapan Algoritma Model Regresi pada Angka New Active Cases Covid-19 di Indonesia,” vol. 01, no. 1, 2022, doi: 10.26418/juara.v1i1.59941.
- [15] N. Nyoman, P. Pinata, M. Sukarsa, N. Kadek, and D. Rusjyanthi, “Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python.”