

Analisis Kinerja XGBoost Menggunakan Bayesian Optimization dalam Prediksi Harga Ethereum

Christian Valentino^{a1}, Luh Arida Ayu Rahning Putri^{a2}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹valentino.2308561081@student.unud.ac.id
²rahningputri@unud.ac.id

Abstract

Cryptocurrency is a digital innovation in the financial sector that has revolutionized the global transaction system through blockchain technology. One of the main challenges in the crypto domain today is determining the price of cryptocurrencies, which are highly volatile. Ethereum, one of the largest cryptocurrencies, exhibits complex volatility patterns that require a robust predictive system. This study aims to compare the performance of the standard XGBoost algorithm with XGBoost optimized using Bayesian Optimization in predicting daily Ethereum prices based on time series data from 2016 to June 2025. The dataset includes price-related features such as open, high, low, volume, and percentage price change. The modeling process consists of several stages including feature engineering, time series-based data splitting, and model training. Model performance was evaluated using three primary metrics: MAE, RMSE, and R² Score. The evaluation results show that the standard XGBoost model achieved an MAE of 80.8926 (3.12%), RMSE of 114.1457 (4.40%), and an R² Score of 0.9723. Meanwhile, the optimized model using Bayesian Optimization achieved an MAE of 70.7241 (2.73%), RMSE of 102.5334 (3.96%), and an R² Score of 0.9777. These results indicate that Bayesian Optimization helps improve the model's prediction accuracy. This study concludes that the XGBoost model with a Bayesian Optimization not only delivers superior predictive performance for Ethereum price forecasting but also contributes scientifically by providing empirical evidence of Bayesian Optimization's effectiveness in enhancing model performance for time series-based cryptocurrency prediction and offering a reproducible framework for future financial forecasting research.

Keywords: XGBoost, Bayesian Optimization, Ethereum Price Prediction

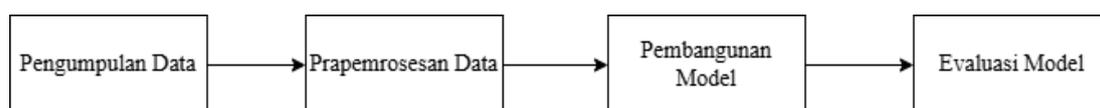
1. Pendahuluan

Kripto merupakan sebuah inovasi digital yang telah merevolusi ekosistem transaksi global dengan teknologi *blockchain*. Seiring meningkatnya minat masyarakat terhadap mata uang kripto seperti Bitcoin dan Ethereum, kebutuhan akan pemahaman dan prediksi pergerakan harga semakin penting. Khususnya pada Ethereum, sebuah mata uang kripto dengan pola volatilitas yang cukup kompleks. Sehingga menjadikannya sebuah tantangan dalam peramalan harga harian. Berbagai pendekatan *machine learning* telah dicoba untuk menjalankan tugas prediksi harga mata uang kripto. Namun, tuning model hingga efisiensi komputasi masih menjadi kendala hingga saat ini [1]. Algoritma XGBoost menjadi salah satu yang efektif dalam regresi dan klasifikasi pada data berdimensi tinggi dengan tingkat variabel yang sangat kompleks [2]. Namun, performa dari algoritma tersebut sangat bergantung pada pemilihan hyperparameternya. Untuk mengatasi permasalahan tersebut, penelitian ini menggunakan *Bayesian Optimization* sebagai cara mencari parameter yang lebih efisien. Berbanding dengan *grid search* dan *random search*, metode *Bayesian Optimization* memanfaatkan fungsi obyektif untuk membangun model probabilistik. Model ini akan memperbarui hasil prediksi berdasarkan hasil yang didapat sebelumnya. Sehingga didapatkan kombinasi parameter terbaik dengan lebih cepat [3].

Pada penelitian terdahulu, algoritma XGBoost telah diterapkan menggunakan pendekatan regresi untuk memprediksi harga logistik tertentu di suatu wilayah, dan menunjukkan performa yang sangat signifikan [4]. Pada penelitian lainnya, Bayesian Optimization telah digunakan sebagai metode tuning dalam memprediksi indeks Dolar Amerika, dan metode ini juga direkomendasikan pada model yang kompleks seperti XGBoost karena tingkat efisiensinya dalam mengurangi kesalahan prediksi [5]. Namun, belum banyak penelitian yang secara langsung menguji pengaruh *Bayesian Optimization* terhadap kinerja model XGBoost dalam konteks prediksi harga Ethereum yang berbasis data deret waktu. Mengacu pada latar belakang tersebut, studi ini dimaksudkan untuk mengevaluasi kinerja ataupun performa dari model XGBoost yang dioptimasi oleh Bayesian Optimization dalam melakukan tugas prediksi harga harian Ethereum. Evaluasinya akan dilakukan dengan bantuan tiga metrik yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R² score* untuk melihat tingkat akurasi model. Studi ini diharapkan dapat memberikan kontribusi yang signifikan dalam kemajuan teknologi khususnya sistem prediksi harga mata uang kripto yang akurat.

2. Metode Penelitian

Penelitian ini dilakukan melalui beberapa tahapan sistematis untuk menciptakan model prediksi yang akurat. Setiap tahapan saling terhubung dan disusun secara berurutan, guna memastikan proses berjalan dengan efektif. Untuk menggambarkan alur kerja penelitian, dirancang diagram proses pada **Gambar 1**.



Gambar 1. Alur Penelitian

Berdasarkan diagram pada **Gambar 1**, ditunjukkan beberapa tahapan utama penelitian yang mencakup pengumpulan data, prapemrosesan data, pembangunan model, dan diakhiri dengan evaluasi kinerja model yang dirancang.

2.1 Pengumpulan Data

Dataset yang digunakan terdiri dari data sekunder yang diperoleh dari situs Investing.com yaitu data harga harian mata uang kripto Ethereum dari tahun Maret 2016 sampai Juni 2025. Dataset yang didapat berformat file “.csv” terdiri dari 3378 baris data dan 7 atribut dengan deskripsi seperti pada **Tabel 1**.

Tabel 1. Atribut Dataset

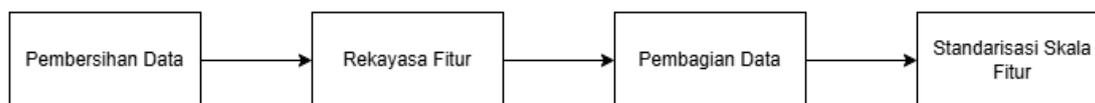
No	Atribut	Deskripsi
1	Date	Tanggal observasi harga Ethereum
2	Price	Harga penutupan harian Ethereum dalam USD. Nilainya menjadi referensi utama dalam memprediksi pergerakan harga.
3	Open	Harga pembukaan harian Ethereum dalam USD. Nilai ini menunjukkan titik awal transaksi dalam satu hari.
4	High	Harga tertinggi yang diraih dalam satu hari.
5	Low	Harga terendah yang diraih dalam satu hari.
6	Vol.	Volume perdagangan yang merepresentasikan total aktivitas trading yang terjadi dalam satu hari.
7	Change%	Persentase perubahan harga dalam satu hari dibandingkan dengan hari sebelumnya.

Tabel 2. Dataset Harga Ethereum dalam USD

Date	Price	Open	High	Low	Vol.	Change%
10/03/2016	11.75	11.20	11.85	11.07	0.00K	4.91%
11/03/2016	11.95	11.75	11.95	11.75	0.18K	1.70%
12/03/2016	12.92	11.95	13.45	11.95	0.83K	8.12%
13/03/2016	15.07	12.92	15.07	12.92	1.30K	16.64%
14/03/2016	12.50	15.07	15.07	11.40	92.18K	- 17.05%
...
04/06/2025	2607.13	2591.81	2675.16	2583.93	578.93K	0.54%
05/06/2025	2414.36	2607.10	2640.14	2394.08	900.97K	- 7.39%
06/06/2025	2475.34	2414.29	2528.84	2385.68	625.00K	2.53%
07/06/2025	2524.66	2476.45	2541.91	2457.03	286.60K	1.99%
08/06/2025	2506.81	2524.66	2525.28	2419.16	266.67K	- 0.71%

2.2 Prapemrosesan Data

Tahap prapemrosesan data mencakup tahapan – tahapan penting yang bertujuan untuk mempersiapkan data agar kemudian layak digunakan pada model. **Gambar 2** berikut menunjukkan diagram alur dari proses – proses yang terjadi pada prapemrosesan yang dilakukan pada penelitian.



Gambar 2. Alur Prapemrosesan Data

a. Pembersihan Data

Tahap pertama dari prapemrosesan data adalah pembersihan dataset yang bertujuan untuk memastikan kualitas serta kelengkapan dari data yang akan digunakan. Dataset yang diperoleh dari situs Investing.com memiliki beberapa nilai kosong. Untuk mengatasi hal tersebut, seluruh entri data yang memiliki nilai NaN akan dihapus. Setelah itu, indeks data direset kembali agar struktur dataset tetap konsisten. Proses ini memastikan agar model tidak menerima input dengan nilai kosong.

b. Rekayasa Fitur

Rekayasa fitur dilakukan untuk membantu model dalam menangkap pola dari data. Pada penelitian ini, dibuat beberapa fitur tambahan berdasarkan nilai masa lalu, seperti Price_lag1, Price_lag2, Vol_lag1, Open_lag1, dan Low_lag1. Setiap fitur tersebut merepresentasikan nilai data dari satu hari sebelumnya. Selain itu, dibuat pula fitur MA3 dan MA5 yang menyimpan rata – rata pergerakan harga dalam 3 dan 5 hari terakhir. Fitur – fitur tersebut dirancang untuk membantu model dalam memahami tren jangka pendek.

c. Pembagian Data

Setelah pembentukan fitur baru, data dipisahkan menjadi dua kategori yaitu data *training* atau latih sebesar 80% dan data *testing* atau uji sebesar 20%. Karena dataset yang digunakan berupa data *time series* perlu dilakukan pembagian secara tidak acak dan sesuai dengan urutan waktunya. Pembagian data ini menggunakan pendekatan *TimeSeriesSplit* dari *library* Scikit-learn. Dengan pendekatan ini, model yang dilatih hanya menggunakan data dari masa lalu dan diuji dengan data masa depan, pendekatan ini

membantu dalam menghindari kebocoran data serta meningkatkan akurasi model.

d. Standarisasi Skala Fitur

Tahap terakhir dari prapemrosesan adalah standarisasi skala data. Proses ini perlu dilakukan karena model seperti XGBoost sangat sensitif terhadap skala fitur input. Proses standarisasi ini mengubah distribusi panjang nilai pada setiap fitur menjadi memiliki rata – rata nol dengan standar deviasi satu. Data latih dan data uji sama – sama ditransformasi agar skala keduanya konsisten satu sama lain. Proses ini membantu meningkatkan stabilitas model saat melakukan pelatihan.

2.3 Pembangunan Model

Dalam melakukan pelatihan model, penelitian ini menggunakan dua pendekatan utama, yaitu menggunakan XGBoost sebagai model utama dalam melakukan prediksi dan menggunakan Bayesian Optimization dalam melakukan tuning hyperparameter untuk meningkatkan performa model.

a. XGBoost

XGBoost atau *Extreme Gradient Boosting* merupakan suatu algoritma *machine learning* yang berbasis pohon keputusan yang dikembangkan untuk meningkatkan akurasi dari metode *Gradient Boosting*. Algoritma ini bekerja dengan membentuk pohon keputusan baru dari kesalahan prediksi sebelumnya secara berulang – ulang. Proses ini membantu model belajar secara bertahap sampai mencapai akurasi tertinggi. Keunggulan utama dari XGBoost terletak pada tingkat efisiensinya dalam menangani data dengan ukuran besar, dengan fitur bawaan yaitu L1 dan L2 yang membantu meminimalisir resiko *overfitting*. Dalam tugas prediksi harga seperti Ethereum dengan sifat non – linear dan fluktuatif, algoritma ini sangat cocok karena mampu memahami pola pergerakan harga yang kompleks [6]. Pada penelitian ini, XGBoost diimplementasikan dengan fungsi *XGBRegressor()* dari pustaka *xgboost* dalam bahasa python. Penelitian ini membuat dua versi model, yakni model dengan parameter bawaan dan model dengan hasil optimasi oleh Bayesian Optimization. Cara ini memungkinkan peneliti dalam melakukan evaluasi kinerja dari model berdasarkan parameter yang berbeda.

b. Bayesian Optimization

Bayesian Optimization merupakan salah satu metode optimasi dengan berbasis probabilitas yang digunakan dalam mencari suatu kombinasi nilai parameter terbaik dari sebuah model dengan jumlah iterasi yang sedikit mungkin. Dalam penelitian ini, metode Bayesian digunakan untuk melakukan tuning pada nilai hyperparameter dari model XGBoost. Proses optimasi berjalan secara iteratif atau berulang dengan dua komponen, yaitu model Surrogate dan fungsi akuisisi.

- Model Surrogate
Model Surrogate digunakan untuk memodelkan atau merepresentasikan hubungan antara kombinasi dari nilai parameter dan performa model. Dalam metode Bayesian, model surrogate bekerja dengan membentuk distribusi probabilitas berdasarkan hasil evaluasi dari kombinasi parameter yang telah dicoba sebelumnya. Informasi – informasi yang didapat kemudian digunakan oleh model untuk menentukan kombinasi nilai parameter baru yang belum pernah dicoba sebelumnya [7]. Pendekatan ini mampu mengurangi jumlah percobaan yang diperlukan tanpa mengurangi tingkat akurasi model[8].
- Fungsi Akuisisi
Fungsi akuisisi digunakan untuk menentukan dan memilih titik atau kombinasi dari nilai parameter baru yang memiliki tingkat probabilitas tinggi dalam menghasilkan performa model yang lebih baik. Pada penelitian ini, fungsi akuisisi yang digunakan adalah *Expected Improvement* (EI). Fungsi ini akan menentukan peluang sebuah

kombinasi parameter akan memberikan peningkatan performa dari sebelumnya atau tidak. Proses ini membuat pencarian nilai parameter tidak dilakukan secara acak, melainkan berfokus satu bagian dengan melihat informasi historis. Dalam penelitian ini, titik – titik baru merujuk pada kombinasi nilai hyperparameter dari model XGBoost.

Pencarian nilai parameter dilakukan dengan menggunakan ruang pencarian yang telah ditentukan berupa rentang nilai pada masing – masing parameter. Penggunaan rentang membantu algoritma dalam mengeksplorasi banyak kemungkinan kombinasi parameter secara cepat.

Optimasi nilai parameter dilakukan menggunakan fungsi *BayesSearchCV* dari *library scikit-optimize* dari bahasa Python. Fungsi ini akan mengevaluasi setiap kombinasi parameter dengan metode 3 – *Fold Cross Validation* dan menggunakan metrik *Mean Absolute Error* sebagai fungsi objektif atau *scoring*. Proses akan berlangsung hingga mencapai jumlah iterasi maksimal yang telah ditentukan atau akan berhenti lebih awal apabila tidak terjadi peningkatan performa yang signifikan dalam beberapa iterasi secara berturut – turut. Hal ini menandakan bahwa kombinasi nilai parameter terakhir telah mencapai kondisi stabil dan tidak lagi memberikan peningkatan performa pada model. Penggalan kode yang digunakan ditunjukkan pada **Tabel 3**.

Tabel 3. Penggalan Kode Program Bayesian

Baris	Potongan Kode
1	search_spaces = {
2	'n_estimators': Integer(300, 500),
3	'max_depth': Integer(3, 5),
4	'learning_rate': Real(0.05, 0.1, prior='log-uniform'),
5	'subsample': Real(0.7, 0.9),
6	'colsample_bytree': Real(0.7, 0.9),
7	}
8	tscv = TimeSeriesSplit(n_splits=3)
9	
10	opt = BayesSearchCV(
11	estimator= XGBRegressor(objective='reg:squarederror', random_state=42),
12	search_spaces=search_spaces,
13	cv=tscv,
14	n_iter=25,
15	scoring='neg_mean_squared_error',
16	random_state=42
17)

2.4 Evaluasi Model

Tahapan terakhir dari penelitian ini adalah evaluasi model yang dirancang menggunakan data uji yang telah dibuat sebelumnya. Evaluasi model ini dilakukan menggunakan tiga metrik utama untuk menilai seberapa baik model dalam menjalankan tugas prediksi. Ketiga metrik tersebut antara lain.

a. Mean Absolute Error (MAE)

Mean Absolute Error atau MAE berfungsi sebagai pengukur nilai rata – rata dari selisih yang tercipta antara hasil prediksi dari model dengan nilai riilnya. Metrik ini akan memberikan besaran kesalahan pada setiap satuan target. Nilai yang diberikan oleh MAE tidak menunjukkan arah kesalahan baik kearah positif ataupun kearah negatif. Berikut merupakan persamaan matematika dari *Mean Absolute Error*.

$$MAE = \frac{1}{n} \sum_i^n |y_i - \hat{y}_i| \quad (1)$$

Keterangan :

- n : banyaknya sampel data yang diuji
- y_i : nilai sebenarnya pada sampel data ke – i
- \hat{y}_i : nilai prediksi dari model pada sampel data ke – i
- $|y_i - \hat{y}_i|$: nilai absolut dari selisih antara nilai sebenarnya dan nilai prediksi

b. Root Mean Squared Error (RMSE)

Root Mean Squared Error atau RMSE digunakan untuk menghitung nilai akar dari rata – rata kuadrat simpangan antara nilai hasil prediksi model dan nilai riilnya. Metrik ini sangat cocok digunakan untuk model yang dirancang untuk prediksi harga karena sangat sensitif dengan kesalahan yang ekstrem dan mampu untuk menangkap variasi yang besar. Berikut merupakan persamaan matematika dari *Root Mean Squared Error*.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

Keterangan :

- n : banyaknya sampel data yang diuji
- y_i : nilai sebenarnya pada sampel data ke – i
- \hat{y}_i : nilai prediksi dari model pada sampel data ke – i
- $(y_i - \hat{y}_i)^2$: nilai kuadrat dari selisih antara nilai sebenarnya dan nilai prediksi model

c. Koefisien Determinasi (R²)

Koefisien determinasi atau *R² score* digunakan untuk menghitung dan menunjukkan banyak atau besarnya variasi data yang dapat dipahami oleh model. Rentang nilainya berkisar pada 0 sampai 1 di mana apabila angka yang dihasilkan semakin mendekati angka 1, maka model yang dirancang memiliki hasil prediksi yang semakin baik. Berikut merupakan persamaan matematika dari koefisien determinasi atau *R²*.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

Keterangan :

- y_i : nilai sebenarnya pada sampel data ke – i
- \hat{y}_i : nilai prediksi dari model pada sampel data ke – i
- \bar{y} : nilai rata – rata dari semua target
- $\sum (y_i - \hat{y}_i)^2$: jumlah kuadrat dari kesalahan prediksi
- $\sum (y_i - \bar{y})^2$: jumlah dari variasi data yang ada

3. Hasil dan Pembahasan

Pelatihan model pertama dilakukan dengan menggunakan algoritma XGBoost tanpa melalui proses optimasi pada parameternya. Kombinasi parameter ini dipilih berdasarkan referensi dan praktik umum dasar dalam penerapan XGBoost, khususnya untuk menangani data *time series*. Nilai – nilai parameter ini digunakan sebagai *baseline* untuk membandingkan seberapa jauh peningkatan performa yang dicapai setelah proses tuning. Nilai parameter pada model awal

XGBoost ditunjukkan pada **Tabel 3**.

Tabel 3. Parameter Awal Model XGBoost

Parameter	Nilai
n_estimator	100
max_depth	2
learning_rate	0.3
subsample	1.0
colsample_bytree	1.0

Model ini selanjutnya dilatih menggunakan data latih dan diuji dengan data uji. Evaluasinya dilakukan dengan tiga metrik utama yaitu MAE, RMSE, dan R^2 Score. Hasil evaluasi pada model awal ditunjukkan pada **Tabel 4**.

Tabel 4. Hasil Evaluasi Model Awal

Metrik	Nilai	Persentase
MAE	80.8926	3.12%
RMSE	114.1457	4.40%
R^2	0.9723	97.23%

Meskipun hasil evaluasi dengan parameter *default* menunjukkan performa model yang cukup baik, hasil yang ditunjukkan pada **Tabel 4**. masih memperlihatkan adanya kesalahan prediksi rata – rata mencapai 80 dolar, dengan tingkat akurasi model sekitar 97% variasi data. Nilai akurasi tersebut menunjukkan bahwa model cukup baik dalam menangkap pola harga yang fluktuatif, namun masih bisa ditingkatkan kembali. Oleh karena itu, dilakukan tuning hyperparameter dengan pendekatan Bayesian Optimization. Rentang nilai parameter yang digunakan serta kombinasi nilai parameter terbaik ditunjukkan pada **Tabel 5**. Kombinasi parameter ini diperoleh pada iterasi ke – 22 dari total 25 iterasi yang dilakukan. Proses tuning dilakukan dengan menentukan kombinasi terbaik menggunakan model probabilistik yang dilihat berdasarkan hasil evaluasi sebelumnya. Evaluasi tiap – tiap kombinasi parameter dari setiap iterasi dilakukan dengan *3-fold cross-validation*, untuk memastikan hasil validasi yang sesuai dengan urutan kronologi dari data.

Tabel 5. Rentang Nilai Parameter dan Kombinasi Parameter Terbaik

Parameter	Rentang	Nilai Parameter Terbaik
n_estimators	300 - 500	319
max_depth	3 – 5	3
learning_rate	0.05 – 0.1	0.0500
subsample	0.7 – 0.9	0.8066
colsample_bytree	0.7 – 0.9	0.9000

Model XGBoost kemudian dilatih kembali menggunakan kombinasi nilai parameter terbaik yang sudah didapatkan. Kemudian evaluasi dilakukan menggunakan metrik yang sama, hasilnya ditunjukkan pada **Tabel 6**.

Tabel 6. Hasil Evaluasi Model Setelah Dioptimasi

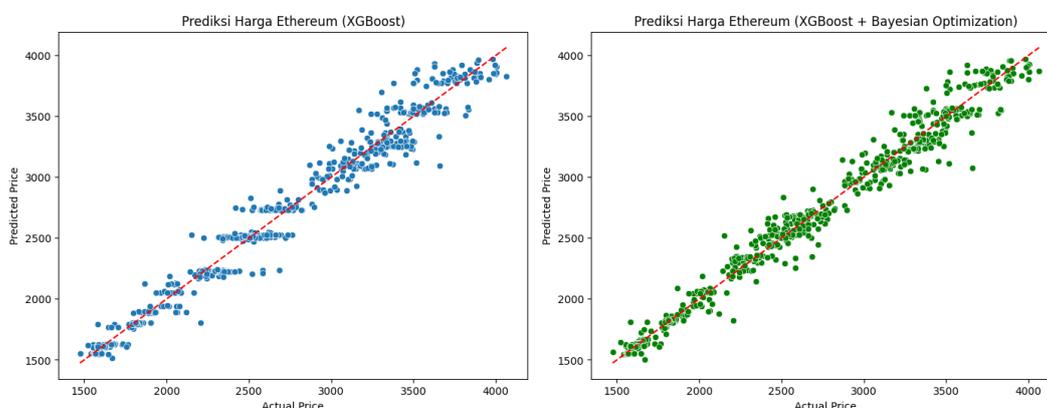
Metrik	Nilai	Persentase
MAE	70.7241	2.73%
RMSE	101.5334	3.96%
R ²	0.9777	97.77%

Setelah membandingkan hasil evaluasi pada **Tabel 4.** dan **Tabel 6.** terlihat bahwa model XGBoost mengalami peningkatan performa pada seluruh metrik setelah dioptimasi. Nilai MAE sebesar 70.7241 atau 2.73% menandai bahwa rata – rata kesalahan hasil prediksi harga meleset pada kisaran 70 dolar dari harga sebenarnya. Mengingat harga Ethereum yang mencapai angka ribuan dolar, nilai kesalahan ini tergolong kecil dan dapat dianggap cukup baik dalam konteks harga mata uang kripto yang fluktuatif. Sementara, pada nilai RMSE yang sebesar 102.5334 atau 3.96% menandai bahwa adanya beberapa kesalahan hasil prediksi yang meleset lebih dari nilai rata – rata yang didapat. Nilai dari RMSE ini sangat berguna dalam melihat apakah model memiliki nilai *outlier*, dan pada kasus kali ini nilai RMSE yang didapat tergolong aman. RMSE sangat penting digunakan pada data *time series* seperti harga Ethereum, karena volatilitas tajam pada pergerakan harga dapat mengganggu stabilitas model apabila tidak ditangani dengan baik.

Terakhir, nilai R² Score yang mencapai 0.9777 menunjukkan bahwa model yang dirancang mampu merepresentasikan sekitar 97.77% dari variasi data harga Ethereum. Angka ini membuktikan bahwa model tidak hanya cocok dengan data yang ada, tetapi juga dapat menangkap pola pergerakan harga dengan baik. Semakin tinggi nilai R², semakin baik model dalam memahami kolerasi antara fitur – fitur seperti harga, volume, dan tren jangka pendek.

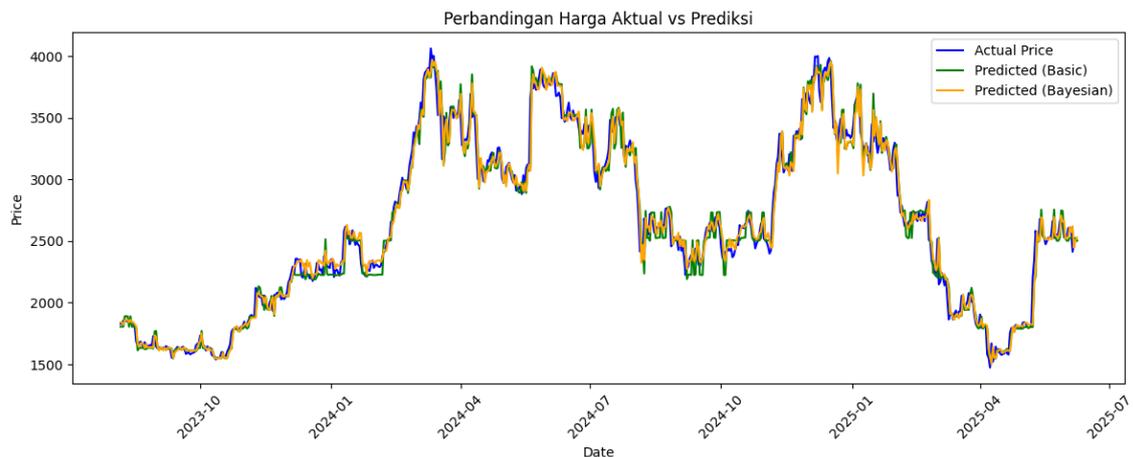
Kombinasi nilai parameter terbaik yang diperoleh dari proses tuning menunjukkan model yang bekerja optimal saat menggunakan *n_estimators* yang tinggi dengan kedalaman *max_depth* yang sedang. Nilai *learning_rate* yang rendah memberikan tingkat stabilitas ketika proses berlangsung. Sedangkan, nilai *subsample* dan *colsample_bytree* yang tinggi membantu meningkatkan kemampuan generalisasi model. Secara keseluruhan kombinasi tersebut membantu model dalam menyeimbangkan tingkat kompleksitas dan akurasi tanpa resiko *overfitting*.

Dibandingkan dengan metode seperti *Grid Search* ataupun *Random Search*, *Bayesian Optimization* memiliki keunggulan dalam menentukan kombinasi parameter. Metode ini memanfaatkan model probabilistik dari hari pembelajaran sebelumnya dan mengeksplor serta beradaptasi terhadap ruang pencarian. Sehingga jumlah iterasi yang dibutuhkan dalam mencari kombinasi terbaik menjadi lebih sedikit dan efisien dibandingkan dengan metode lainnya.



Gambar 3. Scatter Plot Hasil Prediksi Model XGBoost Sebelum dan Sesudah Optimasi

Scatter Plot pada **Gambar 3**, memperlihatkan hubungan antara sumbu x yaitu nilai atau harga sebenarnya dan sumbu y yaitu hasil prediksi. Grafik sebelah kiri menunjukkan prediksi XGBoost tanpa optimasi, sementara grafik sebelah kanan menunjukkan hasil prediksi dari model yang telah dioptimasi menggunakan Bayesian Optimization. Garis putus – putus merah menandakan hasil prediksi yang tepat dengan harga sebenarnya. Dapat dilihat pada *scatter plot* sebelah kanan, titik – titik yang muncul lebih rapat terhadap garis putus – putus dibandingkan dengan *scatter plot* sebelah kiri. Visualisasi ini menunjukkan tingkat deviasi yang lebih rendah serta distribusi eror yang lebih stabil pada model yang telah dioptimasi.



Gambar 4. Visualisasi Deret Waktu

Dari visualisasi pada **Gambar 4**, dapat dilihat perbandingan antara harga sebenarnya dengan hasil prediksi dari kedua model. Garis biru pada grafik merepresentasikan harga sebenarnya, garis hijau merepresentasikan hasil prediksi dari model pertama, dan garis oranye merepresentasikan hasil prediksi dari model yang dioptimasi. Terlihat pada grafik bahwa kedua model dapat menangkap pergerakan harga yang polanya cukup fluktuatif dengan baik, seperti lonjakan harga yang terjadi pada awal tahun 2024. Tetapi dari kedua garis yang menggambarkan hasil prediksi, garis oranye memiliki pergerakan yang lebih halus dan konsisten dalam mengikuti tren harga. Hasil ini menandai bahwa model yang dioptimasi mampu melakukan generalisasi dan adaptasi yang baik terhadap pergerakan harga yang sangat kompleks.

4. Kesimpulan

Penelitian ini membandingkan dan mengevaluasi performa yang dihasilkan oleh model XGBoost standar dengan XGBoost yang dioptimasi oleh Bayesian Optimization dalam memprediksi harga harian mata uang kripto yaitu Ethereum. Dengan memanfaatkan data dari tahun 2016 hingga Juni 2025, model XGBoost yang digunakan mampu mendapatkan nilai MAE sebesar 80.8926 atau sebesar 3.12%, RMSE sebesar 114.1457 atau sebesar 4.40%, dan R^2 Score sebesar 0.9723. Sementara itu, model yang dioptimasi mampu memberikan performa yang lebih baik dengan nilai MAE sebesar 70.7241 atau 2.73%, RMSE sebesar 102.5334 atau sebesar 3.96%, dan R^2 Score sebesar 0.9777. Visualisasi dan hasil evaluasi menunjukkan bahwa model yang dirancang mampu beradaptasi dan memahami pergerakan tren harga dengan baik meskipun terdapat fluktuasi. Secara keseluruhan, XGBoost yang dioptimasi dengan Bayesian terbukti efektif dalam menjalankan tugas peramalan harga mata uang kripto yang berbasis deret waktu. Kedepannya, disarankan untuk menambahkan fitur lain seperti sentimen pasar dan melakukan pelatihan model secara berkala agar model lebih peka terhadap dinamika dan fluktuasi pasar yang terus berubah.

Daftar Pustaka

- [1] I. S. Kervanci, M. F. Akay, and E. Özceylan, "Bitcoin price prediction using LSTM, GRU and hybrid LSTM-GRU with bayesian optimization, random search, and grid search for the next days," *Journal of Industrial and Management Optimization*, vol. 20, no. 2, pp. 570–588, 2024, doi: 10.3934/jimo.2023091.
- [2] T. K. Samson, "Comparative Analysis of Machine Learning Algorithms for Daily Cryptocurrency Price Prediction," *Information Dynamics and Applications*, vol. 3, no. 1, pp. 64–76, Mar. 2024, doi: 10.56578/ida030105.
- [3] V. Chithanuru and M. Ramaiah, "Proactive detection of anomalous behavior in Ethereum accounts using XAI-enabled ensemble stacking with Bayesian optimization," *PeerJ Comput Sci*, vol. 11, p. e2630, Mar. 2025, doi: 10.7717/peerj-cs.2630.
- [4] A. Purnomo, A. G. Putrada, R. Habibi, and S. Syafranita, "MDI and PI XGBoost regression-based methods: regional best pricing prediction for logistics services," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 22, no. 5, p. 1157, May 2024, doi: 10.12928/telkomnika.v22i5.26037.
- [5] A. F. Hamdani, D. Swanjaya, and R. Helilintar, "Facebook Prophet Model with Bayesian Optimization for USD Index Prediction," *JUITA : Jurnal Informatika*, vol. 11, no. 2, p. 293, Nov. 2023, doi: 10.30595/juita.v11i2.17880.
- [6] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [7] Y. Liu, B. Li, and H. Liang, "Building Electricity Prediction Using BILSTM-RF-XGBOOST Hybrid Model with Improved Hyperparameters Based on Bayesian Algorithm," *Electronics (Basel)*, vol. 14, no. 11, p. 2287, Jun. 2025, doi: 10.3390/electronics14112287.
- [8] E. S. Pour, H. Jafari, A. Lashgari, E. Rabiee, and A. Ahmadisharaf, "Cryptocurrency Price Prediction with Neural Networks of LSTM and Bayesian Optimization," *European Journal of Business and Management Research*, vol. 7, no. 2, pp. 20–27, Mar. 2022, doi: 10.24018/ejbmr.2022.7.2.1307.