

Perbandingan Linear Regression dan Random Forest untuk Prediksi Harga Cryptocurrency

Performance Evaluation of Machine Learning Models for Cryptocurrency Price Prediction

Hadid Abdilla¹, Heni Sulistiani²

Program Studi Informatika, Fakultas Teknik Ilmu Komputer, Universitas Teknokrat Indonesia,
Bandar Lampung, Indonesia

Email: hadidabdilla20@gmail.com¹, henisulistiani@teknokrat.ac.id²

Received 22 June 2026; Revised 1 August 2026; Accepted 14 August 2026

Abstrak - Perkembangan *Cryptocurrency* yang pesat menjadikan aset digital seperti BNB dan ETH semakin diminati oleh investor, namun volatilitas harga yang tinggi juga meningkatkan risiko kerugian. Oleh karena itu, diperlukan metode prediksi yang dapat membantu memahami arah pergerakan harga secara lebih terukur. Penelitian ini bertujuan untuk memperkirakan pergerakan harga aset *Cryptocurrency* secara berbasis data. Penelitian ini juga bertujuan untuk mengukur kemampuan model dalam menangkap pola harga pada aset yang bersifat volatil menggunakan pendekatan *Machine Learning*, yaitu *Random Forest Regression* dan *Linear Regression*. Data yang digunakan berupa data historis periode 2020-2025. Proses pengolahan data meliputi preprocessing data, penerapan model *Machine Learning* dan evaluasi kinerja model menggunakan MSE, R^2 , dan MAPE. Hasil penelitian menunjukkan bahwa model *Random Forest Regression* dan *Linear Regression* memiliki performa yang berbeda dalam memprediksi harga *Cryptocurrency* BNB dan ETH. *Linear Regression* menghasilkan MAPE 5,14% pada BNB dan 8,17% pada ETH. Sementara itu, *Random Forest Regression* menunjukkan performa yang lebih rendah dengan MAPE 15,02% pada BNB dan 9,65% pada ETH. Nilai tersebut menunjukkan tingkat kesalahan prediksi yang bervariasi. Dengan demikian, *Linear Regression* dinilai lebih mampu menangkap pola data yang kompleks dan fluktuatif, sehingga lebih direkomendasikan sebagai model prediksi harga *Cryptocurrency* dalam penelitian ini.

Kata kunci : *Random Forest Regression; Linear Regression; Cryptocurrency; BNB; ETH; Machine Learning*

Abstract - The rapid development of *Cryptocurrency* has made digital assets such as BNB and ETH increasingly attractive to investors; however, High price volatility also increases the risk of financial loss. Therefore, a reliable prediction method is required to better understand price movements in a more measurable way. This study aims to predict *Cryptocurrency* price movements based on historical data and to evaluate the ability of *Machine Learning* models to capture patterns in Highly volatile assets using *Random Forest Regression* and *Linear Regression* approaches. The dataset used consists of historical price data from 2020 to 2025. The data processing stage includes preprocessing, model implementation, and performance evaluation using Mean Squared Error (MSE), R^2 , and relative error. The results show that *Random Forest Regression* and *Linear Regression* exhibit different performances in predicting the prices of BNB and ETH. *Linear Regression* produces a relative error of 5.14% for BNB and 8.17% for ETH, while *Random Forest Regression* shows Lower performance with relative errors of 15.02% for BNB and 9.65% for ETH. These results indicate varying levels of prediction accuracy across models and datasets. Overall, *Linear Regression* is considered more capable of capturing complex and fluctuating data patterns and is therefore more recommended as a *Cryptocurrency* price prediction model in this study.

Keywords : *Random Forest Regression; Linear Regression; Cryptocurrency; BNB; ETH; Machine Learning*

1. PENDAHULUAN

Kemajuan teknologi dan informasi yang terjadi pada era modern telah membawa perubahan yang signifikan dalam berbagai aspek kehidupan masyarakat. Perubahan tersebut memengaruhi pola perilaku masyarakat yang kini cenderung mengutamakan kemudahan, kecepatan, sifat yang dinamis, dan efisiensi dalam menjalankan berbagai aktivitas [1]. Salah satu dampak dari perkembangan tersebut dapat dilihat pada sektor keuangan, yang turut mengalami transformasi melalui hadirnya berbagai inovasi berbasis teknologi digital. Dalam beberapa tahun terakhir, pasar mata uang *Cryptocurrency* berkembang menjadi salah satu instrumen keuangan yang sangat dinamis dan menarik bagi investor maupun peneliti [2]. Meningkatnya minat terhadap aset digital ini tidak terlepas dari semakin luasnya adopsi teknologi dalam kehidupan sehari-hari masyarakat. Perhatian terhadap perkembangan mata uang kripto ini selaras dengan perkembangan lingkungan masyarakat yang semakin kental dengan penggunaan teknologi dan alat-alatnya [3]. Di sisi lain, pertumbuhan pasar *Cryptocurrency* juga diiringi oleh berbagai tantangan dalam proses analisis dan pengambilan keputusan investasi. Karakteristiknya yang mencakup volatilitas tinggi, likuiditas 24/7, serta pengaruh berbagai faktor fundamental dan teknikal menjadikan prediksi pergerakan harga serta penentuan sinyal jual-beli sebagai tantangan yang kompleks [4].

Sebagai respons terhadap tantangan tersebut, pendekatan *Machine Learning* (ML) telah banyak dimanfaatkan untuk mengatasi kompleksitas tersebut karena kemampuannya dalam memodelkan pola *non-linear* dan ketergantungan temporal pada data keuangan digital [5]. *Machine learning* dapat didefinisikan sebagai aplikasi komputer dan algoritma matematika yang diadopsi dengan cara pembelajaran yang berasal dari data dan menghasilkan prediksi di masa yang akan datang [6]. Dalam konteks investasi, penggunaan *Machine Learning* berperan penting dalam membantu investor memperoleh keuntungan optimal sekaligus meminimalkan risiko melalui analisis data historis dan prediksi [7]. Pembelajaran dalam *Machine Learning* ini dilakukan melalui serangkaian proses yang bertujuan membangun kecerdasan model, yang secara umum terdiri atas tahap pelatihan (training) dan tahap pengujian (testing) [8]. Berbagai algoritma, seperti *Random Forest Regression*, *Multiple Regression*, dan *Long Short-Term Memory* (LSTM), telah menunjukkan potensi yang signifikan dalam memprediksi harga kripto dengan tingkat akurasi yang tinggi [9].

Sejumlah penelitian terdahulu juga memperkuat temuan tersebut. Penggunaan LSTM terbukti mampu menangkap tren dan fluktuasi harga dengan baik, ditunjukkan oleh nilai RMSE sebesar 0,042 dan MAPE sebesar 18,7% tanpa indikasi overfitting [10]. Selain itu, *Random Forest Regression* menghasilkan nilai MAPE sebesar 1,50% dengan tingkat akurasi mencapai 98,50%, menunjukkan kinerja yang sangat unggul [11]. Sementara itu, metode *Multiple Regression* juga menunjukkan performa yang baik dengan skor model sebesar 0,99 dan RMSE sebesar 3,73 pada data pengujian [12]. Temuan-temuan ini mengindikasikan bahwa keberhasilan model prediksi tidak hanya ditentukan oleh pemilihan algoritma, tetapi juga oleh metode evaluasi yang digunakan dalam mengukur kemampuan model merepresentasikan data.

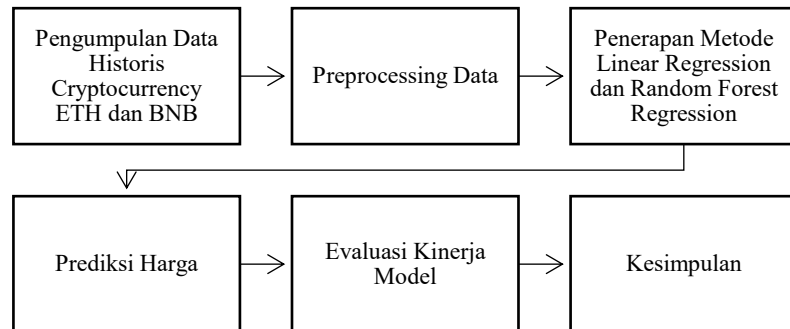
Meskipun demikian, sebagian besar penelitian masih berfokus pada penggunaan satu algoritma secara individual tanpa melakukan perbandingan mendalam antara model linear dan non-linear dalam konteks data *Cryptocurrency* yang fluktuatif [13]. Selain itu, kajian mengenai penggunaan variabel teknikal dasar seperti *Open*, *High*, *Low*, dan *Volume* masih terbatas, padahal variabel-variabel tersebut memiliki potensi korelasi yang tinggi, khususnya antara *High*, *Low*, dan *Close*, yang dapat menimbulkan permasalahan multikolinearitas pada model tertentu [14]. Kondisi ini menunjukkan pentingnya evaluasi terhadap kemampuan model dalam menangani hubungan antar variabel tersebut [15].

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk melakukan analisis komparatif antara *Linear Regression* dan *Random Forest Regression* dalam memprediksi harga *Cryptocurrency*, khususnya pada aset *Ethereum* (ETH) dan *Binance Coin* (BNB), menggunakan data historis periode 2020–2025. Evaluasi dilakukan menggunakan metrik *Mean Squared Error*

(MSE) dan *Koefisien Determinasi* (R^2) untuk mengukur tingkat kesalahan prediksi serta kemampuan model dalam merepresentasikan variasi data. Dengan demikian, penelitian ini tidak hanya berfokus pada tingkat akurasi, tetapi juga pada kemampuan model dalam menangani hubungan antar variabel input, sehingga diharapkan dapat memberikan pemahaman yang lebih komprehensif dalam pemilihan metode prediksi yang sesuai.

2. METODE

Kerangka penelitian yang digunakan dalam penelitian ini ditunjukkan pada Gambar 1 berikut.



Gambar 1 Alur penelitian

Berdasarkan Gambar 1, alur penelitian yang digunakan dalam proses prediksi harga *Cryptocurrency* ETH dan BNB. Penelitian diawali dengan tahap pengumpulan data historis, kemudian dilanjutkan dengan preprocessing data untuk membersihkan dan menyiapkan data sebelum digunakan. Setelah itu, dilakukan penerapan metode *Linear Regression* dan *Random Forest Regression* untuk proses prediksi harga. Hasil prediksi kemudian dievaluasi pada tahap evaluasi kinerja model untuk mengetahui tingkat akurasi model yang digunakan. Tahap terakhir adalah penarikan kesimpulan berdasarkan hasil penelitian yang telah diperoleh.

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa data historis harga *Cryptocurrency* Ethereum (ETH) dan Binance (BNB) yang mencakup variabel *Open*, *High*, *Low*, *Close*, dan *Volume*. Data diperoleh dari *CoinMarketCap*, yang merupakan platform penyedia data *Cryptocurrency* publik yang banyak digunakan sebagai rujukan dalam penelitian akademik. Data historis harga ETH dan BNB dikumpulkan dalam periode waktu tertentu yaitu dari tahun 2020 hingga 2025 dan digunakan sebagai input dalam proses pelatihan serta pengujian model *Machine Learning*. Pemilihan kedua koin tersebut dilakukan karena ETH dan BNB merupakan *Cryptocurrency* dengan tingkat likuiditas tinggi dan *Volume* perdagangan yang besar, sehingga mampu merepresentasikan pergerakan pasar *Cryptocurrency* secara umum dalam mendukung evaluasi kinerja model prediksi harga *Cryptocurrency*.

2.2. Preprocessing Data

Ada beberapa tahap pada pra-pemrosesan data ini, yang pertama mengubah format tanggal menjadi tipe *datetime*, mengurutkan data berdasarkan tanggal, serta menghapus data kosong agar data siap digunakan dalam proses analisis dan prediksi.

Pada tahap kedua dilakukan pembersihan data dengan menghapus nilai yang hilang (*missing values*) guna menjaga kualitas dan kelengkapan data yang digunakan dalam penelitian. Penanganan *Outlier* dilakukan menggunakan metode *Interquartile Range* (IQR), dengan mengidentifikasi dan menghapus data yang berada di luar rentang kuartil pertama (Q1) dan kuartil ketiga (Q3). Langkah ini bertujuan untuk mengurangi pengaruh nilai ekstrem yang dapat

memengaruhi kinerja model, mengingat karakteristik data *Cryptocurrency* yang memiliki volatilitas tinggi.

Tahap ketiga menyesuaikan dengan pendekatan deret waktu (time series), dilakukan pembuatan fitur lag, di mana variabel pada periode sebelumnya (t-1) digunakan untuk memprediksi nilai pada periode berikutnya (t). Variabel independen yang digunakan meliputi *Open*, *High*, *Low*, dan *Volume* dalam bentuk lag, sedangkan variabel dependen adalah harga penutupan (*Close*).

Pada tahap keempat ini data dibagi menjadi data latih dan data uji dengan perbandingan 80:20 secara kronologis untuk menghindari kebocoran informasi (data *leakage*) dan memastikan evaluasi model dilakukan pada data yang belum pernah dilihat sebelumnya.

2.3 Linear Regression

Linear Regression merupakan salah satu metode analisis statistik yang digunakan untuk memodelkan hubungan antara variabel dependen (Y) dengan satu atau lebih variabel independen (X). Tujuan utama dari metode ini adalah membangun sebuah model *Linear* yang dapat digunakan untuk melakukan prediksi serta memahami seberapa besar pengaruh variabel independen terhadap variabel dependen. *Linear Regression* bekerja dengan asumsi bahwa hubungan antara variabel X dan Y bersifat *Linear*. Artinya, perubahan pada nilai X diharapkan menyebabkan perubahan yang searah dan proporsional pada Y. Model ini sederhana, mudah diinterpretasikan, dan sering digunakan sebagai metode dasar dalam analisis data maupun *Machine Learning*.

Algoritma *Linear Regression* menggunakan model persamaan berikut.

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon \quad (1)$$

Berdasarkan persamaan (1), Y merupakan variabel dependen berupa harga penutupan (*Close*), sedangkan X_1 – X_n merupakan variabel independen yang terdiri dari harga pembukaan (*Open*), harga tertinggi (*High*), harga terendah (*Low*), dan volume perdagangan (*Volume*). Simbol *a* merupakan konstanta (*intercept*) yang menunjukkan nilai prediksi harga penutupan, b_1 – b_n merupakan koefisien regresi yang menunjukkan besar dan arah pengaruh masing-masing variabel bebas terhadap harga penutupan (*Close*), sedangkan ε merupakan nilai error atau residual, yaitu selisih antara nilai aktual dan hasil prediksi model.

2.4 Random Forest Regression

Random Forest Regression adalah salah satu metode dalam *Machine Learning* yang digunakan untuk memprediksi angka atau nilai tertentu misalnya, harga mata uang *Cryptocurrency* seperti *ethereum* di masa depan. Model ini bekerja dengan cara membuat banyak pohon keputusan (*decision tree*) dan kemudian menggabungkan hasilnya. Jadi bukan hanya satu pohon yang digunakan, tapi ratusan bahkan ribuan pohon kecil yang masing-masing membuat perkiraan. Hasil akhirnya adalah rata-rata dari semua pohon tersebut, sehingga prediksi jadi lebih akurat dan tidak mudah salah karena data yang berubah-ubah.

Algoritma *Random Forest Regression* menggunakan model Persamaan berikut.

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N f_i(x) \quad (2)$$

Pada persamaan *Random Forest Regression* ini, $\hat{y}$ merupakan nilai prediksi akhir yang dihasilkan model, *N* adalah jumlah pohon keputusan (*decision tree*) yang digunakan dalam proses prediksi, Σ merupakan simbol penjumlahan seluruh hasil prediksi dari masing-masing pohon, $f_i(x)$ adalah hasil prediksi yang diberikan oleh pohon keputusan ke-*i* terhadap data masukan *x*, sedangkan *i* menunjukkan indeks atau urutan pohon keputusan mulai dari 1 hingga *N*. Dengan

cara ini, hasil prediksi yang diperoleh menjadi lebih stabil dan lebih akurat karena mempertimbangkan pendapat dari banyak model.

Model *Random Forest Regression* pada penelitian ini diimplementasikan menggunakan kelas *RandomForestRegressor* dari pustaka *scikit-learn* dengan konfigurasi hyperparameter $n_estimators = 100$ (jumlah pohon keputusan yang dibangun dalam forest) dan $random_state = 42$ untuk menjamin reproducibility hasil, yang diterapkan secara konsisten pada kedua aset (BNB dan ETH). Hyperparameter lainnya menggunakan nilai bawaan (default) dari pustaka *scikit-learn*, yaitu $max_depth = None$ (setiap pohon dikembangkan hingga seluruh daun bersifat murni atau hingga jumlah sampel pada suatu simpul kurang dari $min_samples_split$), $min_samples_split = 2$, $min_samples_leaf = 1$, $max_features = 1.0$ (seluruh fitur dipertimbangkan pada setiap pemisahan/split), dan $bootstrap = True$ (setiap pohon dilatih menggunakan sampel data hasil *bootstrap*). Pencantuman konfigurasi hyperparameter ini dimaksudkan untuk menjamin bahwa eksperimen yang dilakukan dapat direproduksi (reproducible) oleh peneliti lain.

2.5 Prediksi Harga

Pada tahap prediksi harga, kedua model yang digunakan, yaitu *Linear Regression* dan *Random Forest Regression*, diterapkan untuk memprediksi harga penutupan (*Close*) *Cryptocurrency* BNB dan ETH. Proses prediksi dilakukan dengan menggunakan variabel *Open*, *High*, *Low*, dan *Volume* pada periode sebelumnya ($t-1$) sebagai input untuk memprediksi nilai *Close* pada periode saat ini (t), sehingga model hanya memanfaatkan informasi historis dan terhindar dari data leakage. Selanjutnya, hasil prediksi dari masing-masing model dibandingkan dengan data aktual pada data uji yang bersifat kronologis untuk mengukur tingkat kesalahan prediksi. Evaluasi dilakukan menggunakan metrik MSE, R^2 -Score dan MAPE. Perbandingan hasil dari kedua metode ini digunakan untuk menentukan model yang lebih baik dalam menangkap pola pergerakan harga *Cryptocurrency* secara lebih realistis.

2.6 Mean Squared Error (MSE)

Mean Squared Error (MSE) merupakan salah satu metrik evaluasi yang paling umum digunakan dalam model regresi, termasuk pada algoritma *Random Forest Regression*. MSE digunakan untuk mengukur rata-rata selisih kuadrat antara nilai aktual dengan nilai prediksi yang dihasilkan oleh model. Dengan mengkuadratkan selisihnya, MSE memastikan bahwa semua kesalahan bernilai positif dan memberikan penalti lebih besar terhadap kesalahan yang besar. Secara matematis, MSE dapat dituliskan dengan rumus:

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

Pada persamaan MSE, simbol *MSE* menyatakan nilai rata-rata kuadrat kesalahan prediksi, sedangkan n menunjukkan jumlah data yang digunakan dalam proses pengujian. Simbol Σ merupakan operasi penjumlahan terhadap seluruh nilai error yang telah dikuadratkan. Nilai y_i merepresentasikan nilai aktual pada data ke- i , sementara $\hat{y}_i$ merupakan nilai prediksi yang dihasilkan model pada data ke- i . Selisih antara nilai aktual dan nilai prediksi dinyatakan sebagai $(y_i - \hat{y}_i)$, yang disebut sebagai error, kemudian dikuadratkan menjadi $(y_i - \hat{y}_i)^2$ untuk menghilangkan nilai negatif dan memberikan penalti yang lebih besar terhadap kesalahan prediksi yang besar. Nilai MSE diperoleh dengan menjumlahkan seluruh kuadrat error dan membaginya dengan jumlah data pengujian. Semakin kecil nilai MSE yang dihasilkan, maka semakin baik kemampuan model dalam menghasilkan prediksi yang mendekati nilai aktual.

2.7 R-squared (R^2 Score)

Metrik yang sering digunakan untuk mengevaluasi performa model regresi adalah *R-squared* (R^2). Nilai R^2 menunjukkan seberapa besar proporsi variasi dari data aktual yang dapat

dijelaskan oleh model prediksi. R^2 digunakan untuk mengukur sejauh mana model dapat meniru pola yang ada pada data asli.

Rumus R^2 dapat ditulis sebagai berikut:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

Nilai R^2 diperoleh dengan membandingkan jumlah kuadrat kesalahan prediksi terhadap jumlah kuadrat total variasi data. Dalam persamaan tersebut, y_i merupakan nilai aktual, $\hat{y}_i$ merupakan nilai prediksi model, dan $\bar{y}$ merupakan rata-rata nilai aktual. Nilai R^2 yang semakin mendekati 1 menunjukkan bahwa model memiliki kemampuan yang semakin baik dalam menjelaskan variasi data, sedangkan nilai yang mendekati 0 menunjukkan bahwa kemampuan model dalam menjelaskan data masih rendah.

2.8 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) digunakan sebagai salah satu metrik evaluasi untuk mengukur tingkat kesalahan prediksi model dalam bentuk persentase. MAPE dihitung dengan mencari rata-rata nilai absolut dari selisih antara nilai aktual dan nilai prediksi yang kemudian dibagi dengan nilai aktual. Penggunaan nilai absolut bertujuan agar kesalahan prediksi yang bernilai positif maupun negatif tidak saling menghilangkan, sehingga nilai yang diperoleh merepresentasikan besarnya kesalahan prediksi secara keseluruhan. Semakin kecil nilai MAPE yang dihasilkan, maka semakin baik kemampuan model dalam melakukan prediksi terhadap data yang diuji. Persamaan MAPE dirumuskan sebagai berikut:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (5)$$

Pada persamaan tersebut, (MAPE) menyatakan nilai *Mean Absolute Percentage Error*, (n) merupakan jumlah data pada data pengujian (testing data), sedangkan $\sum_{i=1}^n$ menunjukkan proses penjumlahan nilai kesalahan dari data ke-1 hingga data ke-(n). Simbol y_i menyatakan nilai aktual variabel target, yaitu harga penutupan (Close) cryptocurrency pada data ke-(i), sementara $\hat{y}_i$ menyatakan nilai prediksi yang dihasilkan oleh model Random Forest Regression maupun Linear Regression pada data ke-(i). Selisih antara nilai aktual dan nilai prediksi dinyatakan sebagai $y_i - \hat{y}_i$, kemudian operator nilai absolut $\left| \frac{y_i - \hat{y}_i}{y_i} \right|$ digunakan untuk menghilangkan tanda negatif sehingga yang dihitung hanya besarnya kesalahan prediksi. Hasil perbandingan tersebut dibagi dengan nilai aktual y_i , kemudian dirata-ratakan terhadap seluruh data pengujian dan dikalikan dengan 100% sehingga diperoleh rata-rata persentase kesalahan prediksi model.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini tools yang digunakan yaitu Google Collaboratory dengan bahasa pemrograman *Python*. Proses pengolahan data pada penelitian ini menggunakan bantuan pustaka *Numpy*, *Pandas* dan *Matplotlib*. Data yang digunakan mencakup rentang waktu dari tahun 2020 sampai 2025, dengan jumlah total 311 data mingguan seperti terlihat pada Tabel 1 dan Tabel 2 berikut.

Tabel 1 Data historis BNB

<i>Date</i>	<i>Open</i>	<i>High</i>	<i>Low</i>	<i>Close</i>	<i>Volume</i>
06/01/2020	14,11	15,47	14,11	15,32	231150000
13/01/2020	15,32	18,43	14,99	17,36	266286000
20/01/2020	17,39	18,35	16,48	17,43	248646000
27/01/2020	17,44	18,84	17,27	18,44	256209000

Tabel 1 menunjukkan data historis *Cryptocurrency* BNB yang digunakan dalam penelitian. Data tersebut terdiri dari beberapa atribut, yaitu *Date* sebagai tanggal pencatatan data, *Open* sebagai harga pembukaan, *High* sebagai harga tertinggi, *Low* sebagai harga terendah, *Close* sebagai harga penutupan, dan *Volume* sebagai jumlah transaksi perdagangan dalam periode tertentu. Data historis ini digunakan sebagai dasar dalam proses analisis dan prediksi harga *Cryptocurrency* BNB karena setiap variabel memiliki pengaruh terhadap pergerakan harga.

Tabel 2 Data historis ETH

<i>Date</i>	<i>Open</i>	<i>High</i>	<i>Low</i>	<i>Close</i>	<i>Volume</i>
06/01/2020	136,3	146,82	135,52	145,87	8528000000
13/01/2020	146,22	178,52	142,9	166,96	13713000000
20/01/2020	166,9	171	156,74	168,07	9262000000
27/01/2020	168	193,08	166,9	188,61	14054000000

Tabel 2 memperlihatkan sampel data historis *Cryptocurrency* ETH yang digunakan sebagai bahan penelitian dalam proses ini. Data yang ditampilkan mencakup beberapa variabel penting, yaitu harga pembukaan (*Open*), harga tertinggi (*High*), harga terendah (*Low*), harga penutupan (*Close*), serta *Volume* transaksi perdagangan (*Volume*) pada setiap periode pengamatan. Variabel-variabel tersebut digunakan untuk menggambarkan kondisi pergerakan pasar ETH dari waktu ke waktu. Data *Cryptocurrency* memiliki karakteristik fluktuatif dengan volatilitas tinggi, sehingga diperlukan metode pemodelan yang mampu menangkap pola pergerakan secara kompleks. Sebelum pemodelan, data melalui tahap pra-pemrosesan yang meliputi konversi format waktu, pengurutan kronologis, dan penghapusan nilai yang hilang. *Outlier* ditangani menggunakan metode *Interquartile Range* (IQR). Selanjutnya, diterapkan pendekatan *lag* dengan menggunakan variabel pada periode sebelumnya ($t-1$) untuk memprediksi nilai *Close* pada periode saat ini (t) guna menghindari data *leakage*. Data kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20 secara kronologis.

3.1 Preprocessing Data

Pada tahap preprocessing data *Random Forest Regression* dan *Linear Regression*, penanganan *Outlier* dilakukan menggunakan metode *Interquartile Range* (IQR) untuk mengurangi pengaruh nilai ekstrem yang dapat mengganggu kinerja model. Proses ini diawali dengan menghitung kuartil pertama (Q1) dan kuartil ketiga (Q3) dari setiap variabel, kemudian menentukan nilai IQR sebagai selisih antara Q3 dan Q1. Selanjutnya, batas bawah dan batas atas ditentukan masing-masing sebesar $Q1 - 1,5 \times IQR$ dan $Q3 + 1,5 \times IQR$. Data yang berada di luar rentang tersebut dianggap sebagai *Outlier* dan dihapus dari dataset. Penerapan metode ini bertujuan untuk menjaga kualitas data dengan menghilangkan nilai yang tidak representatif, tanpa mengubah distribusi utama data secara signifikan. Dengan demikian, data yang digunakan dalam proses pemodelan menjadi lebih stabil dan mampu merepresentasikan pola pergerakan harga *Cryptocurrency* secara lebih akurat. Hasil tersebut ditunjukkan pada tabel-tabel berikut.

Table 3 Data Sebelum IQR (BNB)

Statistik	<i>Open</i>	<i>High</i>	<i>Low</i>	<i>Close</i>	<i>Volume</i>
count	310	310	310	310	310
mean	392.708613	407.033874	353.152337	383.329065	11217910000

std	346.921399	273.185101	238.887338	258.738805	9553715000
min	10.420000	13.270000	0.214400	10.430000	138494000
max	4450.260000	1370.540000	1054.070000	1303.110000	61537000000

Tabel 3 menunjukkan data BNB sebelum dilakukan proses *Interquartile Range* (IQR). Statistik deskriptif ini digunakan untuk memberikan gambaran umum mengenai distribusi data pada setiap variabel, yaitu *Open*, *High*, *Low*, *Close*, dan *Volume*. Berdasarkan tabel tersebut, jumlah data (*count*) pada setiap variabel sebanyak 310 data.

Table 4 Data Sesudah IQR (BNB)

Statistik	<i>Open</i>	<i>High</i>	<i>Low</i>	<i>Close</i>	<i>Volume</i>
count	291	291	291	291	291
mean	364.937388	387.940210	341.532558	367.998179	9545686000
std	244.125848	255.615884	228.036479	243.776694	6554963000
min	10.420000	13.270000	0.214400	10.430000	138494000
max	1084.330000	1086.590000	933.500000	1048.670000	27158000000

Tabel 4 merupakan hasil penerapan metode *Interquartile Range* (IQR) pada dataset *Cryptocurrency* BNB menunjukkan adanya penghapusan data *outlier* untuk meningkatkan kualitas data. Sebanyak 19 data berhasil dihapus pada tahap *preprocessing* karena teridentifikasi sebagai data yang tidak sesuai atau termasuk *Outlier*, sehingga jumlah data yang digunakan dalam proses penelitian menjadi 291 data. Setelah dilakukan pembersihan data, diperoleh nilai rata-rata (*mean*) pada variabel *Close* sebesar 367.99, yang menunjukkan rata-rata harga penutupan BNB. Selain itu, nilai median sebesar 319.60 menunjukkan bahwa sebagian besar data harga penutupan berada di sekitar nilai tersebut

Table 5 Data Sebelum IQR (ETH)

Statistik	<i>Open</i>	<i>High</i>	<i>Low</i>	<i>Close</i>	<i>Volume</i>
count	310	310	310	310	310
mean	2151.390871	2323.416255	1981.490387	2168.336032	128538200000
std	1186.849649	1262.310641	1084.119784	1183.352219	108167400000
min	123.360000	141.940000	3.290000	123.320000	5292000000
max	4779.760000	4953.730000	4485.090000	4779.640000	131200000000

Tabel 5 menunjukkan statistik deskriptif data *Cryptocurrency* ETH sebelum dilakukan proses *Interquartile Range* (IQR). Statistik deskriptif tersebut digunakan untuk melihat karakteristik dan penyebaran data pada setiap variabel, yaitu *Open*, *High*, *Low*, *Close*, dan *Volume*. Berdasarkan tabel, jumlah data pada masing-masing variabel sebanyak 310 data. Nilai rata-rata (*mean*) pada variabel *Close* sebesar 2168.336032 dengan standar deviasi sebesar 1183.352219, yang menunjukkan bahwa pergerakan harga ETH memiliki tingkat variasi yang cukup tinggi.

Table 6 Data Sesudah IQR (ETH)

Statistik	<i>Open</i>	<i>High</i>	<i>Low</i>	<i>Close</i>	<i>Volume</i>
count	300	300	300	300	300
mean	2099.063233	2266.512797	1943.386333	2122.039467	117309400000
std	1159.373447	1231.352429	1068.658351	1160.073040	73970400000
min	123.360000	141.940000	3.290000	123.320000	5292000000
max	4779.760000	4891.700000	4485.090000	4626.350000	308652000000

Berdasarkan metode *Interquartile Range* (IQR) pada Tabel 6 data *Cryptocurrency* ETH menunjukkan adanya penghapusan data *outlier* untuk meningkatkan kualitas data. Pada data

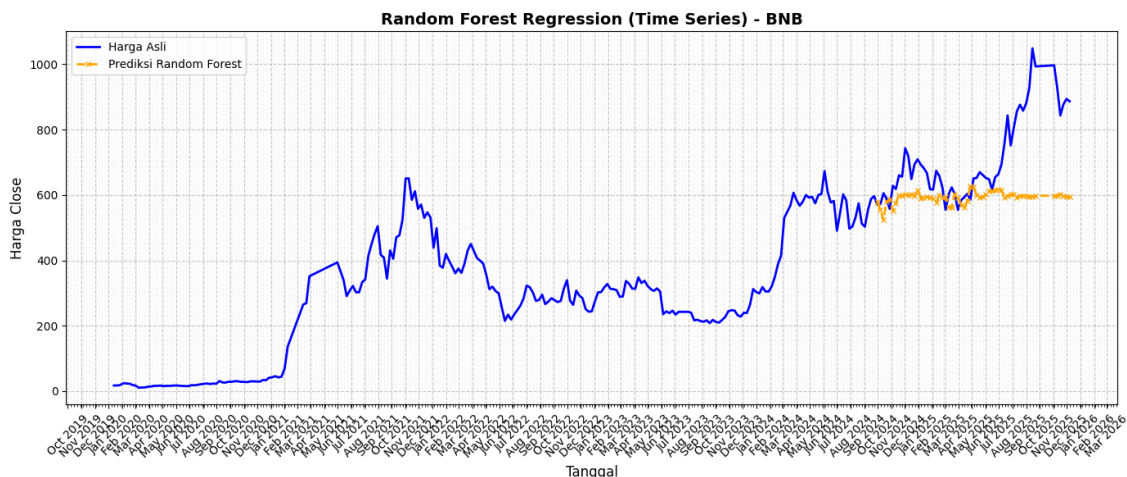
ETH, sebanyak 10 data berhasil dihapus pada tahap preprocessing karena teridentifikasi memiliki nilai yang menyimpang dari pola data utama, sehingga jumlah data yang digunakan dalam penelitian menjadi 300 data. Setelah proses pembersihan data dilakukan, diperoleh nilai rata-rata (*mean*) pada variabel *Close* sebesar 2122.03 yang menunjukkan rata-rata harga penutupan ETH selama periode pengamatan. Selain itu, nilai median sebesar 2029.34 menunjukkan bahwa sebagian besar data harga penutupan berada di sekitar nilai tersebut, sehingga distribusi data menjadi lebih stabil dan lebih baik untuk digunakan. Proses IQR berhasil mengurangi nilai ekstrem tanpa mengubah distribusi utama data, sementara standar deviasi yang relatif tinggi menunjukkan bahwa karakteristik volatilitas tetap terjaga.

Selanjutnya, dilakukan transformasi lag dengan menggunakan variabel pada periode sebelumnya ($t-1$) untuk memprediksi nilai *Close* pada periode saat ini (t), sehingga menghindari data leakage. Data kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20 secara kronologis.

3.2 Penerapan Model dan Hasil Prediksi

3.2.1 *Random Forest Regression*

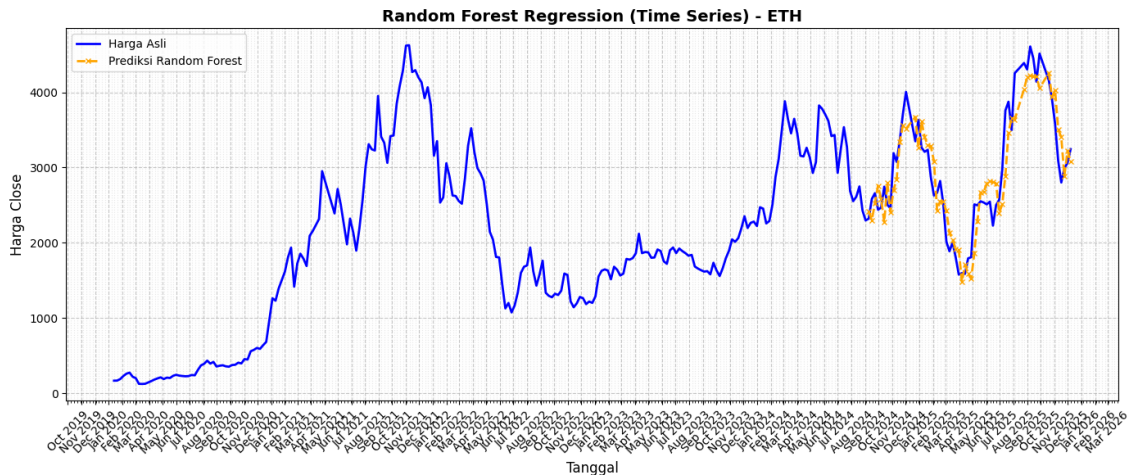
Pada tahap ini, model yang telah dilatih menggunakan data latih (training data) digunakan untuk melakukan prediksi pada data uji (testing data). Proses prediksi dilakukan untuk memperoleh nilai estimasi harga penutupan (*Close*) berdasarkan pola yang telah dipelajari. Hasil prediksi yang diperoleh kemudian disusun bersama data aktual sehingga dapat dilakukan perbandingan secara langsung antara nilai prediksi dan nilai sebenarnya pada periode pengujian. Perbandingan ini digunakan untuk melihat kesesuaian pola pergerakan antara hasil prediksi model dan data aktual dalam bentuk deret waktu. Berikut grafik perbandingannya.



Gambar 2 Grafik Hasil Prediksi dan Perbandingan Data (BNB)

Berdasarkan Gambar 2, garis berwarna biru menunjukkan data harga aktual, sedangkan garis berwarna oranye menunjukkan hasil prediksi *Cryptocurrency* BNB menggunakan metode *Random Forest Regression*. Pada periode awal pengamatan, khususnya September hingga Oktober 2024, garis prediksi terlihat cukup dekat dengan garis harga aktual sehingga model masih mampu mengikuti pola umum pergerakan harga. Namun, mulai November 2024 hingga akhir periode pengamatan, terlihat adanya peningkatan jarak antara kedua garis. Garis prediksi cenderung berada pada kisaran nilai yang relatif tetap, sedangkan harga aktual mengalami fluktuasi yang lebih besar dan menunjukkan tren kenaikan yang cukup tajam. Perbedaan paling mencolok terlihat pada periode Juli hingga September 2025, ketika harga aktual meningkat

signifikan hingga melampaui nilai 1.000, sementara hasil prediksi tetap berada di sekitar kisaran 595–603. Secara keseluruhan dapat disimpulkan bahwa model masih belum cukup baik dalam mengikuti pola harga pada kondisi stabil, hal ini ditunjukkan oleh garis prediksi berwarna oranye yang cenderung tidak mengikuti perubahan garis harga aktual berwarna biru secara konsisten. Grafik di atas mulai mengalami deviasi yang lebih besar ketika terjadi fluktuasi harga yang ekstrem.



Gambar 3 Grafik Hasil Prediksi dan Perbandingan Data (ETH)

Berdasarkan Gambar 3, garis berwarna biru menunjukkan data harga aktual *Ethereum* (ETH), sedangkan garis berwarna oranye menunjukkan hasil prediksi menggunakan metode *Random Forest Regression*. Pada beberapa periode awal pengamatan, khususnya September hingga Oktober 2024, garis prediksi terlihat cukup dekat dengan garis harga aktual sehingga model masih mampu mengikuti pola umum pergerakan harga ETH. Namun, mulai November 2024 hingga awal tahun 2025, terlihat adanya peningkatan jarak antara kedua garis, yang menunjukkan bahwa hasil prediksi mulai menyimpang dari data aktual. Perbedaan tersebut semakin jelas ketika harga ETH mengalami penurunan tajam pada Maret hingga April 2025, di mana garis prediksi tidak sepenuhnya mengikuti arah penurunan harga aktual. Selanjutnya, pada periode Mei hingga Juli 2025 terjadi kenaikan harga yang cukup signifikan, terutama pada pertengahan Juli 2025 ketika harga aktual meningkat hingga sekitar 3.759,47, sementara hasil prediksi hanya berada di kisaran 2.890,06. Pada Agustus hingga September 2025, harga aktual kembali bergerak naik hingga mendekati 4.600, sedangkan garis prediksi tetap berada pada kisaran 4.200-an, sehingga selisih antara kedua garis masih terlihat cukup besar.

Hasil prediksi harga *Cryptocurrency* BNB dan ETH menggunakan *Random Forest Regression* menunjukkan bahwa model mampu mengikuti pola umum pergerakan harga, namun tingkat akurasi masih bervariasi. Pada BNB, hasil prediksi cenderung mengikuti arah harga aktual, tetapi terdapat selisih yang cukup besar pada beberapa periode akibat fluktuasi ekstrem. Sementara itu, pada ETH, prediksi terlihat lebih stabil dan lebih mendekati nilai aktual pada beberapa periode, meskipun tetap terjadi deviasi saat perubahan harga yang tajam. Secara keseluruhan, model *Random Forest* cukup baik dalam menangkap pola utama pergerakan harga, namun masih memiliki keterbatasan dalam menghadapi volatilitas tinggi pada pasar *Cryptocurrency*.

3.2.2 Evaluasi Model

Pada perhitungan MSE dilakukan dengan menghitung selisih antara nilai aktual dan nilai prediksi pada setiap data pengujian, kemudian selisih tersebut dikuadratkan dan dirata-ratakan. Sebagai contoh data BNB pada tanggal 30 September 2024 data ke-247.

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

y_i = nilai aktual

$\hat{y}_i$ = nilai prediksi

$e_i = y_i - \hat{y}_i$

e_i^2 = kuadrat galat

n = jumlah data pengujian

Nilai aktual dan prediksi data ke-247:

$$y_{247} = 570,54$$

$$\widehat{y}_{247} = 580,1537$$

Perhitungan galat:

$$e_{247} = y_{247} - \widehat{y}_{247}$$

$$e_{247} = 570,54 - 580,1537 = -9,6137$$

$$e_{247}^2 = (-9,6137)^2 = 92,42$$

Perhitungan yang sama dilakukan untuk seluruh 58 data pengujian sehingga diperoleh jumlah kuadrat galat:

$$\sum_{i=1}^{58} e_i^2 = 1.652.681,56$$

Karena jumlah data pengujian adalah $n = 58$, maka nilai MSE dihitung sebagai:

$$MSE = \frac{1.652.681,56}{58} = 28.494,51$$

Sehingga diperoleh:

$$MSE_{BNB} = 28.494,51$$

Berdasarkan hasil perhitungan menggunakan rumus *Mean Squared Error* (MSE), diperoleh nilai MSE pada data BNB sebesar 28.494,51. Nilai tersebut menunjukkan bahwa rata-rata kuadrat kesalahan antara harga aktual dan hasil prediksi model *Random Forest* adalah sebesar 28.494,51. Dengan menggunakan prosedur perhitungan MSE yang sama seperti pada data BNB, diperoleh nilai MSE data ETH sebesar 108191,84.

Selanjutnya perhitungan *koefisien determinasi* (R^2) dilakukan dengan membandingkan jumlah kuadrat galat prediksi terhadap jumlah kuadrat variasi data aktual terhadap rata-ratanya. Sebagai contoh pada data BNB tanggal 15 September 2025,

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

y_i = nilai aktual

$\hat{y}_i$ = nilai prediksi

$\bar{y}$ = rata-rata nilai aktual

Dari data BNB diperoleh rata-rata nilai aktual:

$$\bar{y} = 708,44$$

Nilai aktual dan prediksi data ke-297:

$$y_{297} = 1048,67$$

$$\widehat{y}_{297} = 595,5937$$

Hitung galat prediksi:

$$e_{297} = y_{297} - \widehat{y}_{297}$$

$$e_{297} = 1048,67 - 595,5937 = 453,0763$$

Hitung kuadrat galat:

$$e_{297}^2 = (453,0763)^2 = 205.278,16$$

Hitung variasi terhadap rata-rata:

$$(y_{297} - \bar{y})^2 = (1048,67 - 708,44)^2$$

$$= (340,23)^2 = 115.756,45$$

Perhitungan yang sama dilakukan untuk seluruh 58 data pengujian sehingga diperoleh:

Jumlah kuadrat galat (RSS)

$$RSS = \sum_{i=1}^{58} (y_i - \widehat{y}_i)^2 = 1.652.681,56$$

Jumlah kuadrat variasi data (TSS)

$$TSS = \sum_{i=1}^{58} (y_i - \bar{y})^2 = 925.654,31$$

Selanjutnya nilai R² dihitung sebagai berikut:

$$R^2 = 1 - \frac{1.652.681,56}{925.654,31}$$

$$R^2 = 1 - 1,7854$$

$$R^2 = -0,7854$$

Sehingga diperoleh:

$$R_{BNB}^2 = -0,7854$$

Karena kuadrat galat lebih besar daripada variasi data terhadap rata-ratanya, nilai *Residual Sum of Squares* (RSS) menjadi lebih besar daripada *Total Sum of Squares* (TSS), Setelah seluruh data pengujian dihitung dengan prosedur yang sama dan disubstitusikan ke dalam rumus, diperoleh nilai R² sebesar -0,7854. Dengan menggunakan prosedur perhitungan *koefisien determinasi* (R²) yang sama seperti pada data BNB, diperoleh nilai R² data ETH sebesar 0,8309. Nilai tersebut menunjukkan bahwa sekitar 83,09% variasi harga ETH dapat dijelaskan oleh model *Random Forest*, sehingga model memiliki kemampuan yang baik dalam merepresentasikan pola pergerakan harga ETH.

Fenomena R^2 negatif ekstrem pada BNB ini juga dapat dijelaskan oleh sifat matematis *Random Forest* itu sendiri. Algoritma ini memprediksi nilai baru dengan merata-ratakan nilai target pada *leaf node* dari setiap pohon keputusan, sehingga prediksinya secara struktural bersifat *bounded* dan tidak dapat melampaui rentang nilai target yang pernah diamati selama proses training. Keterbatasan ekstrapolasi pada model berbasis pohon keputusan ini telah didokumentasikan secara luas dalam literatur pemodelan deret waktu [16], [17]. Karena skema pembagian data pada penelitian ini bersifat kronologis (80:20), seluruh data uji BNB berada pada periode ketika harga melonjak melampaui harga maksimum yang pernah terlihat model selama training, sehingga *Random Forest* secara sistematis melakukan *underprediction* pada seluruh periode uji tersebut dan menghasilkan galat yang sangat besar. Pola ini tidak muncul secara signifikan pada ETH karena harga maksimum pada periode uji ETH masih berada dalam rentang nilai yang telah diamati model selama training. Sebaliknya, *Linear Regression* tidak memiliki keterbatasan ekstrapolasi tersebut karena bentuk fungsinya yang linear memungkinkan model untuk memprediksi nilai di luar rentang data training, sehingga model ini justru lebih unggul dibandingkan *Random Forest* khusus pada kasus BNB, dan bukan karena adanya kesalahan pada implementasi kode.

Pada perhitungan MAPE dilakukan dengan menghitung persentase kesalahan absolut pada setiap data pengujian, kemudian seluruh persentase kesalahan dirata-ratakan. Sebagai contoh pada data BNB tanggal 15 September 2025:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

y_i = nilai aktual

$\hat{y}_i$ = nilai prediksi

n = jumlah data pengujian

Nilai actual dan prediksi data ke-297:

$$y_{297} = 1048,67$$

$$\widehat{y}_{297} = 595,5937$$

Menghitung selisih absolut:

$$|y_{297} - \widehat{y}_{297}| = |1048,67 - 595,5937| = 453,0763$$

$$\left| \frac{y_{297} - \widehat{y}_{297}}{y_{297}} \right|$$

$$= \frac{453,0763}{1048,67}$$

$$= 0,4320485$$

Ubah ke persentase:

$$0,4320485 \times 100\% = 43,20\%$$

Jadi, nilai *Absolute Percentage Error* (APE) untuk data ke-297 adalah:

$$APE_{297} = 43,20\%$$

Perhitungan yang sama dilakukan untuk seluruh 58 data pengujian. Setelah seluruh nilai APE dijumlahkan diperoleh:

$$\sum_{i=1}^{58} APE_i = 871,16$$

Karena jumlah pengujian adalah 58:
 $n = 58$

Maka:

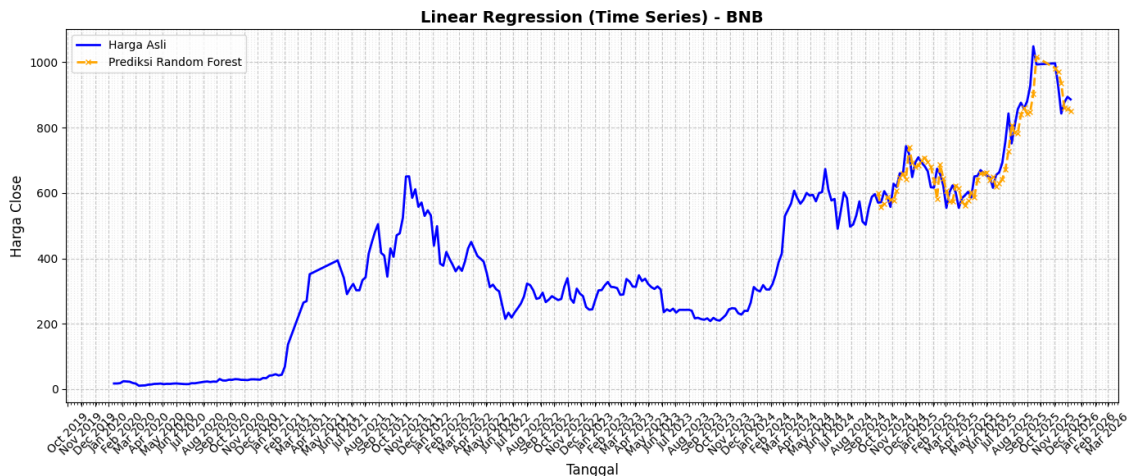
$$MAPE = \frac{871,16}{58} = 15,02\%$$

$$MAPE_{BNB} = 15,02\%$$

Berdasarkan hasil perhitungan menggunakan rumus *Mean Absolute Percentage Error* (MAPE), diperoleh nilai MAPE pada data BNB sebesar 15,02%. Nilai tersebut menunjukkan bahwa rata-rata persentase kesalahan prediksi model *Random Forest* terhadap harga aktual BNB adalah sekitar 15,02%. Perhitungan nilai MAPE pada data ETH dilakukan dengan prosedur yang sama seperti pada data BNB, yaitu dengan menghitung persentase kesalahan absolut antara nilai aktual dan nilai prediksi pada setiap data pengujian, kemudian merata-ratakannya. Berdasarkan hasil perhitungan diperoleh nilai MAPE sebesar 9,65%, yang menunjukkan bahwa rata-rata persentase kesalahan prediksi model *Random Forest* terhadap harga aktual ETH adalah 9,65%.

3.2.3 Linear Regression

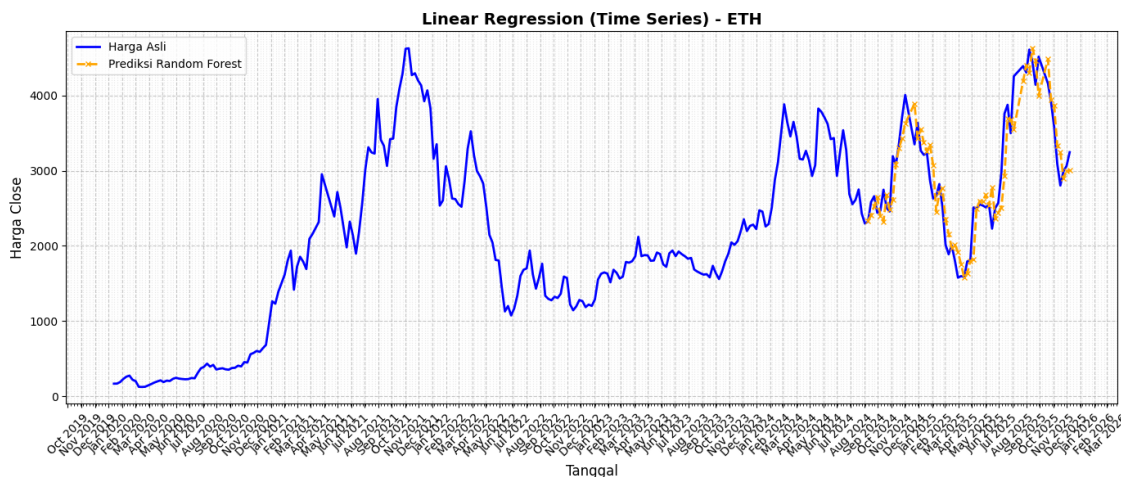
Pada model *Linear Regression*, tahap ini dilakukan dengan menggunakan model yang telah dilatih pada data latih untuk melakukan prediksi pada data uji guna menghasilkan estimasi harga penutupan (*Close*) berdasarkan pola yang telah dipelajari. Hasil prediksi kemudian dibandingkan dengan data aktual untuk melihat kesesuaian pola pergerakan harga pada periode pengujian. Berikut grafik perbandingannya.



Gambar 4 Grafik Hasil Prediksi dan Perbandingan Data (BNB)

Berdasarkan Gambar 4, garis berwarna biru menunjukkan data harga aktual *Binance Coin* (BNB), sedangkan garis berwarna oranye menunjukkan hasil prediksi menggunakan metode *Linear Regression*. Secara visual, garis prediksi tampak lebih mengikuti arah pergerakan harga dibandingkan model sebelumnya, terutama pada periode Oktober 2024 hingga pertengahan tahun 2025. Pada beberapa titik pengamatan, seperti Oktober hingga November 2024 serta Mei hingga

Juni 2025, garis prediksi terlihat cukup dekat dengan garis harga aktual sehingga model mampu merepresentasikan pola pergerakan harga dengan baik pada kondisi pasar yang relatif stabil. Namun, mulai Juli hingga September 2025, ketika harga BNB mengalami kenaikan yang lebih tajam, terlihat adanya peningkatan jarak antara garis prediksi dan garis aktual. Perbedaan tersebut tampak jelas pada pertengahan September 2025 ketika harga aktual mencapai sekitar 1.048,67, sedangkan hasil prediksi berada di sekitar 903,31. Meskipun demikian, pada beberapa periode setelahnya, seperti akhir Agustus hingga November 2025, garis prediksi kembali mendekati garis harga aktual sehingga deviasi yang terjadi relatif lebih kecil dibandingkan saat terjadi lonjakan harga ekstrem. Secara keseluruhan dapat disimpulkan bahwa model *Linear Regression* mampu mengikuti tren umum pergerakan harga BNB, yang ditunjukkan oleh garis prediksi berwarna oranye yang relatif konsisten mengikuti arah garis harga aktual berwarna biru. Akan tetapi, grafik di atas masih menunjukkan adanya deviasi yang lebih besar ketika terjadi fluktuasi harga yang sangat tajam, sehingga akurasi model cenderung menurun pada periode dengan volatilitas pasar yang tinggi.



Gambar 5 Grafik Hasil Prediksi dan Perbandingan Data (ETH)

Berdasarkan Gambar 5, garis berwarna biru menunjukkan data harga aktual *Ethereum* (ETH), sedangkan garis berwarna oranye menunjukkan hasil prediksi menggunakan metode *Linear Regression*. Secara visual, garis prediksi terlihat cukup mengikuti arah pergerakan harga ETH pada beberapa periode pengamatan, terutama pada November 2024 hingga awal Januari 2025 serta pada Mei hingga September 2025. Kedekatan antara garis prediksi dan garis harga aktual pada periode tersebut menunjukkan bahwa model mampu menangkap tren umum pergerakan harga ketika perubahan harga tidak terlalu ekstrem. Namun, pada beberapa periode tertentu masih terlihat deviasi yang cukup besar antara kedua garis. Perbedaan yang mencolok terjadi pada awal November 2024 ketika harga aktual meningkat hingga sekitar 3.191,33, sedangkan hasil prediksi berada di sekitar 2.613,36. Deviasi besar juga tampak pada Maret hingga April 2025 ketika harga ETH mengalami penurunan tajam, di mana garis prediksi cenderung berada di atas garis harga aktual. Selanjutnya, pada pertengahan Juli hingga awal Agustus 2025, harga ETH kembali mengalami lonjakan signifikan hingga mencapai sekitar 4.254,21, sementara hasil prediksi masih berada di kisaran 3.549,62, sehingga jarak antara kedua garis terlihat semakin lebar. Meskipun demikian, pada beberapa titik lain seperti pertengahan Mei 2025, awal September 2025, dan akhir Oktober 2025, garis prediksi kembali mendekati garis harga aktual sehingga model menunjukkan kemampuan yang cukup baik dalam merepresentasikan tren harga. Secara keseluruhan dapat disimpulkan bahwa model *Linear Regression* mampu mengikuti pola tren harga ETH dengan cukup baik pada kondisi pasar yang relatif stabil, yang ditunjukkan oleh garis prediksi berwarna oranye yang umumnya bergerak searah dengan garis harga aktual berwarna biru.

biru. Akan tetapi, grafik di atas menunjukkan bahwa deviasi antara kedua garis meningkat ketika terjadi fluktuasi harga yang tajam, sehingga performa model cenderung menurun pada periode dengan volatilitas pasar yang tinggi.

3.2.4 Evaluasi Model

Berdasarkan prosedur perhitungan MSE, R^2 , dan MAPE yang telah dijelaskan pada bagian sebelumnya di evaluasi model *Random Forest Regression*, model Linear Regression pada data BNB menghasilkan nilai MSE sebesar 2262,69, R^2 sebesar 0,8582, dan MAPE sebesar 5,14%. Nilai tersebut menunjukkan bahwa model mampu menjelaskan sekitar 85,82% variasi harga BNB dengan tingkat kesalahan prediksi yang relatif rendah. Sementara itu, pada data ETH diperoleh nilai MSE sebesar 94079,76, R^2 sebesar 0,8529, dan MAPE sebesar 8,17%, yang menunjukkan bahwa model mampu menjelaskan sekitar 85,29% variasi harga ETH. Meskipun nilai MSE pada ETH lebih besar dibandingkan BNB, nilai R^2 yang hampir sama menunjukkan bahwa model tetap mampu menangkap pola pergerakan harga ETH dengan baik. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa model Linear Regression memiliki performa yang cukup baik dan stabil pada kedua jenis data *cryptocurrency* yang digunakan dalam penelitian ini.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa model *Random Forest Regression* dan *Linear Regression* memiliki performa yang berbeda dalam memprediksi harga *Cryptocurrency* BNB dan ETH. Pada model *Random Forest Regression*, hasil evaluasi menunjukkan bahwa pada data BNB model belum mampu memberikan hasil prediksi yang baik yang ditunjukkan oleh nilai MSE yang cukup tinggi serta nilai R^2 yang negatif, sehingga model tidak dapat menjelaskan variasi data secara optimal. Sebaliknya, pada data ETH model ini menunjukkan kinerja yang lebih baik dengan nilai R^2 yang tinggi dan MAPE yang lebih rendah, yang mengindikasikan bahwa model lebih sesuai digunakan pada data ETH dibandingkan BNB dalam penelitian ini. Sementara itu, pada model *Linear Regression*, hasil evaluasi menunjukkan performa yang lebih stabil dan konsisten pada kedua aset. Pada data BNB, model mampu menjelaskan sebagian besar variasi data dengan nilai R^2 yang tinggi serta MAPE yang rendah, sedangkan pada data ETH meskipun nilai error lebih besar, model tetap mampu menangkap pola pergerakan harga dengan cukup baik karena nilai R^2 masih berada pada tingkat yang tinggi. Secara keseluruhan, *Linear Regression* menunjukkan kinerja yang lebih konsisten dibandingkan *Random Forest* dalam penelitian ini, terutama dalam hal kestabilan hasil prediksi pada dua jenis data yang berbeda dan disebabkan oleh keterbatasan algoritma berbasis pohon keputusan yang tidak mampu melakukan ekstrapolasi di luar rentang data training, sehingga model cenderung melakukan underprediction ketika harga pada data uji meningkat tajam pada data BNB. Oleh karena itu, dapat disimpulkan bahwa pemilihan model sangat bergantung pada karakteristik data, khususnya tingkat volatilitasnya, di mana data yang lebih stabil cenderung lebih cocok menggunakan model linear, sedangkan data yang lebih kompleks dan fluktuatif memerlukan pendekatan model yang lebih adaptif. Penelitian selanjutnya disarankan untuk mengembangkan model dengan menambahkan variabel eksternal seperti indikator teknikal, atau faktor sentimen pasar agar hasil prediksi dapat lebih akurat dan mampu merepresentasikan kondisi pasar *Cryptocurrency* yang sangat dinamis.

REFERENCES

- [1] Kadek Dyah Pramitha Widyarani, Ida Ayu Putu Widiati, and Ni Made Puspasutari Ujianti, "Kajian Yuridis Penggunaan Koin Kripto sebagai Alat Pembayaran di Indonesia," *Jurnal Preferensi Hukum*, vol. 3, no. 2, pp. 300–305, Apr. 2022, doi: 10.55637/jph.3.2.4934.300-305.

- [2] M. A. Tambun *et al.*, “TATA KELOLA PEMBENTUKAN REGULASI TERKAIT PERDAGANGAN MATA UANG KRIPTO (CRYPTOCURRENCY) SEBAGAI ASET KRIPTO (CRYPTO ASSET),” 2022. [Online]. Available: <http://www.hit.bme.hu/~buttyan/courses/BMEVIHIM219/2009/Chaum.BlindSigForPayment>
- [3] M. Najibur Rohman, “Jurnal Ilmiah Ilmu Hukum TINJAUAN YURIDIS NORMATIF TERHADAP REGULASI MATA UANG KRIPTO (CRYPTO CURRENCY) DI INDONESIA,” *Jurnal Supremasi*, vol. 11, no. 2, 2021, [Online]. Available: <https://ejournal.unisbablitar.ac.id/index.php/supremasi>
- [4] M. D. Tria, “ANALISIS RISIKO HARGA SAHAM SYARIAH MENGGUNAKAN DAILY EARNING AT RISK (DEAR) TERHADAP MINAT INVESTOR (Studi pada harga saham harian di Bursa Efek Indonesia Saham Jakarta Islamic Index Periode 2022),” 2024.
- [5] M. Jiménez-Carrión, G. A. Flores-Fernandez, and A. B. Jiménez-Panta, “Efficient Transit Network Design, Frequency Adjustment, and Fleet Calculation Using Genetic Algorithms,” *Journal of Internet Services and Information Security*, vol. 13, no. 4, pp. 26–49, Nov. 2023, doi: 10.58346/JISIS.2023.14.003.
- [6] M. Iqbal, H. M. Nawawi, M. Rezki, A. Hamid, S. Rahayu, and C. Author, “Pendekatan Algoritma Klasifikasi Machine Learning untuk Deteksi Penyakit Demensia,” 2023. [Online]. Available: <http://jurnal.bsi.ac.id/index.php/co-science>
- [7] I. Priyadi, J. Santony, and J. Na’am, “Data Mining Predictive Modeling for Prediction of Gold Prices Based on Dollar Exchange Rates, Bi Rates and World Crude Oil Prices,” *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 2, no. 2, p. 93, Sep. 2019, doi: 10.24014/ijaidm.v2i2.6864.
- [8] A. Roihan, P. Abas Sunarya, and A. S. Rafika, “Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper,” *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 5, no. 1, pp. 75–82, Mar. 2020, Accessed: Jun. 22, 2026. [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ijcit>
- [9] Syarif Aminul Khoiri and Abdul Wahid, “Analisis Kinerja Algoritma Machine Learning dalam Prediksi Harga Cryptocurrency,” *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*, vol. 9, no. 2, pp. 133–141, Aug. 2024, doi: 10.32528/justindo.v9i2.1965.
- [10] R. Saputra *et al.*, “ANALISIS PREDIKSI SAHAM TESLA MENGGUNAKAN ALGORITMA LONG SHORT TERM MEMORY (LSTM),” *Journal of Computer Science and Information Technology (JCSIT)*, vol. 2, no. 1, Dec. 2024, doi: <https://doi.org/10.70248/jcsit.v2i1.1482>.
- [11] S. Saadah and H. Salsabila, “Jurnal Politeknik Caltex Riau Prediksi Harga Bitcoin Menggunakan Metode Random Forest (Studi Kasus: Data Acak Pada Awal Masa Pandemic Covid-19),” 2021. [Online]. Available: <https://jurnal.pcr.ac.id/index.php/jkt/>
- [12] R. Dona Ariyatma and S. Fahmi, “Data Mining Menggunakan Multiple Regression untuk Prediksi Harga Saham Netflix,” *Jurnal SAINTEKOM*, vol. 13, no. 2, pp. 184–192, Sep. 2023, doi: 10.33020/saintekom.v13i2.419.
- [13] M. D. S. Oktavian, R. Parlika, and F. P. Aditiawan, “Seminar Nasional Informatika Bela Negara (SANTIKA) Analisis Studi Literatur: Perbandingan Model Prediksi Harga Bitcoin Berbasis Long Short-Term Memory (LSTM),” *Seminar Nasional Informatika Bela Negara (SANTIKA)*, vol. 5, Dec. 2025, Accessed: Jun. 22, 2026. [Online]. Available: <https://santika.upnjatim.ac.id/submissions/index.php/santika/article/view/879>
- [14] N. S. Pulungan, “Analisis hubungan faktor ekonomi makro dengan indeks harga saham syariah menggunakan Vector Error Correction Model (VECM),” *UIN Syekh Ali Hasan Ahmad Addary Padangsidempuan*, 2024, Accessed: Jun. 17, 2026. [Online]. Available: <http://etd.uinsyahada.ac.id/id/eprint/11128>

- [15] I. W. Sucipta, I. M. Candiasa, and I. G. Sudirtha, “PENGARUH MODEL PEMBELAJARAN BERBASIS MASALAH DAN BENTUK ASESMEN FORMATIF TERHADAP KEMAMPUAN BERPIKIR KRITIS,” *Jurnal Penelitian dan Evaluasi Pendidikan Indonesia*, vol. 13, no. 2, Aug. 2023, doi: <https://doi.org/10.23887/jpepi.v13i2.2660>.
- [16] S. Zhang, Z. Lu, and C. Zheng, “A Simple and Effective Random Forest Modelling for Nonlinear Time Series Data,” Nov. 2025, doi: <https://doi.org/10.48550/arXiv.2511.06544>.
- [17] R. A. Davis and M. S. Nielsen, “Modeling of time series using random forests: Theoretical developments,” *Electron. J. Stat.*, vol. 14, no. 2, pp. 3644–3671, 2020, doi: 10.1214/20-EJS1758.