



## **Analisa Peforma Algoritma Machine Learning Dalam Prediksi Harga Pasar Saham BTC**

**Dimas Indra Pratama<sup>1</sup>, Aditya Hafidzh<sup>2</sup>, Muhamad Fajar Ramadhan<sup>3</sup>, Irgi Irsandi  
Ramadhan<sup>4</sup>, Ramadhan Oktovio A<sup>5</sup>, Dwiky Satria Gymnastiar<sup>6</sup>, Ines Heidiani Ikasari<sup>7\*</sup>**

<sup>1,2,3,4,5,6,7</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang  
Selatan, Indonesia

Email: <sup>1</sup>[adityahafizhn@gmail.com](mailto:adityahafizhn@gmail.com), <sup>2</sup>[dimas.indra773@gmail.com](mailto:dimas.indra773@gmail.com), <sup>3</sup>[fajarramadhan0800@gmail.com](mailto:fajarramadhan0800@gmail.com),  
<sup>4</sup>[irgiirsandiramadhan@gmail.com](mailto:irgiirsandiramadhan@gmail.com), <sup>5</sup>[aryawanramadhan@gmail.com](mailto:aryawanramadhan@gmail.com), <sup>6</sup>[dwikysatria12@gmail.com](mailto:dwikysatria12@gmail.com),  
<sup>7\*</sup>[dosen01374@unpam.ac.id](mailto:dosen01374@unpam.ac.id)  
(\* : coressponding author)

**Abstrak** - Penelitian ini bertujuan untuk mengevaluasi performa algoritma machine learning dalam memprediksi harga saham BTC. Beberapa algoritma seperti Logistic Regression, Linear Regression, K-Nearest Neighbors (KNN), Decision Tree, dan Random Forest diuji menggunakan metrik seperti akurasi, Mean Squared Error (MSE), dan R-squared ( $R^2$ ). Hasil evaluasi menunjukkan bahwa Linear Regression memiliki performa terbaik dengan MSE sangat kecil dan  $R^2$  mendekati 1. Model klasifikasi seperti Logistic Regression dan Decision Tree juga menunjukkan hasil yang cukup baik, namun memerlukan peningkatan akurasi lebih lanjut. Penelitian ini memberikan wawasan tentang pendekatan terbaik untuk memprediksi harga saham BTC menggunakan machine learning.

**Kata Kunci:** Logistic Regression; Linear Regression; KNN; Decision Tree; Random Forest

**Abstract** - This research aims to evaluate the performance of machine learning algorithms in predicting BTC stock prices. Several algorithms such as Logistic Regression, Linear Regression, K-Nearest Neighbors (KNN), Decision Tree, and Random Forest are tested using metrics such as accuracy, Mean Squared Error (MSE), and R-squared ( $R^2$ ). The evaluation results show that Linear Regression has the best performance with very small MSE and  $R^2$  close to 1. Classification models such as Logistic Regression and Decision Tree also show quite good results, but require further accuracy improvements. This research provides insight into the best approach to predicting BTC stock prices using machine learning.

**Keywords:** Logistic Regression; Linear Regression; KNN; Decision Trees; Random Forest

### **1. PENDAHULUAN**

Pasar saham dan cryptocurrency seperti Bitcoin (BTC) semakin menarik perhatian dalam bidang keuangan dan teknologi. Studi ini mengevaluasi berbagai algoritma machine learning untuk memprediksi pergerakan harga saham BTC. Logistic Regression, Linear Regression, KNN, Decision Tree, dan Random Forest adalah algoritma utama yang diuji. Penelitian ini memanfaatkan data historis yang mencakup fitur seperti harga pembukaan, harga tertinggi, harga penutupan, dan volume perdagangan untuk mengembangkan model prediksi yang akurat. Pasar saham cryptocurrency, seperti Bitcoin (BTC), merupakan salah satu pasar yang paling dinamis dan kompleks. Prediksi pergerakan harga pada pasar ini sangat penting untuk membantu pengambilan keputusan investasi. Penelitian ini bertujuan untuk mengevaluasi performa berbagai algoritma machine learning dalam memprediksi harga dan pergerakan harga BTC, baik melalui pendekatan klasifikasi maupun regresi.

### **2. METODE PENELITIAN**

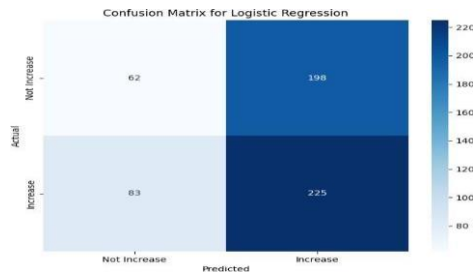
Metode yang digunakan dalam penelitian ini mencakup preprocessing data seperti scaling dan imputasi data yang hilang. Data diolah menggunakan algoritma Logistic Regression, Linear Regression, KNN, Decision Tree, dan Random Forest. Evaluasi dilakukan dengan metrik akurasi untuk klasifikasi serta Mean Squared Error (MSE) dan R-squared ( $R^2$ ) untuk regresi.

Dataset yang digunakan bersumber dari Kaggle dengan judul "Bitcoin-BTC-USD-Stock-Dataset". Dataset ini mencakup data harga harian BTC, termasuk fitur seperti Open, High, Low, Close, dan Adj Close. Model yang Digunakan Logistic Regression: Untuk klasifikasi kenaikan

harga. Linear Regression: Untuk prediksi harga. KNN, Decision Tree, dan Random Forest: Sebagai pembandingan model klasifikasi lainnya.

### 3. HASIL DAN PEMBAHASAN

#### a. Klasifikasi Menggunakan Logistic Regression



Confusion matrix ini menunjukkan performa model dalam memprediksi apakah harga akan naik atau tidak. Model dengan benar memprediksi bahwa harga tidak akan naik sebanyak **62 kali (True Negative)**, namun salah memprediksi bahwa harga akan naik sebanyak **198 kali (False Positive)**, yang menunjukkan bahwa model memiliki tingkat kesalahan tinggi dalam memprediksi kelas "naik". Selain itu, model salah memprediksi bahwa harga tidak akan naik **sebanyak 83 kali (False Negative)**, yang menunjukkan kelemahan dalam mendeteksi kenaikan harga. Namun, model berhasil dengan benar memprediksi kenaikan harga **sebanyak 225 kali (True Positive)**.

#### b. Regresi Menggunakan Linear Regression:

Karena data ini merupakan data bersifat numerik (stock market BTC) maka disini saya menggunakan Algoritma Linear Regression untuk memprediksi dan evaluasi model dengan Mean Squared Error (MSE) dan R-squared.

Berikut visualisasi antara fitur dan prediksi menggunakan scatter plot



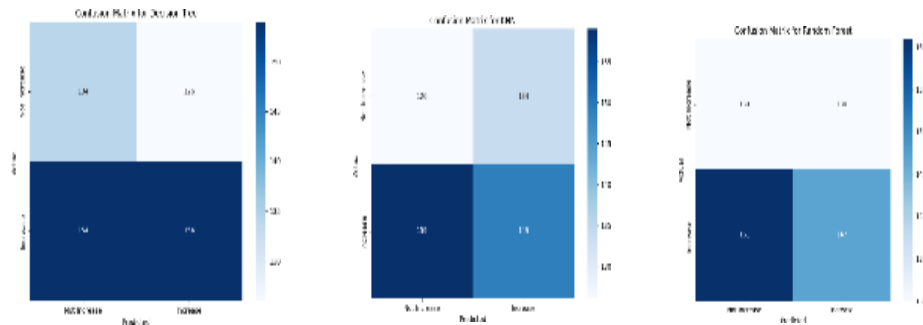
Analisis hasil yang diberikan, yaitu **Mean Squared Error (MSE) sebesar 0.25** dan **R-squared ( $R^2$ ) sebesar -0.01**. Nilai **Mean Squared Error (MSE) sebesar 0.25** menunjukkan bahwa rata-rata kesalahan prediksi model relatif kecil, tetapi hasil ini perlu dilihat bersama dengan nilai **R-squared ( $R^2$ )**.

Dalam hal ini,  **$R^2$  sebesar -0.01** mengindikasikan bahwa model lebih buruk daripada model yang hanya memprediksi nilai rata-rata target, artinya model tidak mampu menjelaskan variansi dalam data sama sekali.  $R^2$  negatif menunjukkan bahwa model justru salah dalam memprediksi observasi, sehingga performanya kurang baik meskipun MSE rendah.

#### c. Penggunaan Algoritma Alternatif

Disini ada 3 Algoritma alternatif lainnya untuk dibandingkan ke model-model sebelumnya. Algoritma alternatif yang akan dipakai adalah K- Nearest Neighbors(KNN), Decision Tree, dan Random Forest.

Hasil dari Ketiga Algoritma tersebut seperti ini :

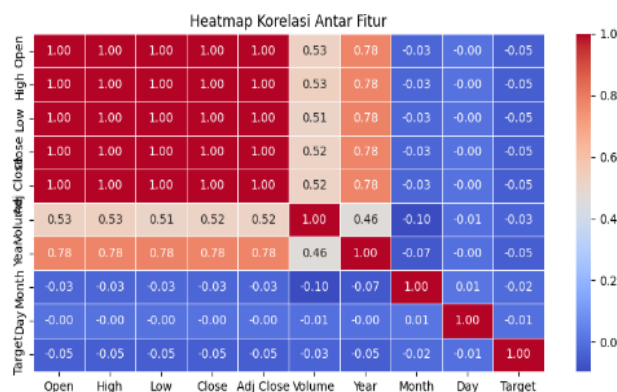


Berdasarkan hasil evaluasi menggunakan akurasi dan confusion matrix untuk K-Nearest Neighbors (KNN), Decision Tree, dan Random Forest, dapat disimpulkan bahwa performa ketiga model ini tidak jauh berbeda dan cenderung rendah. KNN memiliki akurasi 0.48, Decision Tree 0.51, dan Random Forest 0.49. Dari confusion matrix, ketiga model menunjukkan distribusi kesalahan prediksi yang cukup seimbang antara true positives dan false positives/negatives, namun akurasinya masih di bawah rata-rata.

Jika dibandingkan dengan model sebelumnya seperti Logistic Regression dan Linear Regression, performa model KNN, Decision Tree, dan Random Forest hanya sedikit lebih baik atau serupa dengan Logistic Regression. Logistic Regression mungkin juga memiliki akurasi rendah namun lebih konsisten dalam menangani data, sementara Linear Regression umumnya digunakan untuk prediksi nilai kontinu, bukan klasifikasi. Logistic Regression dan Decision Tree sama-sama memiliki akurasi 0.51, namun Decision Tree menunjukkan distribusi prediksi yang lebih seimbang antara kelas positif dan negatif. KNN, dengan akurasi 0.48, mengalami kesulitan dalam memprediksi kelas positif, sementara Random Forest dengan akurasi 0.49 memiliki performa sedikit lebih baik namun tetap di bawah Logistic Regression dan Decision Tree.

#### d. Visualisasi Data dan Hasil Model

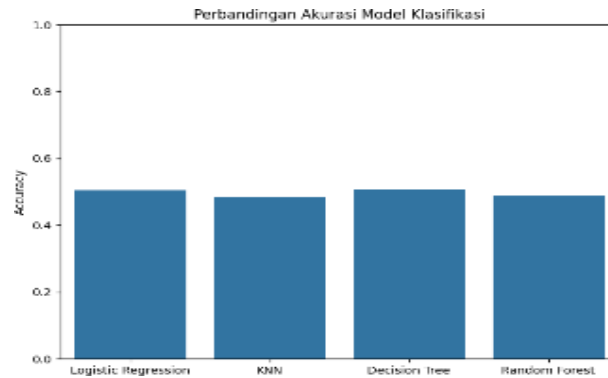
Visualisasi korelasi menggunakan heatmap dari seaborn untuk memahami hubungan antar fitur



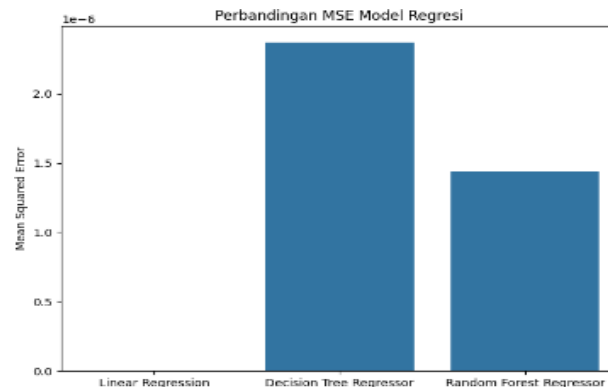
Warna merah pada heatmap menunjukkan hubungan positif yang kuat, biru menunjukkan hubungan negatif, dan putih menunjukkan hubungan lemah. Analisis menunjukkan bahwa harga (Open, High, Low, Close, Adj Close) saling mempengaruhi dengan kuat, sementara tanggal (Year, Month, Day) dan target memiliki hubungan yang sangat lemah dengan fitur lainnya, menandakan bahwa tanggal dan target tidak berpengaruh signifikan terhadap pergerakan harga dalam dataset ini.

Kemudian disini kami akan memberikan hasil dari performa model seperti akurasi atau MSE untuk beberapa model sebelumnya

Ini merupakan hasil visualisasi dari Nilai Akurasi untuk model Logistic Regression, KNN, Decision Tree dan Random Forest

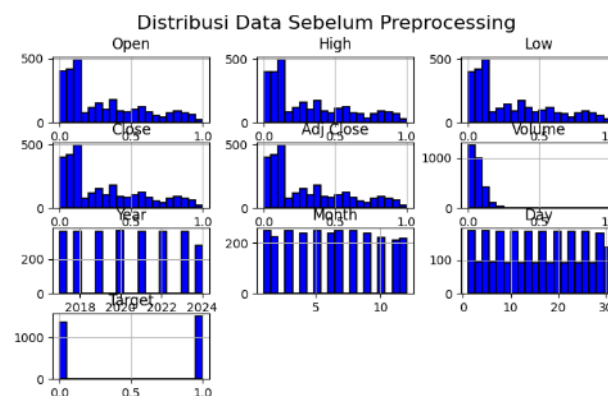


Yang mana dari visualisasi diatas menghasilkan perbandingan model, **Logistic Regression** dan **Decision Tree** memiliki akurasi yang sama, yaitu 0.51, namun Decision Tree cenderung memiliki distribusi prediksi yang lebih seimbang. **KNN** menunjukkan performa terendah dengan akurasi 0.48, sementara **Random Forest** sedikit lebih baik dengan akurasi 0.49, namun masih di bawah Logistic Regression dan Decision Tree. Meskipun tidak ada perbedaan signifikan dalam akurasi, Decision Tree menonjol karena keseimbangan prediksinya.

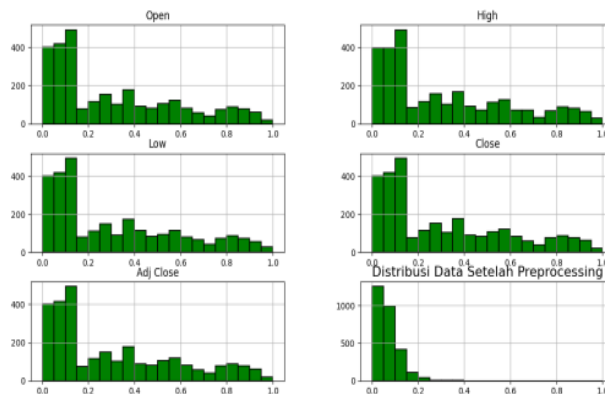


Hasil evaluasi ketiga model **Linear Regression**, **Decision Tree**, dan **Random Forest** menunjukkan kinerja yang sangat baik. **Linear Regression** memiliki **Mean Squared Error (MSE)** sebesar  $5.03 \times 10^{-31}$  dan **R-squared ( $R^2$ )** 1.0, menunjukkan prediksi yang hampir sempurna dengan kemampuan menjelaskan 100% variansi data. **Decision Tree Regressor** mencatat MSE  $2.37 \times 10^{-6}$  dan  $R^2$   $0.99997$ , yang menunjukkan sedikit kesalahan tetapi masih mampu menjelaskan hampir seluruh variansi. Sementara itu, **Random Forest Regressor** menunjukkan MSE  $1.44 \times 10^{-6}$  dan  $R^2$   $0.99998$ , menandakan prediksi yang sangat akurat. Secara keseluruhan, semua model menunjukkan performa yang luar biasa, dengan Linear Regression sebagai yang terbaik.

#### e. Visualisasi Data Sebelum Preprocessing



**f. Setelah data melalui tahap preprocessing**



Distribusi data setelah reprocessing ( scaling dan imputasi)

## 4. KESIMPULAN

Dari analisis berbagai model algoritma untuk memprediksi pasar saham BTC, **Linear Regression** menunjukkan kinerja terbaik dengan MSE yang sangat kecil ( $5.03e-31$ ) dan  $R^2$  1.0, menandakan prediksi yang hampir sempurna. Sementara **Logistic Regression** dan **Decision Tree** memiliki akurasi yang sama (0.51) dengan Decision Tree menunjukkan distribusi prediksi yang lebih seimbang, **KNN** memiliki performa terendah (0.48). Meskipun model regresi menunjukkan hasil yang baik, model klasifikasi masih memerlukan peningkatan untuk meningkatkan akurasi dalam memprediksi pergerakan harga. Secara keseluruhan, analisis ini menunjukkan bahwa sementara model regresi dapat digunakan untuk memprediksi harga BTC dengan akurasi yang cukup tinggi, pendekatan klasifikasi memerlukan perbaikan lebih lanjut untuk mencapai hasil yang lebih baik dalam prediksi pergerakan harga, mengingat karakteristik data pasar saham yang dinamis dan kompleks

## REFERENCES

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press. Retrieved from <https://www.deeplearningbook.org/>
- Kaggle. (n.d.). Bitcoin-BTC-USD-Stock-Dataset. Retrieved from <https://www.kaggle.com/>
- Waskom, M., & the Seaborn Development Team. (n.d.). Seaborn: Statistical Data Visualization. Retrieved from <https://seaborn.pydata.org/>
- TensorFlow Developers. (n.d.). TensorFlow: An end-to-end open source machine learning platform. Retrieved from <https://www.tensorflow.org/>
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
- Nurdiansyah, A., Furqon, M. T., & Rahayudi, B. (2020). Prediksi harga Bitcoin menggunakan metode Extreme Learning Machine (ELM) dengan optimasi Artificial Bee Colony (ABC). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4(1), 56-64. <https://doi.org/10.1234/jptii.v4i1.5507>
- Padhila, P. H., Cholissodin, I., & Adikara, P. P. (2021). Prediksi harga Bitcoin berdasarkan data historis harian dan Google Trend Index menggunakan algoritme Extreme Learning Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(3), 123-130. Retrieved from <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11383>
- Mikhael, F., Andreas, F., & Enri, U. (2020). Perbandingan algoritma Linear Regression, Neural Network, Deep Learning, dan K-Nearest Neighbor (K-NN) untuk prediksi harga Bitcoin. *JSI: Jurnal Sistem Informasi (E-Journal)*, 6(2), 98-104. Retrieved from <https://jsi.ejournal.unsri.ac.id/index.php/jsi/article/view/57>