

## KOMPARASI AKURASI REGRESI LINEAR BERGANDA DAN DECISION TREE PREDIKSI HARGA MOTOR BEKAS BERBASIS WEBSITE

Arya Putrana Kohan<sup>1</sup>, Adhitya Pratama<sup>2</sup>, Andika Satria Pratama<sup>3</sup>, Tri Ramadhanti<sup>4</sup>, Victor Parulian<sup>5</sup>,  
Sigit Wibawa<sup>6</sup>, Muhammad Muharrom<sup>7</sup>

<sup>1,2,3,4,5,6,7</sup> Universitas Bina Sarana Informatika, Bekasi, Indonesia

Responden: aryaputrana21@gmail.com

### ABSTRACT

Two-wheeled vehicles are now a primary necessity for Indonesians. Buying and selling motorcycles, both new and used, is now commonplace. However, many prospective buyers are still unaware of the price range for the type or model of motorcycle they wish to buy or sell. To help estimate these prices, a Machine Learning (ML) approach can be used with the Supervised Learning method, which trains models using statistical analysis algorithms. Prediction is an important aspect of statistics, especially in decision making. This study aims to compare the performance of two machine learning algorithms, namely Multiple Linear Regression and Decision Tree Regressor, in predicting the price of used motorcycles. The data used includes Model, Year, Transmission, Odometer, Type, Tax, Fuel Consumption, Engine, and Price. The dataset was divided into training and testing data with a ratio of 80:20. The results showed that the Multiple Linear Regression algorithm had higher accuracy with an  $R^2$  value of 0.8755 and a lower RMSE of 1328.09 compared to the Decision Tree, which had an  $R^2$  of 0.8422 and an RMSE of 1495.23. Therefore, the Multiple Linear Regression model is more effective in handling linear relationships between variables when predicting the price of used motorcycles, with the comparison results visualised in the form of a heatmap and the predictions displayed on a website.

Keywords: Machine Learning, Prediction, Linear Regression, Decision Tree, Used Motorcycles

### ABSTRAK

Kendaraan roda dua kini menjadi kebutuhan utama bagi masyarakat Indonesia. Aktivitas jual beli motor, baik baru maupun bekas, sudah sangat umum dilakukan. Namun, masih banyak calon pembeli yang belum mengetahui kisaran harga dari tipe atau model motor yang ingin mereka beli maupun jual. Untuk membantu memperkirakan harga tersebut, dapat digunakan pendekatan Machine Learning (ML) dengan metode Supervised Learning, yang berfungsi melatih model menggunakan algoritma analisis statistik. Prediksi merupakan aspek penting dalam statistik, terutama dalam pengambilan keputusan. Penelitian ini bertujuan membandingkan performa dua algoritma machine learning, yaitu Regresi Linear Berganda dan Decision Tree Regressor, dalam memprediksi harga motor bekas. Data yang digunakan meliputi Model, Tahun, Transmisi, Odometer, Jenis, Pajak, Konsumsi BBM, Mesin, dan Harga. Dataset dibagi menjadi data pelatihan dan pengujian dengan perbandingan 80:20. Hasil penelitian menunjukkan bahwa algoritma Regresi Linear Berganda memiliki akurasi lebih tinggi dengan nilai  $R^2$  sebesar 0.8755 dan RMSE lebih rendah sebesar 1328.09 dibandingkan Decision Tree yang memiliki  $R^2$  sebesar 0.8422 dan RMSE 1495.23. Oleh karena itu, model Regresi Linear Berganda lebih efektif dalam menangani hubungan linear antar variabel saat memprediksi harga motor bekas, dengan hasil perbandingan divisualisasikan dalam bentuk heatmap dan prediksinya ditampilkan melalui website.

**Kata Kunci:** Machine Learning, Prediction, Linear Regression, Decision Tree, Used Motorcycles

### Riwayat Artikel :

Tanggal diterima : 31-10-2025

Tanggal revisi : 10-11-2025

Tanggal terbit : 01-12-2025

### DOI :

<https://doi.org/10.31949/infotech.v11i2.16430>

INFOTECH journal by Informatika UNMA is licensed under CC BY-SA 4.0

Copyright © 2025 By Author



1. PENDAHULUAN

Pada keadaan saat ini, kendaraan bermotor sudah menjadi kebutuhan pokok bagi masyarakat terutama kendaraan roda dua atau sepeda motor. Menurut data dari Korlantas Polri kendaraan bermotor mencapai 168.275.423 unit pada bulan April 2025 sumber data (CNA Indonesia, 2025). Namun seiring perkembangan zaman banyak produsen kendaraan roda dua menaikkan kualitas produk mereka dengan memperbarui teknologi pada produk mereka, semakin unggul kualitas produk mereka maka harga dari produk mereka juga mengalami peningkatan. Di sisi lain, masih banyak masyarakat yang mempertahankan atau memperjualbelikan motor keluaran tahun-tahun sebelumnya. Namun, sebagian besar dari mereka belum memiliki acuan yang tepat untuk memprediksi harga jual motor bekasnya. Dalam masalah ini Penggunaan metode Machine Learning memberikan peluang untuk melakukan proses pembelajaran data secara lebih mendalam dan terstruktur, sehingga mampu meningkatkan ketepatan dalam memprediksi harga motor bekas (Leriansyah Putra Nasyuli et al., 2023)

Machine Learning (ML) adalah sebuah metode yang memungkinkan komputer atau mesin untuk belajar secara otomatis dari pengalaman tanpa harus diprogram secara eksplisit. Dalam prosesnya, data digunakan sebagai bahan pembelajaran, sehingga dikenal dengan istilah belajar dari data (learn from data) (Pamulang, n.d.). Pembelajaran pada machine learning dapat dilakukan menggunakan beragam jenis algoritma sesuai dengan tujuan analisisnya, ada 3 jenis yaitu Supervised Learning, Unsupervised Learning, dan Reinforcement Learning. Supervised Learning merupakan metode belajar yang diberikan dataset berupa informasi input, output yang diinginkan, sehingga sistem dapat mempelajari berdasarkan data yang sudah ada. Sistem ini akan mencari pola dari dataset, lalu pola tersebut akan dijadikan acuan untuk kumpulan data berikutnya (Artikel et al., 2021). Unsupervised Learning ini bersifat deskriptif dan akan berguna untuk mengelompokkan sebuah data, algoritma ini tidak mendapatkan training dataset, dikarenakan algoritma ini bukan bersifat prediktif, sehingga membutuhkan sebuah pembelajaran dari data yang sudah ada (Abijono et al., 2021). Reinforcement Learning adalah sebuah algoritma yang mengakumulasikan informasi dalam bentuk sinyal untuk memilih langkahlangkah yang menuju hasil yang di inginkan dengan tingkat maksimal (Nurhalizah et al., 2024). Prediksi adalah hasil dari kegiatan memperkirakan atau memprediksikan nilai pada masa yang akan datang dengan menggunakan data pada masa lalu, prediksi ini menunjukkan apa yang akan terjadi pada suatu keadaan dan merupakan input untuk proses perencanaan dan pengambilan sebuah keputusan (Hasibuan et al., 2022).

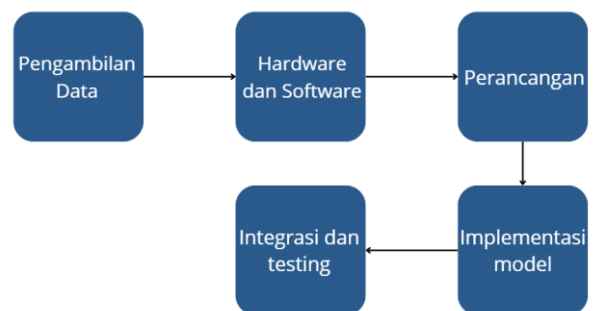
Model regresi adalah sebuah teknik statistik yang digunakan untuk membuat prediksi dengan cara memodelkan hubungan fungsional dan matematis antara dua atau lebih variabel. Penelitian ini

bertujuan untuk memprediksi volume penjualan properti PT XYZ menggunakan data historis perusahaan. Penggunaan model regresi linear dalam memproyeksikan penjualan properti pada penelitian ini menunjukkan hasil yang sangat akurat. Tingkat keakurasian ini didukung oleh hasil pengujian yang memenuhi standar melalui nilai Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Mean Absolute Percentage Error (MAPE) yang rendah untuk seluruh tipe properti yang dianalisis (Ayuni & Fitriyah, 2020). Model decision tree adalah model data processing yang memprediksi hasil di masa depan melalui pembangunan model klasifikasi atau regresi yang disajikan dalam bentuk struktur pohon. Secara fundamental, Decision Tree berfungsi sebagai model visual yang memungkinkan pemahaman yang rasional, bertahap, dan terstruktur mengenai proses pengambilan keputusan, di mana setiap langkahnya dapat divisualisasikan secara jelas dalam diagram pohon keputusan (Nurani et al., 2023). Namun demikian, masih relatif sedikit penelitian tentang perbandingan model regresi linear dengan decision tree, dan kami melakukan penelitian ini untuk membandingkan hasil dari Koefisien Determinasi ( $R^2$ ) dan Root Mean Squared Error (RMSE) dari kedua model tersebut.

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma Regresi Linear Berganda dan Decision Tree Regression dalam memprediksi harga motor bekas berdasarkan kriteria yang diinginkan. Data yang digunakan merupakan data historis harga motor yang diolah dan disajikan melalui sebuah layanan berbasis website (Hasibuan et al., 2022). Evaluasi terhadap performa model dilakukan menggunakan metrik Root Mean Squared Error (RMSE) dan R-Squared ( $R^2$ ) guna mengetahui tingkat akurasi dan efisiensi dari masing-masing algoritma. Melalui penelitian ini, diharapkan dapat diperoleh model prediksi harga motor bekas yang lebih akurat dan dapat membantu pengguna dalam memperkirakan harga secara efektif (Hallan & Fajri, 2025).

2. METODE

2.1. Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen komparatif. Tujuannya adalah membangun model prediksi harga motor bekas dan membandingkan performa dua algoritma

Supervised Learning utama dalam bidang Machine Learning, yaitu Regresi Linear Berganda dan Decision Tree Regressor. Perbandingan ini bertujuan mengidentifikasi model mana yang memberikan tingkat akurasi prediksi terbaik. Berikut adalah rencana yang digunakan untuk semua langkah dan keperluan penelitian ini (Afrilia S et al., 2024).

a. **Pengambilan Data**

Pada tahap ini, data yang digunakan berasal dari sebuah situs yang bernama Kaggle.com. Kaggle adalah salah satu situs yang terkenal di dunia Data Science dan Machine Learning, dan data ini sangat membantu ilmuwan di seluruh dunia (Meliala et al., 2023). Data tersebut mencakup beberapa informasi mengenai harga, dan spesifikasi motor bekas. Dataset awal berjumlah 199 data, setelah melalui proses pembersihan dan seleksi berdasarkan kelengkapan atribut seperti Model, Tahun, Transmisi, Odometer, Jenis, Pajak, Konsumsi BBM, Mesin, dan Harga, diperoleh 199 data yang digunakan dalam penelitian. Pada gambar ke-2 adalah struktur data yang digunakan.

```
Data columns (total 9 columns):
# Column Non-Null Count Dtype
---
0 model 199 non-null object
1 tahun 199 non-null int64
2 harga 199 non-null int64
3 transmisi 199 non-null object
4 odometer 199 non-null int64
5 jenis 199 non-null object
6 pajak 199 non-null int64
7 konsumsiBBM 199 non-null float64
8 mesin 199 non-null int64
dtypes: float64(1), int64(5), object(3)
memory usage: 14.1+ KB
```

Gambar 2. Pengenalan struktur data yang digunakan

b. **Hardware dan Software**

Pada tahap ini dilakukan untuk mengidentifikasi kebutuhan perangkat keras dan perangkat lunak yang digunakan dalam penelitian (Meliala et al., 2023). Pengembangan ini menggunakan spesifikasi hardware (Laptop) sebagai berikut:

- i. CPU: Amd Ryzen 3 3250U
- ii. RAM: 16 Gb DDR 4
- iii. Penyimpanan: 256 Gb
- iv. GPU: AMD Radeon Graphics
- v. Sistem Operasi: windows 11 64-bit

Perangkat lunak yang digunakan dalam pengembangan dan analisis data pada penelitian ini sebagai berikut:

- i. Bahasa Pemrograman: Python
- ii. Library Machine Learning: Scikitlearn, Numpy, Matplotlib, Seaborn, dan Pandas.
- iii. Framework: Laravel 10
- iv. Bahasa Pemrograman Website: PHP

c. **Perancangan**

Pada tahap perancangan model terdapat pengenalan data dan preprocessing data. Pada pengenalan data menampilkan isi dari dataset. Pada tahap preprocessing data untuk mempersiapkan data agar dapat digunakan dalam pembentukan model yaitu terdiri dari pemilihan variabel berdasarkan tingkat korelasi, dan pemisahan data menjadi data pelatihan dan data pengujian.

a) Model Regresi Lienar Berganda

Secara matematis, persamaan regresi linear berganda dapat dituliskan sebagai berikut:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_nX_n + \epsilon$$

Keterangan:

- i. Y= variabel dependen atau nilai yang diprediksi
- ii.  $\alpha / \beta_0$ = konstanta (intercept), yaitu nilai Y ketika seluruh X bernilai nol
- iii.  $\beta_1, \beta_2, \dots, \beta_n$ = koefisien regresi yang menunjukkan seberapa besar pengaruh masing-masing variabel X terhadap Y
- iv.  $X_1, X_2, \dots, X_n$ = variabel independen (prediktor)
- v.  $\epsilon$ = komponen error atau kesalahan prediksi

Tabel 1. Contoh kode python pada regresi linear berganda

Kode	Penjelasan
#Split data X_train,X_test,y_train, y_test = train_test_split( X, y, test_size=0.2,rand om_state=42)	Ini digunakan untuk membagi dataset menjadi data training dan data testing, test_size=0.2 adalah bernilai 20% dan memiliki data training sebesar 80%.
#Latih model model=LinearRegressi on() model.fit(X_train, y_train)	Untuk memanggil model regresi linear, dan untuk melakukan proses pelatihan (training).
#Evaluasi Model y_pred = model.predict(X_test) r2 = r2_score (y_test, y_pred) rmse = np.sqrt(mean_squared_ error(y_test, y_pred))	Untuk menghasilkan nilai prediksi dari model yang diuji, dan untuk mengukur seberapa baiknya nilai yang dihasilkan.

b) Model Decision Tree

Selain regresi linear berganda, penelitian ini juga menggunakan model Decision Tree Regression sebagai pembanding untuk memprediksi harga motor bekas. Model ini bekerja dengan cara membangun struktur pohon keputusan yang secara bertahap membagi data menjadi beberapa kelompok. Proses pembentukan pohon dilakukan dengan memilih titik pemisahan (split) yang menghasilkan nilai kesalahan terkecil.

Tabel 2. Contoh kode python pada decision tree

Kode	Penjelasan
# Pisahkan fitur dan target X = df[['model','tahun','transmisi','odometer','jenis','pajak','konsumsiBBM','mesin']] y = df['harga']	X (fitur / variabel independen) → faktor-faktor yang digunakan untuk memprediksi.
# Split data X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)	Ini digunakan untuk membagi dataset menjadi data training dan data testing, test_size=0.2 adalah bernilai 20% dan memiliki data training sebesar 80%.
# Buat dan latih model model = DecisionTreeRegressor model.fit(X_train, y_train)	Untuk memanggil model regresi linear, dan untuk melakukan proses pelatihan (training).
# Evaluasi y_pred = model.predict(X_test) r2 = r2_score(y_test, y_pred) rmse = np.sqrt(mean_squared_error(y_test, y_pred))	Untuk menghasilkan nilai prediksi dari model yang diuji, dan untuk mengukur seberapa baiknya nilai yang dihasilkan.

Setelah model terbentuk akan dievaluasi menggunakan Root Mean Squared Error (RMSE) dan Koefisien Determinasi ( $R^2$ ) sebagai ukuran performa. Nilai RMSE menunjukkan tingkat kesalahan prediksi model dalam satuan yang sama dengan data harga motor, sedangkan  $R^2$  menunjukkan seberapa besar variasi data yang dapat dijelaskan oleh model.

Formula  $R^2$  dalam sistematis dinyatakan sebagai berikut:

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}$$

Keterangan:

- i.  $R^2$ = Koefisien determinasi
- ii.  $Y_i$ = Nilai aktual ke-i
- iii.  $\hat{Y}_i$ = Nilai prediksi ke-i
- iv.  $\bar{Y}$ = Nilai rata-rata dari data actual
- v.  $n$ = Jumlah total data

Formula RMSE dalam sistematis dinyatakan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

Keterangan:

- i. RMSE = Root Mean Squared Error
- ii.  $Y_i$ = Nilai aktual ke-i
- iii.  $\hat{Y}_i$ = Nilai prediksi ke-i
- iv.  $n$ = Jumlah total data

**d. Implementasi Model**

Setelah data dipisahkan, langkah berikutnya adalah melatih model menggunakan algoritma Regresi Linear Berganda dan Decision Tree Regression. Dataset dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian dengan perbandingan 80:20. Data pelatihan digunakan untuk membangun model agar dapat mengenali pola hubungan antara variabel bebas (fitur) dan variabel terikat (harga motor bekas), sedangkan data pengujian berfungsi untuk menguji kemampuan model dalam memprediksi data baru yang belum pernah dipelajari sebelumnya (Jl et al., 2023). Proses pembuatan model pada penelitian ini terdiri dari tiga tahap:

- i. Analisis Data Historis – Tahap ini bertujuan untuk memahami karakteristik dari data harga motor bekas, termasuk distribusi data, hubungan antar variabel, serta mendeteksi adanya data pencilan (outlier).
- ii. Pemilihan dan Penerapan Algoritma – Dua algoritma machine learning, yaitu Regresi Linear Berganda dan Decision Tree Regression, diterapkan untuk membangun model prediktif. Pemilihan algoritma didasarkan pada kemampuan masing-masing untuk menangani data dengan pola linear maupun non-linear.
- iii. Transformasi dan Evaluasi Model – Data historis diproses dan dimasukkan ke dalam model untuk menghasilkan prediksi harga motor bekas. Setelah model dibuat, evaluasi dilakukan dengan menggunakan metrik Root Mean Squared Error (RMSE) dan Koefisien Determinasi ( $R^2$ ) guna mengukur akurasi dan kinerja model.

Dengan mengikuti tahapan-tahapan tersebut, model yang dihasilkan diharapkan mampu memberikan hasil prediksi harga motor bekas secara akurat dan dapat diintegrasikan ke dalam sistem berbasis web.

**e. Integrasi dan Testing**

Proses integrasi dan pengujian sistem bertujuan untuk mengetahui apakah perangkat lunak sudah sesuai dengan desain, dan fungsionalitas dari aplikasi apakah berjalan dengan baik atau tidak, dengan adanya tahapan ini maka dapat mencegah terjadinya sebuah kesalahan, bug atau error pada program yang telah dibuat (Sari et al., 2024).

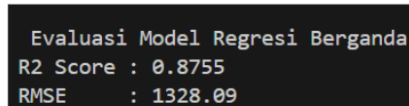
Tahapan-tahapan tersebut dilakukan agar model regresi linear dan decision tree yang dihasilkan mampu memberikan hasil prediksi harga motor bekas secara akurat dan dapat diintegrasikan dalam sistem berbasis web.

**3. HASIL DAN PEMBAHASAN**

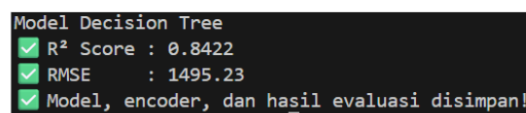
Penelitian ini menggunakan dataset yang berisi 199 data motor bekas dari berbagai merek dengan atribut seperti Model, Tahun, Transmisi, Odometer, Jenis, Pajak, Konsumsi BBM, Mesin, dan Harga. Setelah melalui proses pembersihan data, dilakukan pembagian data menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20%.

Proses latihan ini dilakukan menggunakan dua algoritma yaitu Regresi Linear Berganda dan Decision Tree Regeression, kedua model ini diuji coba dengan menggunakan metrik Koefisien Determinasi ( $R^2$ ) dan Root Mean Squared Error (RMSE) untuk dapat menilai tingkat akurasi dan kesalahan dari prediksi. Berikut gambar 3 adalah hasil dari perhitungan  $R^2$  dan RMSE model Regresi Lienar berganda dan Decision Tree.

Model Regresi Lienar



Model Decision Tree



**Gambar 3. Hasil perhitungan  $R^2$  dan RMSE**

**Tabel 3. Menampilkan hasil evaluasi dari kedua model**

Model	Koefisien Determinasi ( $R^2$ )	Root Mean Squared Error (RMSE)
Regresi Linear Berganda	0.8755	1328.09
Decision Tree	0.8422	1495.23

Dari table 3 di atas, terlihat bahwa model Regresi Linear Berganda memiliki nilai  $R^2$  yang lebih tinggi dibandingkan dengan model Decision Tree Regression. Artinya, model Regresi Linear Berganda mampu menjelaskan sekitar 87,55% variasi data harga motor bekas, sedangkan model Decision Tree hanya mampu menjelaskan sekitar

84,22% variasi data. Sementara itu, pada hasil evaluasi RMSE, regresi linear berganda memperoleh nilai RMSE sebesar 1328.09, sedangkan Decision Tree sebesar 1495.23. Nilai RMSE yang lebih kecil terdapat pada regresi linear berganda dan ini menunjukkan bahwa selisih rata-rata antara nilai prediksi dan nilai aktual harga motor bekas lebih rendah, sehingga dapat dikatakan kesalahan prediksi pada model ini lebih kecil jika dibandingkan dengan Decesion Tree. Secara keseluruhan berdasarkan dua metrik dapat disimpulkan bahwa model Regresi Linear Berganda memberikan hasil prediksi yang lebih akurat dan stabil dibandingkan dengan model Decision Tree Regression untuk dataset harga motor bekas yang digunakan dalam penelitian ini.

**3.1. Hasil Analisis Regresi Linear Berganda**

Model Regresi Linear Berganda bekerja dengan mencari garis terbaik yang merepresentasikan hubungan linier antara variabel independen (misalnya tahun, odometer, pajak, mesin, konsumsiBBM) dengan variabel dependen (harga). Pada hasil  $R^2$  yang sebesar 0.8755 ini menunjukkan bahwa sebagian besar variasi harga motor dapat dijelaskan oleh variabel-variabel tersebut. Pada nilai RMSE sebesar 1328.09 menunjukkan rata-rata selisih prediksi harga terhadap nilai sebenarnya relatif kecil, yang artinya model ini cukup akurat. Beberapa variabel yang kemungkinan besar memiliki pengaruh paling kuat terhadap harga yaitu:

- a. Tahun - semakin baru tahun produksi, maka harga akan semakin tinggi.
- b. Odometer - semakin tinggi jarak tempuh yang dilewati, maka harga akan cenderung menurun.
- c. Kapasitas mesin (CC) - Motor dengan kapasitas mesin yang lebih besar memiliki harga yang cenderung lebih tinggi.
- d. Jenis dan transmisi - dapat mempengaruhi harga karena segmentasi pasar.

**3.2. Hasil Analisis Decision Tree Regression**

Model decision tree pada penelitian ini menghasilkan nilai  $R^2$  sebesar 0.8422 dan nilai RMSE sebesar 1495.23, ini hasil yang sedikit rendah dibandingkan dengan model regresi linear berganda, hal ini menjadi tanda bahwa Decision tree ini tidak terlalu efektif untuk dataset ini. Ada beberapa kemungkinan yang menjadi penyebabnya yaitu:

- a. Jumlah data yang sangat terbatas (199) membuat struktur pohon keputusan cenderung overfitting terhadap data latih. Overvitting adalah masalah yang dimana model belajar data pelatihan secara lebih detail, seperti noise dan data anomali, sehingga tidak dapat melakukan prediksi yang akurat pada data baru yang tidak dikenal.
- b. Variasi antar atribut relatif kecil, sehingga data (split) pada setiap cabang pohon tidak

menghasilkan peningkatan akurasi yang signifikan.

- c. Pola hubungan antar variabel dalam data ini tidak terlalu kompleks, sehingga pendekatan linear lebih efisien dan akurat.

Meski begitu, model ini tetap memiliki keunggulan dalam interpretasi hierarkis dan kemampuan menangani data non-linear. Namun pada penelitian ini struktur hubungan antara variabel tidak terlalu cukup kompleks untuk memerlukan data non-linear.

### 3.3. Perbandingan Kinerja Kedua Model

Berdasarkan evaluasi pada tabel 3 diatas, dapat disimpulkan bahwa Regresi Linear Berganda memiliki performa lebih baik dibandingkan dengan Decision Tree Regression pada dataset yang digunakan ini. Perbandingan detailnya adalah sebagai berikut:

- a. Akurasi model ( $R^2$ ): Regresi linear berganda lebih tinggi dengan hasil ( $0.8755 > 0.8422$ ).
- b. Tingkat Kesalahan (RMSE): Regresi linear berganda lebih rendah dengan memiliki hasil ( $1328.09 < 1495.23$ ).
- c. Kemampuan Non-Linear: Decision tree lebih unggul dalam mendeteksi pola non-linear, tetapi pada hasilnya belum optimal pada dataset yang kecil.
- d. Kestabilan model: Regresi linear lebih stabil terhadap variasi data kecil, sedangkan Decision tree bisa mengalami overfitting jika tidak diatur kedalaman pohonnya.

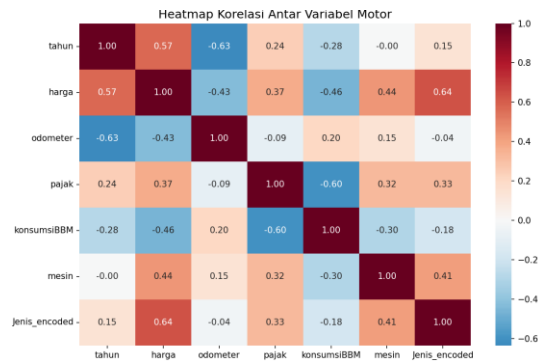
### 3.4. Visualisasi Korelasi Antar Variabel

Penelitian ini menggunakan visualisasi heatmap korelasi antar variabel dari dataset harga motor bekas. Heatmap digunakan untuk melihat sejauh mana hubungan antar variabel input (fitur) terhadap variabel target yaitu harga motor. Nilai korelasi berkisar antara -1 hingga +1, di mana:

- a. Nilai positif (+) menunjukkan hubungan searah antara dua variabel (jika satu naik, yang lain ikut naik).
- b. Nilai negatif (-) menunjukkan hubungan berlawanan arah (jika satu naik, yang lain turun).
- c. Nilai mendekati 0 menunjukkan tidak adanya hubungan linear yang kuat.

### 3.5. Analisis Heatmap Korelasi Pada Model Regresi Linear Berganda

Pada heatmap untuk model Regresi linear berganda hubungan antar variabel

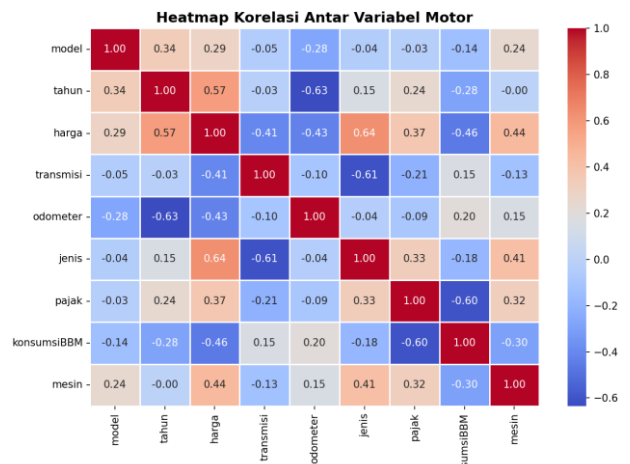


Gambar 4. Heatmap Korelasi antar variabel Regresi Linear Berganda

- a. Korelasi antara tahun dan harga (0.57) tetap menjadi indikator paling dominan dalam menentukan harga motor.
- b. Odometer (-0.43) juga tetap menjadi faktor penurunan harga.
- c. Variabel mesin (0.44) dan jenis\_encoded (0.64) memperlihatkan hubungan positif sedang hingga kuat, menunjukkan semakin besar kapasitas mesin dan jenis motor tertentu memiliki harga lebih tinggi.
- d. Korelasi antar variabel input seperti tahun - odometer (-0.63) juga terlihat cukup tinggi secara negatif, menunjukkan adanya potensi multikolinearitas, yang perlu diperhatikan dalam model regresi linear.

### 3.6. Analisis Heatmap Korelasi Pada Model Decision Tree Regression

Pada heatmap untuk model Decision Tree Regression, korelasi antar variabel menunjukkan bahwa:



Gambar 5. Heatmap Korelasi antar variabel Decision tree

- a. Tahun memiliki korelasi positif terhadap harga sebesar 0.57, menunjukkan bahwa semakin baru tahun motor maka harganya cenderung lebih tinggi.
- b. Odometer memiliki korelasi negatif cukup kuat sebesar -0.43, artinya semakin besar jarak tempuh motor, maka harga jualnya menurun.
- c. Jenis motor (fitur jenis) menunjukkan korelasi positif sebesar 0.64 terhadap harga.

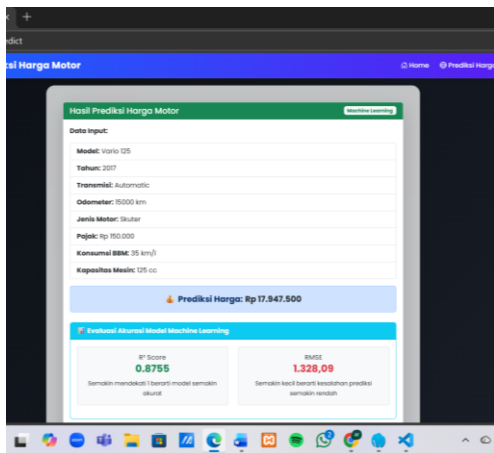
Ini menandakan tipe motor (sport, matic, bebek) berpengaruh terhadap harga jual.

- d. Konsumsi BBM dan pajak memiliki korelasi yang lebih rendah terhadap harga (sekitar 0.3–0.4), sehingga pengaruhnya tidak sebesar tahun dan jenis.

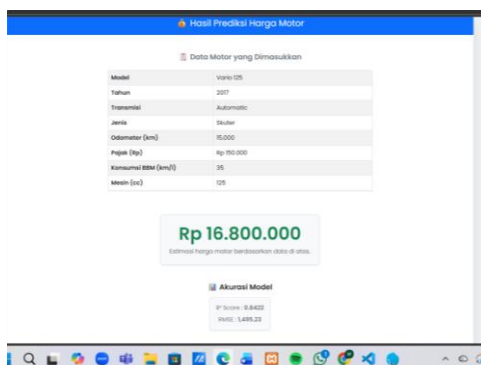
Karena Decision Tree tidak terlalu sensitif terhadap hubungan linear antar variabel, model ini dapat menangkap pola non-linear dan interaksi antar fitur yang tidak selalu tercermin dalam korelasi.

Dapat dikatakan bahwa analisis korelasi yang mempengaruhi positif terhadap harga motor bekas ada tiga yaitu: tahun, jenis, mesin. Dapat dipastikan juga bahwa odometer memiliki pengaruh negatif terhadap harga penjualan motor bekas. Walaupun kedua model menggunakan dataset yang sama, Regresi Linear Berganda lebih bergantung pada korelasi linear, sedangkan Decision Tree mampu menangkap hubungan non-linear yang tidak selalu tampak dalam analisis korelasi ini.

3.7. Visualisasi Website



Gambar 6. Tampilan website dari Regresi Linear Berganda



Gambar 7. Tampilan website dari Decision Tree Regression

Pada gambar nomor 11 dan gambar nomor 12 menampilkan halaman hasil dari prediksi yang dapat digunakan oleh user. User atau pengguna dapat memasukan 8 data seperti (Model motor, tahun motor, jenis transmisi, jenis motor, odometer, harga pajak, konsumsi bbm, dan mesin). Dari hasil di atas,

sistem akan menampilkan estimasi harga motor bekas berdasarkan data yang dimasukkan serta menampilkan nilai akurasi model yang digunakan (Panggabean et al., 2020).

4. KESIMPULAN

Berdasarkan hasil penelitian pada gambar 9 dan tabel 1 yang dilakukan mengenai perbandingan algoritma Regresi Linear Berganda dan Decision Tree Regression dalam memprediksi harga motor bekas berbasis website, dapat disimpulkan bahwa Regresi Linear Berganda memiliki kinerja yang lebih unggul. Model ini memperoleh nilai R<sup>2</sup> sebesar 0,8755 dan RMSE sebesar 1.328,09, yang menunjukkan tingkat akurasi tinggi serta kesalahan prediksi yang lebih rendah dibandingkan dengan Decision Tree Regression, yang memiliki R<sup>2</sup> sebesar 0,8422 dan RMSE sebesar 1.495.23. Hasil tersebut mengindikasikan bahwa Regresi Linear Berganda lebih mampu menggambarkan hubungan antar variabel secara stabil dan efisien, sedangkan Decision Tree Regression meskipun memberikan hasil yang cukup baik, cenderung lebih peka terhadap perubahan data. Penerapan kedua model dalam sistem berbasis website juga berhasil menampilkan hasil prediksi secara informatif, cepat, dan mudah dipahami oleh pengguna. Secara keseluruhan, dapat disimpulkan bahwa Regresi Linear Berganda merupakan model yang paling efektif dan optimal untuk digunakan dalam sistem prediksi harga motor bekas berbasis website pada penelitian ini.

PUSTAKA

Abijono, H., Santoso, P. & Anggreini, N. L. (2021). Algoritma Supervised Learning Dan Unsupervised Learning Dalam Pengolahan Data. *Jurnal Teknologi Terapan: G-Tech*, 4(2), 315–318. <https://doi.org/10.33379/gtech.v4i2.635>

Afrilia S, N., Frazna Az-Zahra, F. & Prajoko, P. (2024). Prediksi Hasil Panen Wortel Menggunakan Algoritma Regresi Linear Berganda. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(5), 10255–10262. <https://doi.org/10.36040/jati.v8i5.10954>

Artikel, R., Luh, N., Chandra, P., Rahman, R. A., Venyutzky, R. & Rakhmawati, N. A. (2021). Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning. 7(April), 47–58.

Ayuni, G. N. & Fitriannah, D. (2020). Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ. *Jurnal Telematika*, 14(2), 79–86. <https://doi.org/10.61769/telematika.v14i2.321>

CNA Indonesia. (2025). Kendaraan bermotor

- Indonesia tembus 168 jutaan, paling banyak di mana? In *CNA Indonesia (cna.id)*. [https://www.cna.id/indonesia/kendaraan-bermotor-indonesia-tembus-168-jutaan-paling-banyak-di-mana-31031?utm\\_source=chatgpt.com](https://www.cna.id/indonesia/kendaraan-bermotor-indonesia-tembus-168-jutaan-paling-banyak-di-mana-31031?utm_source=chatgpt.com)
- Hallan, R. R. & Fajri, I. N. (2025). Prediksi Harga Rumah menggunakan Machine Learning Algoritma Regresi Linier. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 7(1), 57–62. <https://doi.org/10.47233/jteksis.v7i1.1732>
- Hasibuan, E., Informasi, S., Ilmu, F., Informasi, T., Gunadarma, U., Margonda, J., No, R., Cina, P. & Jawa, D. (2022). Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web. *Jurnal Ilmiah Komputasi*, 21(4), 595–602. <https://doi.org/10.32409/jikstik.21.4.3327>
- Jl, A., Iskandar, W., Baru, K., Percut, K., Tuan, S., Serdang, K. D. & Utara, S. (2023). *Analisis Prediksi Harga Rumah di Bandung Menggunakan Regresi Linear Berganda Rafif Nauval Tuah Siregar Vijay Sitorus Universitas Negeri Medan Willy Pramudia Ananta perbandingan melalui penalaran berbasis kasus , " yang dilakukan oleh I-Cheng Yeh , Tzu- yang berkaitan . dengan faktor-faktor yang mempengaruhi prediksi harga rumah , dengan. 1(6).*
- Leriansyah Putra Nasyuli, Imran Lubis & Andi Marwan Elhanafi. (2023). Penerapan Model Machine Learning Algoritma Gradient Boosting dan Linear Regression Melakukan Prediksi Harga Kendaraan Bekas. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 2(2), 299–310. <https://doi.org/10.70340/jirsi.v2i2.56>
- Meliala, S., Elektro, T., Teknik, F. & Malikussaleh, U. (2023). *PENGIRAAN POSE MODEL MANUSIA PADA REPETISI KEBUGARAN AI*. 9(1), 11–19.
- Nurani, A. T., Setiawan, A. & Susanto, B. (2023). Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma. *Jurnal Sains Dan Edukasi Sains*, 6(1), 34–43. <https://doi.org/10.24246/juses.v6i1p34-43>
- Nurhalizah, R. S., Ardianto, R. & Purwono, P. (2024). Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review. *Jurnal Ilmu Komputer Dan Informatika*, 4(1), 61–72. <https://doi.org/10.54082/jiki.168>
- Pamulang, U. (n.d.). *KOMMAS : Jurnal Pengabdian Kepada Masyarakat PENGENALAN SEJAK DINI SISWA SMP TENTANG MACHINE LEARNING UNTUK KLASIFIKASI GAMBAR DALAM MENGHADAPI KOMMAS : Jurnal*
- Pengabdian Kepada Masyarakat. 7–15.*
- Panggabean, D. S. O., Buulolo, E. & Silalahi, N. (2020). Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 56. <https://doi.org/10.30865/jurikom.v7i1.1947>
- Sari, N. N., Anisah, T. T. & Fitriani, R. (2024). Implementasi Machine Learning untuk Prediksi Harga Laptop Menggunakan Algoritma Regresi Linear Berganda Machine Learning Implementation for Laptop Price Prediction Using Multiple Linear Regression Algorithm. *Jurnal Manajemen Informatika (JAMIKA)*, 14(2), 162–177. <https://doi.org/10.34010/jamika.v14i2.12923>