

PREDIKSI PERGERAKAN HARGA ETHEREUM MENGGUNAKAN MACHINE LEARNING DENGAN ALGORITMA RANDOM FOREST DAN XGBOOST

I Made Candra Girinata^{1*}, Budi Styawan², Arwin Wahyu Saputra³, M Aidil Arif⁴, Arnoldus Janssen Dahur⁵

^{1,2,3,4,5} Teknologi Rekayasa Komputer Jaringan, Politeknik Negeri Tanah Laut

E-mail: candragirinata@politla.ac.id^{1*}, budistyawan@politla.ac.id², arwinwahyu@politla.ac.id³,
maidilarif@politla.ac.id⁴, arnoldusdahur@politla.ac.id⁵

INFO ARTIKEL

Sejarah Artikel

Diterima : 21/10/2025

Direvisi : 03/11/2025

Diterbitkan : 01/12/2025

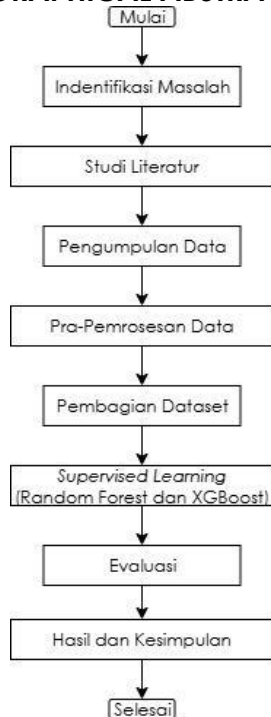
*Corresponding author

candragirinata@politla.ac.id

DOI: 10.70247/jumistik.v4i2.222

<https://ojs.amiklps.ac.id>

GRAPHICAL ABSTRACT



ABSTRACT

The rapid development of crypto assets, particularly Ethereum, demands accurate price prediction models to support investment strategies and risk management. This study aims to analyze and compare the performance of two ensemble machine learning algorithms, Random Forest (RF) and XGBoost, in predicting the daily price of Ethereum. A historical ETH/USD dataset of 3,423 observations from September 2016 to July 2025 was obtained from the Bitfinex platform. After undergoing data pre-processing and temporal feature engineering, the dataset was split with an 80:20 ratio for training and testing. The models were evaluated using Root Mean Square Error (RMSE) and the Coefficient of Determination (R^2). The experimental results show that XGBoost significantly outperformed Random Forest, with an RMSE of 134.63 and an R^2 of 0.958. In comparison, Random Forest yielded an RMSE of 208.45 and an R^2 of 0.899. This finding indicates that the boosting mechanism in XGBoost is more effective at capturing the complexity and volatility of crypto market data.

Keywords: Price Prediction, Ethereum, Machine Learning, XGBoost, Random Forest.

ABSTRAK

Perkembangan aset kripto yang pesat, khususnya Ethereum, menuntut adanya model prediksi harga yang akurat untuk mendukung strategi investasi dan manajemen risiko. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja dua algoritma machine learning ensemble, yaitu Random Forest (RF) dan XGBoost, dalam memprediksi harga harian Ethereum. Dataset historis ETH/USD sebanyak 3.423 observasi dari periode September 2016 hingga Juli 2025 diperoleh dari platform Bitfinex. Setelah melalui tahap pra-pemrosesan data dan rekayasa fitur temporal, dataset dibagi dengan rasio 80:20 untuk pelatihan dan pengujian. Model dievaluasi menggunakan metrik Root Mean Square Error (RMSE) dan Koefisien Determinasi (R^2). Hasil eksperimen menunjukkan bahwa XGBoost secara signifikan mengungguli Random Forest, dengan nilai RMSE 134.63 dan R^2 0.958. Sebagai perbandingan, Random Forest menghasilkan RMSE 208.45 dan R^2 0.899. Temuan ini mengindikasikan bahwa mekanisme boosting pada XGBoost lebih efektif dalam menangkap kompleksitas dan volatilitas data pasar kripto.

Kata kunci: Prediksi Harga, Ethereum, Machine Learning, XGBoost, Random Forest.

© 2025 Penerbit STMIK Amika Soppeng. All rights reserved

PENDAHULUAN

Perkembangan crypto asset dalam beberapa tahun terakhir telah menarik minat besar dari kalangan investor, pengembang, dan institusi keuangan. Banyaknya aplikasi yang muncul untuk memudahkan *crypto asset* belakangan ini menunjukkan adanya peningkatan minat pengguna yang signifikan di bidang *Cryptocurrency*. Fakta ini juga didukung dengan selarasnya jumlah kenaikan harga Bitcoin berbanding lurus dengan lonjakan unduhan aplikasi exchange dan jumlah pengguna baru yang sangat tinggi[1].

Peningkatan harga Bitcoin yang dipicu oleh pembelian sebagai aset investasi juga mencerminkan perkembangan *cryptocurrency* yang ditandai dengan inovasi platform seperti Ethereum, yang memperluas fungsi dan kegunaan teknologi blockchain. Jika Bitcoin menarik minat melalui nilai spekulatifnya, Ethereum menawarkan utilitas lebih jauh dengan *smart contracts* dan *dApps*, sehingga mendorong adopsi di sektor-sektor praktis seperti keuangan terdesentralisasi (DeFi) dan manajemen rantai pasok [2].

Decentralized Finance (DeFi) telah menunjukkan pertumbuhan luar biasa dengan Total Value Locked (TVL) melonjak dari \$50 juta pada tahun 2020 menjadi lebih dari \$100 miliar di tahun 2024, membuktikan daya tariknya sebagai alternatif terhadap sistem keuangan tradisional [3]. DeFi bukan hanya tren tetapi juga evolusi keuangan yang dengan dukungan teknologi dan kebijakan tepat dapat menjadi tulang punggung sistem keuangan masa depan, khususnya di jaringan *Ethereum*. Pertumbuhan ini juga didorong oleh kemampuan *Ethereum* dalam menyediakan infrastruktur yang aman dan terdesentralisasi untuk berbagai aplikasi keuangan, mulai dari *lending*, *borrowing*, hingga *yield farming*. Tren harga *Ethereum* sangat dipengaruhi oleh perkembangan dalam ekosistemnya, seperti pembaruan protokol dan perubahan jaringan, yang secara langsung berdampak pada utilitas dan adopsinya[4].

Dengan pertumbuhan kapitalisasi pasar yang meningkat mencapai ratusan miliar dolar dalam kurun waktu yang singkat, dapat disimpulkan pentingnya untuk memprediksi harga *Ethereum* yang akurat untuk manajemen risiko dan strategi investasi [5]. Volatilitas harga *cryptocurrency* yang tinggi menciptakan peluang sekaligus risiko besar bagi investor [6]. Prediksi yang tepat dapat membantu dalam pengambilan keputusan investasi, mitigasi kerugian, dan optimalisasi portofolio.

Beberapa penelitian sebelumnya telah mencoba memprediksi harga *cryptocurrency* menggunakan pendekatan statistik seperti *Autoregressive Integrated Moving Average* (ARIMA) namun, metode tradisional ini sering kali kurang efektif dalam menangkap pola non-linear dan dinamika pasar yang kompleks [7]. Oleh karena itu, pendekatan berbasis pembelajaran mesin (*machine learning*) mulai banyak digunakan karena kemampuannya dalam memproses data berskala besar dan mengidentifikasi pola tersembunyi dari data historis [8].

Di antara berbagai algoritma pembelajaran mesin, *Random Forest* dan *XGBoost* telah terbukti dapat digunakan dalam prediksi deret waktu finansial[9]. Algoritma *machine learning* seperti *Random Forest* dan *XGBoost* telah menjadi alat penting untuk memprediksi pergerakan harga aset digital [10], [11]. *Random Forest* bekerja dengan membuat banyak pohon keputusan secara bersamaan, dimana setiap pohon memberikan pendapatnya, kemudian hasil akhir ditentukan berdasarkan suara terbanyak. Cara ini membuat prediksinya lebih stabil dan mengurangi kesalahan. Keunggulan utamanya adalah kemampuannya bekerja dengan baik meskipun datanya tidak sempurna atau mengandung *noise* [12], [13]. *XGBoost* mengambil pendekatan yang berbeda. Algoritma ini membangun pohon keputusan satu per satu secara berurutan, dimana setiap pohon baru belajar dari kesalahan pohon sebelumnya. Proses ini membuat prediksinya semakin akurat seiring bertambahnya jumlah pohon [14].

Random Forest dikenal karena ketangguhannya terhadap *overfitting* dan kemampuannya menangani data dengan fitur beragam [13]. Sementara itu, *XGBoost* diakui keunggulannya dalam optimisasi kecepatan dan akurasi [14]. Namun, belum banyak penelitian yang secara khusus membandingkan kinerja kedua algoritma ini dalam memprediksi harga *Ethereum*.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma *Random Forest* dan *XGBoost* dalam memprediksi pergerakan harga *Ethereum*. Hasil penelitian ini diharapkan dapat memberikan kontribusi bagi investor, analis pasar, dan pengembang dalam menyusun strategi berbasis data yang lebih akurat. Selain itu, temuan ini juga dapat menjadi dasar pengembangan model prediktif yang lebih canggih untuk aset digital di masa depan.

TINJAUAN PUSTAKA

Perkembangan metode *machine learning* dalam analisis data kompleks telah terbukti efektif melalui berbagai penelitian sebelumnya. Studi yang dilakukan oleh Wulandari et al. [15] berhasil membandingkan kinerja Naive Bayes dan SVM untuk analisis sentimen isu politik di Twitter, sementara penelitian yang dilakukan oleh Rahayu et al. [16] mendemonstrasikan kemampuan SVM dalam mengklasifikasikan konten sensitif terkait terorisme. Meskipun berfokus pada data tekstual, temuan-temuan tersebut memberikan landasan berharga untuk pendekatan komparatif algoritma *machine learning*.

Melanjutkan tradisi akademik tersebut, penelitian ini mengalihkan fokus kepada prediksi harga *Ethereum* sebagai aset kripto yang memiliki karakteristik unik. Penelitian tentang prediksi harga mata uang kripto menggunakan pembelajaran mesin telah menjadi bidang yang mendapatkan perhatian secara signifikan seiring dengan volatilitas pasar yang ekstrem dan potensi keuntungan finansial yang

menjanjikan. Sifat pasar kripto yang non-linear, dipengaruhi oleh faktor teknis seperti volume perdagangan dan *moving averages*, fundamental (regulasi, adopsi institusional), serta sentimen sosial media, menuntut pendekatan pemodelan yang lebih canggih dibandingkan metode statistik tradisional. Dalam beberapa tahun terakhir, berbagai model pembelajaran mesin dari yang sederhana seperti *Linear Regression* hingga algoritma kompleks berbasis ensemble seperti *Random Forest* (RF) dan *XGBoost* telah dieksplorasi untuk menangkap kompleksitas ini.

Penelitian yang dibuat oleh Tian [17] mengeksplorasi prediksi harga dua aset kripto, yaitu *Binance Coin* (BNB) dan *Huobi Token* (HT), dengan memanfaatkan empat kelompok fitur yaitu indikator teknis, aset kripto lain, faktor makroekonomi, dan komoditas tradisional serta tiga model *machine learning* (*Random Forest* (RF), *Long Short-Term Memory* (LSTM), dan *XGBoost*) dalam empat jangka waktu prediksi (1, 3, 7, dan 30 hari). Metode penelitiannya melibatkan perbandingan performa berdasarkan dua aspek diantaranya evaluasi ketahanan terhadap outlier menggunakan selisih MAPE dan Median APE, dan evaluasi performa rata-rata dengan MSE. Hasilnya menunjukkan bahwa prediksi 1 hari unggul dalam stabilitas (selisih APE terendah) dan akurasi (MSE rata-rata terbaik: 0.0117), sementara prediksi 30 hari paling tidak stabil. Di antara model, LSTM memiliki konsistensi terburuk dengan selisih performa ekstrem tertinggi, sedangkan RF dan *XGBoost* menunjukkan ketahanan serupa terhadap outlier.

Penelitian selanjutnya yang dilakukan oleh Zuo [18] membandingkan kinerja lima model *machine learning* *Support Vector Machine* (SVM), *Random Forest*, *Neural Network*, *XGBoost*, dan *LightGBM* untuk memprediksi harga Bitcoin menggunakan dataset historis. Metode penelitiannya melibatkan pembagian dataset menjadi *training set* dan *test set* (rasio 7:3), dengan perubahan harga saham sebagai variabel target dan variabel lainnya sebagai input. Model dievaluasi menggunakan metrik MSE, RMSE, MAE, MAPE, dan R^2 untuk mengukur akurasi prediksi. Hasilnya menunjukkan bahwa *XGBoost* unggul dalam semua metrik evaluasi, diikuti secara berurutan oleh *LightGBM*, *Random Forest*, SVM, dan *Neural Network* yang memiliki kinerja terburuk dengan MSE puluhan kali lebih tinggi. Keunggulan *XGBoost* terletak pada kemampuannya menangani data berdimensi tinggi dan hubungan nonlinier, sementara *Neural Network* membutuhkan tuning lebih intensif. Simpulan penelitian menyatakan *XGBoost* sebagai pilihan optimal untuk prediksi harga Bitcoin, meskipun pemilihan model dalam praktik harus mempertimbangkan karakteristik data untuk hasil yang lebih baik.

Perbandingan hasil algoritma *XGBoost* dengan beberapa algoritma juga telah dilakukan didalam penelitian Terence et al. [19] yang membandingkan kinerja empat algoritma *Machine Learning* untuk memprediksi harga Bitcoin, yaitu *Long Short-Term Memory* (LSTM), *K-Nearest Neighbours* (KNN), *Random Forest Classifier* (RFC), dan *Extreme Gradient Boosting* (*XGBoost*). Tujuan penelitian ini adalah membantu investor dan trader dalam pengambilan keputusan dengan menyediakan

prediksi harga yang akurat. Metode penelitian menggunakan data historis Bitcoin seperti volume, harga open, high, dan low sebagai variabel input. Evaluasi performa dilakukan berdasarkan parameter akurasi, *precision*, *recall*, dan F1-score. Hasil penelitian menunjukkan bahwa *XGBoost* mencapai akurasi di atas 92% dan unggul dalam semua parameter evaluasi dibandingkan ketiga model lainnya, sehingga terbukti sebagai algoritma paling efektif untuk prediksi harga Bitcoin dalam studi ini.

Studi penggunaan algoritma *Random Forest* juga telah dilakukan oleh Asbullah et al. [20] untuk memprediksi harga harian *cryptocurrency* *Binance* dengan menggunakan dataset yang mencakup 1.992 observasi dari periode 1 Januari 2018 hingga 15 Juni 2023, dengan fokus pada PT. Tennen Depositary Indonesia. Metode penelitian diimplementasikan menggunakan *Python*, di mana efektivitas prediksi dievaluasi berdasarkan *Mean Absolute Percentage Error* (MAPE) dan *Root Mean Squared Error* (RMSE). Hasil penelitian menunjukkan bahwa algoritma *Random Forest* mampu memberikan prediksi yang akurat dengan rata-rata MAPE sebesar 1,38% dan RMSE sebesar 4,38, yang mengindikasikan keandalan model dalam menghadapi volatilitas pasar *cryptocurrency*. Keunikan sistem ini terletak pada kemampuannya menangani kompleksitas dan volatilitas pasar, meskipun keterbatasan data historis dan fluktuasi pasar tetap menjadi tantangan. Temuan ini dapat menjadi dasar pengembangan model yang lebih adaptif dan kokoh dalam merespons dinamika pasar yang terus berkembang.

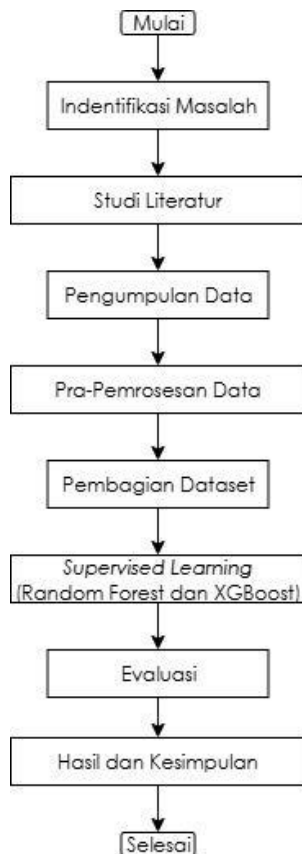
Dari berbagai studi terkait prediksi harga aset kripto, dapat dilihat bahwa *Random Forest* (RF) dan *XGBoost* memiliki keunggulan komparatif dalam hal akurasi dan stabilitas prediksi. *XGBoost* secara konsisten menunjukkan performa terbaik dalam berbagai penelitian dengan kemampuan unggul dalam menangani data berdimensi tinggi dan hubungan nonlinier, seperti yang dibuktikan oleh Tian [17], Zuo [18], dan Terence et al. [19] dengan berbagai metrik evaluasi. Sementara itu, *Random Forest* juga menunjukkan hasil yang kompetitif, terutama dalam ketahanan terhadap outlier dan volatilitas pasar, sebagaimana ditunjukkan dalam penelitian Asbullah et al. [20] dengan nilai MAPE dan RMSE yang relatif rendah. Kedua algoritma ini terbukti lebih efektif dibandingkan model lain seperti LSTM, SVM, atau *Neural Network* yang memerlukan *tuning* lebih intensif dan cenderung kurang stabil dalam beberapa kasus. Temuan-temuan ini mengindikasikan bahwa kemampuan *XGBoost* dalam menangani kompleksitas data dan ketahanan *Random Forest* terhadap noise pasar dapat menjadi landasan pengembangan model prediksi yang lebih *robust*.

Atas dasar itu, maka pada penelitian ini menerapkan kedua algoritma tersebut secara komprehensif dalam memprediksi fluktuasi harga *Ethereum* aset kripto utama yang volatilitasnya menjadi pusat perhatian investor di pasar digital. Kajian ini memiliki urgensi tinggi karena masih terbatasnya penelitian yang membandingkan secara mendalam performa algoritma prediktif untuk aset kripto spesifik, terutama dengan pendekatan *machine*

learning yang berfokus pada analisis pola pasar dan faktor-faktor penentu harga di ekosistem *blockchain*.

METODOLOGI PENELITIAN

Diagram alur penelitian pada Gambar 1 menyajikan panduan sistematis bagi peneliti dalam melaksanakan seluruh rangkaian penelitian. Bagan ini berfungsi sebagai pedoman operasional yang menjamin keterkaitan logis antar setiap tahapan penelitian sekaligus memastikan konsistensi pelaksanaan dengan desain penelitian yang telah ditetapkan.



Gambar 1. Diagram Alur Penelitian
Sumber: Hasil Olah Data

Indentifikasi Masalah

Perkembangan aset kripto dalam beberapa tahun terakhir telah menciptakan gelombang minat yang signifikan di kalangan investor, pengembang, dan lembaga keuangan. Melihat lonjakan kapitalisasi pasar yang mampu menembus ratusan miliar dolar dalam waktu relatif singkat, prediksi harga Ethereum yang akurat menjadi kebutuhan krusial untuk manajemen risiko dan perencanaan investasi [5]. Tingginya volatilitas yang menjadi ciri khas pasar *cryptocurrency* menawarkan potensi keuntungan besar namun sekaligus mengandung risiko yang tidak kecil [6]. Keberadaan model prediksi yang handal dapat menjadi alat vital dalam mendukung keputusan investasi, meminimalkan potensi kerugian, serta meningkatkan kinerja portofolio. Studi ini secara khusus mengkaji perbandingan kinerja dua algoritma

pembelajaran terawasi, *Random Forest* (RF) dan *XGBoost* dalam memprediksi harga Ethereum.

Studi Literatur

Peneliti mengkaji berbagai literatur akademis dan hasil penelitian terdahulu yang relevan dengan topik memprediksi aset kripto untuk mendapatkan landasan teoritis dan pemahaman menyeluruh tentang rencana penelitian yang akan dilakukan.

Pengumpulan Data

Tahap pengumpulan data menempati posisi sentral sebagai faktor penentu kesuksesan pelaksanaan penelitian. Metode pengumpulan data pada penelitian ini menggunakan data sekunder, dimana data diperoleh dari platform pertukaran data dan analisis kripto aset yaitu Bitfinex.

a. Pra-Pemrosesan Data

Pra-pemrosesan data merupakan tahap fundamental dalam *pipeline machine learning* yang berfungsi untuk mentransformasikan data mentah menjadi struktur terstandarisasi yang siap digunakan untuk pelatihan model, di mana dalam konteks prediksi harga aset kripto seperti Ethereum, proses ini menjadi semakin krusial mengingat karakteristik unik data finansial berbasis *time-series* yang meliputi volatilitas tinggi, adanya noise pasar, dan ketidak stasioner. Tahapan ini tidak hanya mencakup pembersihan data dasar seperti penanganan *missing values* dan *outlier*, tetapi juga melibatkan teknik-teknik khusus seperti rekayasa fitur temporal.

b. Pembagian Dataset

Proses pembagian data dilakukan setelah tahap pembersihan, dimana *dataset* dipisahkan menjadi dua kelompok utama. Kelompok pertama adalah *data training* memiliki persentase sebesar 70% dari jumlah data yang kemudian akan digunakan untuk membangun model prediktif melalui proses pembelajaran algoritma, sementara kelompok kedua adalah *data testing* sebesar 30% dari jumlah data yang akan berfungsi sebagai validasi akhir untuk mengukur kemampuan generalisasi model terhadap data baru yang belum pernah diproses sebelumnya.

c. Supervised Learning

Setelah data dibagi menjadi subset pelatihan dan pengujian, tahap berikutnya adalah membangun model prediktif menggunakan algoritma *Random Forest* (RF) dan *XGBoost*. Kedua algoritma ini dipilih karena keunggulannya dalam menangani *data time-series* finansial yang kompleks [19], [20]. *Random Forest* dengan pendekatan *ensemble* berbasis bagging-nya mampu menangkap hubungan non-linear antar variabel pasar sekaligus mengurangi risiko *overfitting* melalui mekanisme voting dari *multiple decision trees*, sementara *XGBoost* yang mengadopsi teknik *gradient boosting* secara iteratif mampu mengoptimalkan *loss function* untuk data yang sangat dinamis seperti pergerakan harga kripto. Baik *Random Forest* (RF) maupun *XGBoost* diimplementasikan

secara paralel namun terpisah dalam membangun model prediksi harga Ethereum, dimana masing-masing algoritma bekerja dengan pendekatan yang berbeda. RF menggunakan metode *ensemble* berbasis bagging dengan *multiple decision trees* yang bekerja secara independen, sementara XGBoost mengadopsi teknik *boosting* yang berfokus pada koreksi error bertahap - untuk kemudian hasil prediksi dari kedua model tersebut dibandingkan secara komprehensif melalui berbagai metrik evaluasi.

Evaluasi

Metrik evaluasi seperti RMSE (*Root Mean Square Error*), MAE (*Mean Absolute Error*), dan *R-squared* diterapkan secara komprehensif untuk menganalisis performa prediktif kedua model[21], dimana RMSE mengukur besarnya deviasi prediksi dengan memberikan bobot lebih besar pada *error* yang signifikan, MAE menilai rata-rata kesalahan absolut tanpa mempertimbangkan arah deviasi sehingga cocok untuk mengevaluasi konsistensi prediksi, sedangkan *R-squared* mengkuantifikasi seberapa baik variasi harga Ethereum dapat dijelaskan oleh model, ketiga metrik ini saling melengkapi dalam menentukan algoritma mana antara *Random Forest* dan XGBoost yang mampu memberikan akurasi lebih tinggi sekaligus stabilitas prediksi yang lebih baik dalam menangkap volatilitas harga ETH yang fluktuatif, terutama pada kondisi pasar ekstrem seperti saat bull run atau market crash.

Hasil dan Kesimpulan

Pada tahap konklusi penelitian ini, menyajikan hasil analisis prediksi harga *Ethereum* menggunakan dua algoritma *machine learning*, yaitu *Random Forest* (RF) dan XGBoost, yang dievaluasi melalui metrik RMSE, MAE, dan *R-squared* untuk mengukur tingkat akurasi prediksi. Penelitian ini menghasilkan kesimpulan komparatif mengenai efektivitas masing-masing algoritma dalam memprediksi volatilitas harga ETH, sekaligus memberikan rekomendasi untuk pengembangan model yang lebih robust di masa depan. Hasil penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap pengembangan teknologi prediksi harga aset kripto, khususnya dalam menghadapi fluktuasi pasar yang tinggi, melalui pendekatan *machine learning* yang teruji dan dapat diandalkan untuk pengambilan keputusan investasi.

HASIL DAN PEMBAHASAN

Proses Pengumpulan Data

Penelitian ini menggunakan data historis *Ethereum* (ETH) yang diperoleh dari platform pertukaran data dan analisis kripto aset Bitfinex, khususnya pasangan perdagangan ETH/USD (Bitfinex_ETHUSD_d), dengan total 3.423 observasi

harian yang mencakup rentang waktu hampir sembilan tahun, dimulai dari 3 September 2016 pukul 00:00 hingga 22 Juli 2025 pukul 00:00. Dataset tersebut mencakup berbagai variabel kunci, termasuk harga pembukaan (*Open*), harga tertinggi (*High*), harga terendah (*Low*), harga penutupan (*Close*), serta volume perdagangan yang terbagi menjadi Volume_ETH (jumlah ETH yang diperdagangkan) dan Volume_Currency (nilai perdagangan dalam USD). Fitur-fitur ini menjadi dasar analisis untuk memprediksi pergerakan harga Ethereum. Sebelum dilakukan analisis prediksi harga, data historis ETH/USD terlebih dahulu melalui tahap pra-pemrosesan data untuk memastikan kualitas data dan konsistensi variabel.

Tabel 1. Sample Data Perdagangan ETH

open	high	low	close	Volume USD	Volume ETH
4447.1	4482.5	4440.9	4481.1	232744	51.93904014
4551.7	4672.9	4375	4449.4	32310893.5	7261.854081

Sumber: Bitfinex

Pra-Pemrosesan Data

Pra-pemrosesan data merupakan tahap fundamental dalam analisis data deret waktu finansial, yang bertujuan untuk mentransformasi data mentah menjadi struktur terstandarisasi yang kompatibel dengan algoritma prediktif. Pada data historis kripto seperti ETH/USD, proses ini menjadi kritical karena karakteristik datanya yang mengandung *noise*, *missing values*, dan ketidakkonsistenan temporal. Tahapan Pra-pemrosesan data tidak hanya meningkatkan kualitas data, tetapi juga memastikan validitas relasi kausal antara variabel prediktor dengan target, sehingga secara signifikan berkontribusi pada performa model *machine learning*. Dalam penelitian ini, pra-pemrosesan data diterapkan pada 3.423 observasi harian dengan tahapan sebagai berikut:

- **Seleksi Variabel**
Proses diawali dengan seleksi variabel dimana hanya fitur inti seperti *timestamp*, OHLC (*Open, High, Low, Close*), Volume USD dan Volume ETH yang dipertahankan, sementara metadata seperti unix dan simbol dieliminasi karena tidak relevan untuk pemodelan prediktif.
- **Penanganan Missing Values**
Zero-filling untuk kolom volume dan *weighted price* (merepresentasikan absennya aktivitas perdagangan). Jika tidak ada catatan perdagangan di suatu waktu, nilainya diisi 0 karena angka nol artinya memang tidak terjadi transaksi sama sekali (bukan data yang hilang). Selain itu juga dilakukan pendekatan *forward-fill* pada data OHLC, karena harga aset seperti *Ethereum* harus selalu ada (kontinu), nilai yang hilang diisi dengan mengambil harga terakhir yang tercatat sebelumnya. Misalnya, jika harga "Close" hari ini kosong, sistem akan mengisi

dengan harga "Close" kemarin agar grafik harganya tidak terputus.

- **Rekayasa Fitur Temporal**

Dalam tahap ini dilakukan pemrosesan data deret waktu yang bertujuan menciptakan fitur-fitur prediktif baru dengan memanfaatkan informasi historis dan pola temporal untuk meningkatkan akurasi model. Penggunaan *time lag* memainkan peran penting dalam meningkatkan akurasi prediksi, khususnya dalam konteks analisis deret waktu dan pemodelan *machine learning*[22]. Teknik ini berfungsi sebagai alat fundamental untuk mengevaluasi pengaruh nilai-nilai historis (*lagged*) suatu variabel terhadap nilai terkini atau masa depan, sehingga memungkinkan model untuk menangkap dinamika temporal seperti tren, pola musiman, dan autokorelasi. Dalam konteks prediksi harga aset kripto seperti *Ethereum*, penerapan *time lag* memungkinkan model untuk mengidentifikasi pola fluktuasi harga jangka pendek dan menengah, serta respons terhadap faktor eksternal yang tercermin dalam data historis, yang pada akhirnya berkontribusi signifikan terhadap peningkatan kinerja prediktif model.

Pada penelitian ini, dilakukan penambahan 7 fitur lag untuk setiap variabel OHLC (Open, High, Low, Close) dan volume perdagangan yang merepresentasikan pola historis selama seminggu terakhir, dimana setiap fitur lag mencatat nilai variabel pada hari ke- $t-1$ hingga $t-7$, serta pembentukan variabel target berupa *Prediction_Target* yang merepresentasikan harga penutupan hari berikutnya ($t+1$). Proses pembuatan fitur *lag* ini secara alami menghasilkan *missing values* pada 7 baris awal dataset akibat ketiadaan data historis sebelum periode tersebut, sementara pada baris terakhir muncul *missing value* pada kolom *Prediction_Target* karena ketiadaan data masa depan ($t+1$). Untuk menjaga konsistensi data[16] maka dilakukan penghapusan baris yang mengandung *missing values* sehingga menyebabkan total observasi akhir berkurang dari 3.423 menjadi 3.415 baris data yang siap digunakan untuk pemodelan prediktif.

Tabel 2. Sample Data Hasil Pengolahan

...	Open_b_7	High_b_7	Low_b_7	Close_b_7	Volum_e_ETH_b_7	Volum_e_USD_b_7	Prediction_Target
...	2939	3013.6	2933.5	2968.1	15832.97664	46993857.97	3762.9
...	2968.4	3078.8	2961.2	3014.4	14055.34567	42368434	3754.6

Sumber: Hasil Olah Data

a. Pembagian Dataset

Dari total 3.415 data yang telah melalui tahap pra-pemrosesan, dilakukan pembagian *dataset* menjadi dua subset utama dengan proporsi 80:20, yaitu sebanyak 2.732 data (80%) dialokasikan sebagai *data training* dan 683 data (20%) sebagai data

testing. Skema pembagian ini dipilih untuk memastikan model dilatih pada data historis yang cukup dan diuji pada data yang benar-benar belum pernah dilihat, sehingga dapat dievaluasi kemampuannya dalam memprediksi pergerakan harga *Ethereum* secara generalisasi.

b. Supervised Learning

Implementasi *supervised learning* dalam penelitian ini dilakukan menggunakan bahasa pemrograman *Python* yang dijalankan pada platform *Google Colab* untuk menerapkan algoritma *Random Forest* dan *XGBoost*. *Python* dipilih karena kemampuannya dalam menyediakan *library machine learning* yang komprehensif seperti *Scikit-learn* dan *XGBoost*, yang memungkinkan implementasi efisien dari kedua algoritma tersebut dalam membangun model prediktif berbasis data historis. Melalui platform *Google Colab*, proses komputasi yang intensif dapat dijalankan secara optimal dengan memanfaatkan infrastruktur *cloud* tanpa bergantung pada sumber daya lokal. Implementasi kode untuk pelatihan model meliputi tahap konfigurasi *hyperparameter* dan pelatihan model dengan *data training* seperti yang ditunjukkan pada Gambar 2.

```
from sklearn.ensemble import RandomForestRegressor
import xgboost as xgb
models = {
    "Random Forest": RandomForestRegressor(n_estimators=1000, random_state=5, max_depth=3),
    "XGBoost": xgb.XGBRegressor(
        n_estimators=1000,
        random_state=5,
        max_depth=3
    )
}
for name, model in models.items():
    model.fit(X_train, np.ravel(y_train))
    predictions = model.predict(X_test)
```

Gambar 2. Kode RF dan XGBoost

Sumber : Hasil Olah Kode (Phyton)

Kode di atas mengimplementasikan dua model *machine learning ensemble* untuk tugas regresi, yaitu *Random Forest* dan *XGBoost*, yang dikonfigurasi dengan parameter yang identik. Masing-masing menggunakan 1000 pohon keputusan ($n_estimators=1000$), kedalaman maksimum 3 ($max_depth=3$), dan random state 5 ($random_state=5$) untuk memastikan hasil yang dapat direproduksi dan mencegah *overfitting* dengan membatasi kompleksitas setiap pohon individu. Selanjutnya, setiap model dalam *dictionary models* di-loop untuk dilakukan proses pelatihan pada data latih (X_train dan y_train yang diubah menjadi array 1D) serta menghasilkan prediksi pada data uji (X_test), sehingga memungkinkan perbandingan performa yang fair antara kedua algoritma pada *dataset* yang sama.

c. Evaluasi

Kualitas prediksi dari model yang dibangun dinilai menggunakan dua metrik evaluasi, yaitu *Root Mean Squared Error* (RMSE) dan Koefisien Determinasi (R^2). RMSE digunakan untuk mengukur tingkat akurasi prediksi model dengan menghitung rata-rata akar kuadrat dari selisih antara nilai aktual dan nilai hasil prediksi.

Nilai RMSE yang semakin kecil menunjukkan bahwa hasil prediksi model semakin mendekati nilai sebenarnya. Perumusan RMSE dapat dituliskan sebagai berikut:

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

Sementara itu, koefisien determinasi (R^2) digunakan untuk mengukur seberapa besar proporsi variasi pada variabel target yang dapat dijelaskan oleh model. Nilai R^2 dihitung berdasarkan perbandingan antara jumlah kuadrat selisih antara nilai aktual dan nilai prediksi dengan jumlah kuadrat selisih antara nilai aktual dan rata-rata nilai aktual.

Nilai R^2 berkisar antara 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa model mampu menjelaskan sebagian besar variasi data.

Rumus R^2 dinyatakan sebagai berikut :

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (2)$$

Dengan demikian, nilai RMSE yang rendah dan nilai R^2 yang tinggi menunjukkan bahwa model prediksi memiliki kinerja yang lebih baik dalam menjelaskan variasi hubungan antara variabel input dan output secara akurat.

Hasil eksperimen yang dipaparkan pada Tabel 3, menunjukkan kemampuan dari setiap model dalam memprediksi harga Ethereum. Namun, terdapat perbedaan kinerja yang signifikan antara keduanya.

Tabel 3. Hasil Kinerja dari Setiap Model Machine Learning

Model	RMSE	R2
Random Forest	208.4530652	0.8995679563
XGboost	134.6300608	0.9581071591

Sumber: Hasil Output Program

Berdasarkan hasil evaluasi, XGBoost menunjukkan kinerja prediktif yang lebih unggul dibandingkan *Random Forest* dalam memprediksi harga Ethereum. Koefisien Determinasi (R^2) XGBoost yaitu bernilai 0.958 mengindikasikan kemampuannya menjelaskan 95.8% varians data secara signifikan lebih tinggi daripada *Random Forest* yang bernilai 0.900 atau 90.0%. Hal ini menunjukkan tingkat kesesuaian (*goodness-of-fit*) model XGBoost yang lebih baik terhadap data, sehingga mampu menangkap hubungan yang lebih kompleks antara variabel prediktor dan harga Ethereum. Di sisi lain, nilai *Root Mean Square Error* (RMSE) XGBoost lebih rendah (134.63) dibandingkan *Random Forest* (208.45) mengonfirmasi keakuratan yang lebih tinggi. Selisih 73.82 ini berarti prediksi XGBoost secara rata-rata lebih mendekati harga aktual, dengan tingkat kesalahan *Random Forest* 55% lebih besar. Secara keseluruhan, kinerja XGBoost yang lebih unggul untuk menganalisis data keuangan yang fluktuatif dan rumit.

KESIMPULAN DAN SARAN

Kedua model, *Random Forest* dan XGBoost, telah membuktikan diri sebagai alat prediktif yang valid untuk harga Ethereum, dengan kinerja yang melampaui model *baseline*. Namun, berdasarkan evaluasi metrik RMSE dan R^2 , XGBoost secara signifikan lebih unggul daripada *Random Forest*. Keunggulan ini diduga kuat berasal dari mekanisme *boosting* yang adaptif dan penerapan regularisasi yang ketat pada XGBoost, yang membuatnya lebih cocok untuk menangkap kompleksitas dan *noise* yang melekat pada data pasar kripto. Oleh karena itu, untuk aplikasi prediksi harga Ethereum yang memprioritaskan akurasi tinggi, XGBoost direkomendasikan sebagai model pilihan. Penelitian lanjutan dapat mengeksplorasi tuning hyperparameter yang lebih mendalam atau mengintegrasikan model ini dengan pendekatan *deep learning* lainnya untuk akurasi yang lebih baik lagi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam pelaksanaan penelitian ini, khususnya kepada institusi dan pembimbing yang telah memberikan arahan dan fasilitas.

DAFTAR PUSTAKA

- [1] M. Aquilina, G. Cornelli, J. Frost, and L. Gambacorta, "Cryptocurrencies and decentralised finance: functions and financial stability implications".
- [2] M. S. Minu, S. R. Suryaa, J. J. Titus, and S. Sharan, "Empowering Small-Scale Farmers with Decentralized Finance (DeFi) via Ethereum Smart Contracts," in *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, Coimbatore, India: IEEE, July 2024, pp. 683–690. doi: 10.1109/ICSCSS60660.2024.10625462.
- [3] Syarif Hidayatullah State Islamic University Jakarta and Y. Durachman, "Blockchain and the Evolution of Decentralized Finance Navigating Growth and Vulnerabilities," *J. Curr. Res. Blockchain*, vol. 1, no. 3, pp. 166–177, Dec. 2024, doi: 10.47738/jcrb.v1i3.20.
- [4] M. Gupta and A. Singla, "Identification of Expected Growth in Crypto Currency," *Sci. J. Metaverse Blockchain Technol.*, vol. 2, no. 2, pp. 56–67, Aug. 2024, doi: 10.36676/sjmbt.v2.i2.33.
- [5] K. Murray, A. Rossi, D. Carraro, and A. Visentin, "On Forecasting Cryptocurrency Prices: A Comparison of Machine Learning, Deep Learning, and Ensembles," *Forecasting*, vol. 5, no. 1, pp. 196–209, Jan. 2023, doi: 10.3390/forecast5010010.
- [6] S. E. Hardiyanti, "Risk and Return Analysis on Cryptocurrency Investment," *juket*, vol. 4, no. 2, pp. 55–58, Dec. 2024, doi: 10.54076/juket.v4i2.515.

- [7] E. Sener and I. Demir, "Wavelet-Denoised ARIMA and LSTM Method to Predict Cryptocurrency Price," in *2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, Ankara, Turkey: IEEE, Oct. 2022, pp. 577–582. doi: 10.1109/ISMSIT56059.2022.9932727.
- [8] Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab-140401, India, M. Muskaan, P. K. Sarangi, and Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab-140401, India, "A Literature Review on Machine Learning Applications in Financial Forecasting," *J. Technol. Manag. Grow. Econ.*, vol. 11, no. 1, pp. 23–27, July 2020, doi: 10.15415/jtmge.2020.111004.
- [9] S. Elhishi, Z. Rabea, and S. El-Metwally, "Exploring the Predictive Power of Machine Learning Algorithms on Daily Gold Prices:," in *Advances in Finance, Accounting, and Economics*, M. Irfan, M. Elmogy, M. Shabri Abd. Majid, and S. El-Sappagh, Eds., IGI Global, 2024, pp. 131–150. doi: 10.4018/979-8-3693-0082-4.ch008.
- [10] S. Vaka, M. S. Reddy, and S. N. Prabhu, "Hybrid Model for Cryptocurrency Price Prediction using LSTM, Bidirectional LSTM, and XGBoost," in *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, Bengaluru, India: IEEE, Dec. 2024, pp. 925–932. doi: 10.1109/ICICNIS64247.2024.10823223.
- [11] J. Wang, "Research on Stock Price Rise and Fall Prediction Based on Optimization Random Forest," *Appl. Comput. Eng.*, vol. 133, no. 1, pp. 61–67, Jan. 2025, doi: 10.54254/2755-2721/2025.20601.
- [12] T. Zhu, "Analysis on the Applicability of the Random Forest," *J. Phys. Conf. Ser.*, vol. 1607, no. 1, p. 012123, Aug. 2020, doi: 10.1088/1742-6596/1607/1/012123.
- [13] M. Robnik-Šikonja, "Improving Random Forests," in *Machine Learning: ECML 2004*, vol. 3201, J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi, Eds., in *Lecture Notes in Computer Science*, vol. 3201, Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 359–370. doi: 10.1007/978-3-540-30115-8_34.
- [14] D. Tarwidi, S. R. Pudjaprasetya, D. Adytia, and M. Apri, "An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach," *MethodsX*, vol. 10, p. 102119, 2023, doi: 10.1016/j.mex.2023.102119.
- [15] O. M. Wulandari, I. Maulana, F. Syamsudin, and R. Waluyo, "Perbandingan Algoritma Naive Bayes dan SVM dalam Analisis Sentimen Twitter terhadap Isu Ijazah Jokowi Palsu," *J. Manaj. Inform. Sist. Inf. Dan Teknol. Komput. JUMISTIK*, vol. 4, no. 1, pp. 392–400, June 2025, doi: 10.70247/jumistik.v4i1.145.
- [16] N. Rahayu, S. Indri Yani, M. Marwah, and A. Pratama, "Analisis Sentimen terhadap Terorisme pada Platform Twitter menggunakan Support Vector Machine," *J. Manaj. Inform. Sist. Inf. Dan Teknol. Komput. JUMISTIK*, vol. 4, no. 1, pp. 430–441, June 2025, doi: 10.70247/jumistik.v4i1.152.
- [17] S. Tian, "Price Prediction of the Cryptocurrency from Niche Market Based on Random Forest, LSTM and XGBoost," *Adv. Econ. Manag. Polit. Sci.*, vol. 87, no. 1, pp. 97–106, June 2024, doi: 10.54254/2754-1169/87/20240911.
- [18] A. Zuo, "Comparison and Analysis of the Accuracy of Various Machine Learning Algorithms in Bitcoin Price Prediction," *Adv. Econ. Manag. Polit. Sci.*, vol. 70, no. 1, pp. 302–308, Jan. 2024, doi: 10.54254/2754-1169/70/20231726.
- [19] S. Terence, J. Immaculate, Titus, D. B. Naik, and S. Selvarathi, "Machine Learning based Bitcoin Price Prediction," in *2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, Tirunelveli, India: IEEE, Nov. 2024, pp. 1070–1074. doi: 10.1109/ICDICI62993.2024.10810810.
- [20] J. Asbullah and S. Samsudin, "Prediksi Harga Cryptocurrency Binance Berdasarkan Informasi Blockchain dengan Menggunakan Algoritma Random Forest," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 1, p. 260, Jan. 2024, doi: 10.30865/mib.v8i1.7100.
- [21] MBA in Business Analytics, Gannon University, Erie, PA et al., "EVALUATING THE EFFECTIVENESS OF MACHINE LEARNING ALGORITHMS IN PREDICTING CRYPTOCURRENCY PRICES UNDER MARKET VOLATILITY: A STUDY BASED ON THE USA FINANCIAL MARKET," *Am. J. Manag. Econ. Innov.*, vol. 06, no. 12, pp. 15–38, Dec. 2024, doi: 10.37547/tajmei/Volume06Issue12-03.
- [22] A. T. Firat, O. Aygün, M. Gögebakan, C. Ulus, and M. F. Akay, "Developing Demand Forecasting Models for E-Commerce: Analyzing the Impact of Time Lags on Model Performance," *Sci. J. Mehmet Akif Ersoy Univ.*, vol. 8, no. 1, pp. 1–15, June 2025, doi: 10.70030/sjmakeu.1592024.