

Penerapan Algoritma Linear Regression dalam Prediksi Harga Mobil Bekas Berbasis Machine Learning

Pepti Kumala Dewi¹, Yoviapridiansyah²

^{1,2} Teknik Informatika, Universitas Muhammadiyah Bengkulu

Email: ¹peptikumaladewi2802@gmail.com, ²yoviapridiansyah@umb.ac.id

Email Penulis Korespondensi: yoviapridiansyah@umb.ac.id

Abstrak

Penentuan harga mobil bekas sering menjadi permasalahan dalam proses jual beli kendaraan karena masih banyak dilakukan secara subjektif dan belum sepenuhnya memanfaatkan data. Penelitian ini bertujuan untuk membangun model prediksi harga mobil bekas menggunakan algoritma Linear Regression berbasis machine learning agar menghasilkan estimasi yang lebih objektif dan terukur. Data yang digunakan berasal dari platform Kaggle dengan total 8.128 data kendaraan bekas yang mencakup atribut tahun kendaraan, jarak tempuh, kapasitas mesin, daya maksimum, dan jumlah tempat duduk. Tahapan penelitian meliputi pengumpulan data, preprocessing, serta pembagian data menjadi data latih dan data uji dengan rasio 80:20. Model dibangun menggunakan algoritma Linear Regression dan dievaluasi menggunakan metrik Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), serta koefisien determinasi (R^2). Hasil penelitian menunjukkan bahwa model memiliki performa yang baik dengan nilai MAE sebesar 96.477, RMSE sebesar 215.746, dan R^2 sebesar 0,9289, yang menunjukkan bahwa model mampu menjelaskan 92,89% variasi harga mobil bekas. Dengan demikian, model yang dihasilkan dapat digunakan sebagai sistem pendukung keputusan dalam menentukan harga mobil bekas secara lebih akurat.

Kata Kunci: Prediksi harga mobil bekas, *Machine Learning*, *Linear Regression*, *Google Colab*, Evaluasi model.

Abstract

Determining the price of used cars is often a problem in the vehicle buying and selling process, as it is still largely based on subjective judgment and has not fully utilized data. This study aims to develop a used car price prediction model using the Linear Regression algorithm based on machine learning to produce more objective and measurable estimates. The data used were obtained from the Kaggle platform, consisting of 8,128 used car records with attributes including vehicle year, mileage, engine capacity, maximum power, and seating capacity. The research stages include data collection, preprocessing, and data splitting into training and testing sets with a ratio of 80:20. The model was built using the Linear Regression algorithm and evaluated using several metrics, namely Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The results show that the model has good performance, with an MAE of 96,477, an RMSE of 215,746, and an R^2 of 0.9289, indicating that the model can explain 92.89% of the variation in used car prices. Therefore, the developed model can be utilized as a decision support system to assist in determining used car prices more accurately.

Keywords: *Used car price prediction, Machine learning, Linear regression, Google Colab, Model evaluation.*

1. PENDAHULUAN

Kemajuan teknologi informasi dan machine learning telah mendorong perubahan signifikan dalam berbagai bidang, termasuk sektor otomotif. Pertumbuhan ekonomi masyarakat turut mendorong berkembangnya bisnis jual beli mobil bekas yang terus mengalami peningkatan setiap tahunnya[1]. Seiring meningkatnya produksi kendaraan baru, jumlah mobil bekas di pasaran juga semakin bertambah, sehingga persaingan dalam bisnis ini menjadi semakin ketat [2]. Penentuan harga mobil bekas merupakan salah satu permasalahan utama dalam transaksi jual beli. Harga kendaraan dipengaruhi oleh berbagai faktor, seperti usia kendaraan, jarak tempuh, kondisi mesin, kapasitas mesin, serta merek dan tipe kendaraan[3]. Ketidaktepatan dalam menentukan harga dapat menyebabkan kerugian bagi penjual maupun pembeli, sehingga diperlukan metode yang mampu melakukan estimasi harga secara objektif dan akurat[4].

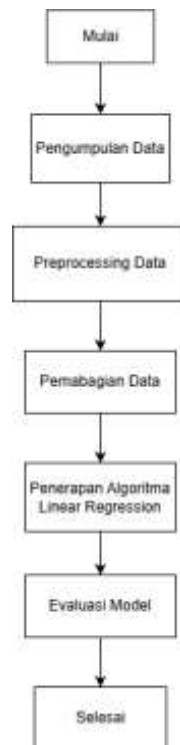
Dalam praktiknya, penentuan harga mobil bekas masih sering dilakukan secara subjektif berdasarkan pengalaman individu, sehingga hasilnya kurang konsisten dan tidak selalu sesuai dengan kondisi pasar[5]. Dengan berkembangnya teknologi, pemanfaatan *machine learning* menjadi solusi yang efektif dalam mengolah data historis untuk menghasilkan prediksi yang lebih akurat [6]. Algoritma *Linear Regression* mampu memodelkan hubungan antara variabel input seperti tahun produksi, jarak tempuh, kapasitas mesin, dan fitur kendaraan lainnya dengan variabel output berupa harga mobil. Melalui pemanfaatan data historis penjualan, model *linear regression* dapat digunakan untuk menghasilkan prediksi harga yang lebih akurat dan berbasis data[7]. Berdasarkan permasalahan tersebut, penelitian ini berfokus pada penerapan algoritma *Linear Regression* berbasis *machine learning* dalam memprediksi harga mobil bekas. Namun demikian, beberapa penelitian sebelumnya masih memiliki keterbatasan dalam pemilihan variabel dan evaluasi performa model, sehingga diperlukan pengembangan model yang lebih optimal untuk meningkatkan akurasi prediksi[8]. Oleh karena itu, penelitian ini bertujuan untuk membangun model prediksi harga mobil bekas menggunakan algoritma *Linear Regression* berbasis *machine learning*, serta mengevaluasi tingkat akurasi model yang dihasilkan. Hasil penelitian ini diharapkan

dapat memberikan estimasi harga yang lebih objektif dan mendukung pengambilan keputusan dalam transaksi jual beli mobil bekas.

2.METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pengembangan model prediksi harga mobil bekas dilakukan secara kuantitatif dengan algoritma *Linear Regression* melalui tahapan pengumpulan serta prapemrosesan data, pembagian dataset, pelatihan, serta evaluasi model yang seluruhnya diproses di Google Colab, sebagaimana tersaji pada Gambar 1.



Gambar 1 Alur Flowchart pada tahapan penelitian

Gambar 1 menerapkan pendekatan kuantitatif, karena seluruh proses analisis dilakukan melalui pengolahan data numerik dan pengukuran performa model secara statistik. Proses analisis dan pemodelan dilakukan pada platform Google Colab, yang dipilih karena menyediakan lingkungan komputasi berbasis cloud, mendukung berbagai pustaka machine learning, serta mempermudah integrasi dengan dataset berukuran besar dan beragam format[7]. Dataset yang digunakan berjumlah 8.128 data. Pada tahap *preprocessing*, dilakukan pembersihan dan penyesuaian data tanpa mengurangi jumlah data, sehingga seluruh data tetap digunakan dalam proses pemodelan [9]. Penggunaan keseluruhan data dilakukan karena dianggap lebih mampu merepresentasikan kondisi sebenarnya dibandingkan penggunaan sebagian data saja. Jumlah data yang besar memungkinkan model untuk mengenali pola hubungan antar variabel secara lebih akurat dan mengurangi potensi bias dalam proses analisis. Dengan memanfaatkan seluruh data yang tersedia, diharapkan hasil pemodelan yang dihasilkan menjadi lebih valid, reliabel, dan mencerminkan kondisi nyata, sehingga dapat meningkatkan keakuratan model prediksi yang dibangun[4]. Dataset ini dipilih karena memiliki variabel yang relevan untuk proses prediksi harga mobil dan sudah banyak digunakan dalam penelitian berbasis regresi serta *machine learning*.

2.2 Pengumpulan Data

Dataset yang dipakai pada kajian ini didapati dari platform Kaggle melalui judul *Car Price Prediction Dataset*. Dataset tersebut berjumlah 8.128 data mobil bekas, di mana setiap data merepresentasikan satu unit kendaraan. Dataset memuat berbagai atribut kendaraan, antara lain nama mobil (*name*), tahun produksi (*year*), harga jual (*selling_price*), jarak tempuh (*km_driven*), jenis bahan bakar (*fuel*), tipe penjual (*seller_type*), jenis transmisi (*transmission*), kepemilikan kendaraan (*owner*), konsumsi bahan bakar (*mileage*), kapasitas mesin (*engine*), daya maksimum (*max_power*), serta jumlah kursi (*seats*). Dataset disimpan dalam format CSV dengan pemisah titik koma (;), kemudian dimuat ke dalam lingkungan Google Colab menggunakan pustaka *pandas* untuk dilakukan analisis lebih lanjut.

Tabel 1. Dataset Harga Mobil Bekas

<i>n</i>	<i>name</i>	<i>year</i>	<i>selling_price</i>	<i>km_driven</i>	<i>fuel</i>	<i>seller_type</i>	<i>transmission</i>	<i>owner</i>	<i>mileage(km/ltr/kg)</i>	<i>engine</i>	<i>max_power</i>	<i>seats</i>
1	Maruti Swift Dzire VDI	2014	450000	145500	Diesel	Individual	Manual	First Owner	23.4	1248.0	74	5.0
2	Skoda Rapid 1.5 TDI Ambition	2014	370000	120000	Diesel	Individual	Manual	Second Owner	21.14	1498.0	103.52	5.0
3	Honda City 2017 - 2020 EXi	2016	158000	140000	Petrol	Individual	Manual	Third Owner	17.7	1497.0	78	5.0
4	Hyundai i20 Sport z Diesel	2010	225000	127000	Diesel	Individual	Manual	First Owner	23.0	1396.0	90	5.0
5	Maruti Swift VXI BSIII	2007	130000	120000	Petrol	Individual	Manual	First Owner	16.1	1298.0	88.2	5.0

2.3 Preprocessing Data

Tahapan preprocessing data bertujuan untuk mempersiapkan dataset sehingga layak dan siap digunakan pada tahap pemodelan. Dataset diperiksa untuk mengetahui struktur, tipe data, serta keberadaan nilai kosong. Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki missing value, sehingga tidak dilakukan penghapusan data. Preprocessing difokuskan pada penyesuaian format data dan transformasi variabel, seperti pembersihan atribut numerik dari satuan serta pengubahan atribut kategorikal ke bentuk data numerik direpresentasikan memakai cara *One-Hot Encoding* supaya diolah secara optimal pada algoritma *Linear Regression*[8].

2.4 Pembagian Data

Dataset hasil prapemrosesan menjadikan data latih serta data uji dari rasio 80:20 melalui fungsi *train_test_split*. Melalui total 8.128 data yang digunakan, sebanyak 6.502 data dialokasikan sebagai data latih yang terdiri dari *X_train* dan *Y_train* sedangkan 1.626 data sisanya digunakan sebagai data uji yang meliputi *X_test* dan *Y_test*. Sasaran pembagian dataset ialah menilai kapabilitas model dalam memprediksi data baru di luar proses pelatihan [10].

2.5 Penerapan Algoritma Linear Regressio

Pada kajian ini, algoritma Linear Regression diterapkan guna menciptakan model yang dipakai dalam memprediksi harga mobil bekas dengan memanfaatkan data historis kendaraan. Variabel dependen yang digunakan adalah harga mobil bekas (*selling_price*), sedangkan variabel independen meliputi tahun kendaraan (*year*), jarak tempuh (*km_driven*), kapasitas mesin (*engine*), daya maksimum mesin (*max_power*), dan jumlah kursi (*seats*)[12]. Karena penelitian ini melibatkan lebih dari satu variabel independen, maka model yang digunakan adalah regresi linear berganda. Hubungan antara variabel independen dan variabel dependen dinyatakan dalam bentuk persamaan regresi linear berganda sebagai berikut:

$$Y = \beta_0 + \beta_1 \chi_1 + \beta_2 \chi_2 + \beta_3 \chi_3 + \beta_4 \chi_4 + \beta_5 \chi_5 + \varepsilon \quad (1)$$

Keterangan variable:

- Y = Variabel dependen, yaitu harga mobil bekas (selling_price)
- β_0 = Intercept atau konstanta, yaitu nilai harga mobil ketika seluruh variabel independen bernilai nol
- β_1 = merupakan koefisien regresi yang merepresentasikan pengaruh variabel tahun kendaraan (year) terhadap model.
- β_2 = merupakan koefisien regresi yang merepresentasikan pengaruh variabel jarak tempuh kendaraan (km_driven) terhadap model.
- β_3 = Koefisien regresi untuk variabel kapasitas mesin (engine)
- β_4 = Koefisien regresi untuk variabel daya maksimum mesin (max_power)
- β_5 = Koefisien regresi untuk variabel jumlah kursi (seats)
- χ_1 = Tahun produksi kendaraan
- χ_2 = Jarak tempuh kendaraan (kilometer)
- χ_3 = Kapasitas mesin kendaraan (cc)
- χ_4 = Daya maksimum mesin (bhp)
- χ_5 = Jumlah kursi kendaraan
- ε = Error term, yaitu selisih antar nilai harga aktual serta nilai harga hasil prediksi model

Setiap parameter regresi (β) merepresentasikan tingkat kontribusi masing-masing variabel bebas dalam memengaruhi nilai harga mobil bekas, dengan asumsi bahwa variabel bebas lainnya berada pada kondisi tetap. Koefisien regresi yang bernilai positif mengindikasikan adanya pengaruh searah, di mana peningkatan nilai variabel tersebut cenderung diikuti oleh kenaikan harga mobil bekas. Sebaliknya, koefisien bernilai negatif menunjukkan pengaruh yang berlawanan arah. Sementara itu, komponen galat (ε) menjelaskan dampak faktor-faktor lain di luar variabel yang dimodelkan yang mungkin berkontribusi terhadap variasi harga mobil bekas tetapi tidak termasuk dalam studi ini [13]. Untuk dataset berukuran besar, proses perhitungan koefisien dan evaluasi model dilakukan menggunakan bantuan perangkat lunak Google Colab guna memastikan efisiensi dan akurasi perhitungan. Perhitungan manual ditampilkan secara terbatas sebagai ilustrasi konsep perhitungan [14].

2.6 Evaluasi Model

Evaluasi performa serta akurasi model dilakukan melalui MAE, MSE, RMSE, serta koefisien determinasi (R^2) setelah Linear Regression dipakai memprediksi harga mobil menggunakan data uji [15]. Metrik-metrik ini digunakan untuk meningkatkan akurasi prediksi dan kinerja model saat mendeskripsikan variasi harga mobil [16].

MSE: Menghitung rerata kuadrat dari kesalahan prediksi

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

MSE : Kinerja prediksi model dinilai semakin baik ketika nilai MSE semakin kecil, sebab metrik ini mengukur rerata kuadrat selisih antar nilai aktual serta nilai prediksi.

RMSE: Berlandaskan MSE, RMSE mengindikasikan masalah pada harga tunggal.

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

RMSE merupakan akar kuadrat dari MSE yang digunakan untuk menggambarkan besarnya kesalahan prediksi dalam satuan yang mempunyai kesamaan seperti data aslinya, sehingga hasil evaluasi menjadi lebih mudah untuk diinterpretasikan.

R-squared (R^2): Menunjukkan seberapa besar variansi harga mobil yang dapat dijelaskan oleh model.

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

Koefisien determinasi (R^2) digunakan untuk menggambarkan proporsi variasi data yang mampu didefinisikan pada model. Nilai R^2 yang ada pada angka 1 memperlihatkan bahwasannya model memiliki kemampuan prediksi yang kian baik.

Keterangan Umum Matrik Evaluasi:

- m :Jumlah observasi data.
- y_1 :Nilai actual harga mobil untuk pengamatan ke- i .
- \hat{y}_i :Prediksi harga mobil untuk pengamatan ke- i .
- \hat{y} : Rata-rata (*mean*) dari actual harga mobil

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Kajian ini ditujukan guna menciptakan model prediksi harga mobil bekas memakai pendekatan *Linear Regression*. Dataset yang digunakan telah dari tahap *preprocessing*, meliputi pembersihan data, penyesuaian tipe data, serta penanganan nilai kosong. Tahap seterusnya ialah melakukan penyebaran dataset ke dalam data latih serta data uji melalui dibandingkan 80:20. Model Linear Regression kemudian dilatih menggunakan beberapa variabel prediktor, meliputi tahun kendaraan (*year*), jarak tempuh (*km_driven*), kapasitas mesin (*engine*), daya maksimum (*max_power*), serta jumlah kursi (*seats*), dengan harga jual kendaraan (*selling_price*) sebagai variabel target. Untuk menilai tingkat ketepatan dan kemampuan model dalam memprediksi harga mobil bekas, Performa model dinilai melalui MAE, MSE, RMSE, serta koefisien determinasi (R^2).

```
import pandas as pd #Import Library
import numpy as np
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

Gambar 3 Library yang digunakan

Gambar 3 menunjukkan data set yang dipakai pada studi ini ada pada 8.128 data mobil bekas, di mana setiap data merepresentasikan satu unit mobil dengan karakteristik tertentu. Dataset ini memuat berbagai atribut yang berkaitan dengan kondisi dan spesifikasi kendaraan, seperti tahun produksi, jarak tempuh, kapasitas mesin, jenis bahan bakar, jenis transmisi, serta tipe penjual. Atribut-atribut tersebut digunakan guna variabel terikat yang diberikan dugaan menghasilkan pengaruh kepada harga jual mobil bekas. Sementara itu, harga mobil bekas dijadikan sebagai variabel dependen atau target yang akan diprediksi menggunakan algoritma Linear Regression. Jumlah data sebanyak 8128 dinilai cukup representatif untuk membangun model prediksi yang stabil serta mampu merepresentasikan pola hubungan antara variabel-variabel penentu dengan harga mobil bekas secara efektif. Pendekatan ini memungkinkan model machine learning guna menghasilkan prediksi yang akurat serta bisa dipertanggungjawabkan secara akademik.

```
# Load dataset (pakai separator ;)
df = pd.read_csv("data_mobil.csv", sep=';')

# Tampilkan 5 data awal
df.head()
```

Gambar 4 Proses Memuat Dataset

Gambar 4 menunjukkan proses memuat dataset menggunakan Python dengan library Pandas. Dataset dibaca dari file *data_mobil.csv* menggunakan pemisah titik koma (;), kemudian ditampilkan lima data teratas dengan *df.head()*. Tahap ini bertujuan untuk melihat struktur awal data dan memastikan bahwa dataset berhasil dimuat dengan benar.

```
df['mileage(km/ltr/kg)'].dtype
df['engine'].dtype
df['max_power'].dtype
|
... dtype('float64')
```

Gambar 5 Proses Pengecekan tipe data

Gambar 5 memperlihatkan proses pengecekan tipe data (datatype) pada beberapa atribut, yaitu *mileage*, *engine*, dan *max_power* menggunakan `.dtype`. Tujuannya adalah memastikan bahwa atribut-atribut tersebut sudah bertipe numerik (misalnya float64) sehingga dapat digunakan dalam proses perhitungan dan pemodelan machine learning.

Tabel 2. Statistik Deskriptif Dataset Harga Mobil Bekas

Statistik	Year	Selling Price	Km Driven	Mileage(km/ltr/kg)	Engine	Seats
Count	8128	8128	8128	7907	7907	7907
Mean	2013.804	638271.8	69819.51	19.418	1458.63	5.417
Std	4.044	806253.4	56550.55	4.037	503.916	0.959
Min	1983	29999	1	0.0	624	2
25%	2011	254999	35000	16.78	1197	5
50%	2015	450000	60000	19.30	1248	5
75%	2017	675000	98000	22.32	1582	5
Max	2020	1000000	2360457	42.00	3604	14

3.2 Hasil Preprocessing

```
year 0
selling_price 0
km_driven 0
mileage(km/ltr/kg) 0
engine 0
...
transmission_Manual 0
owner_Fourth & Above Owner 0
owner_Second Owner 0
owner_Test Drive Car 0
owner_Third Owner 0
```

2074 rows x 1 columns

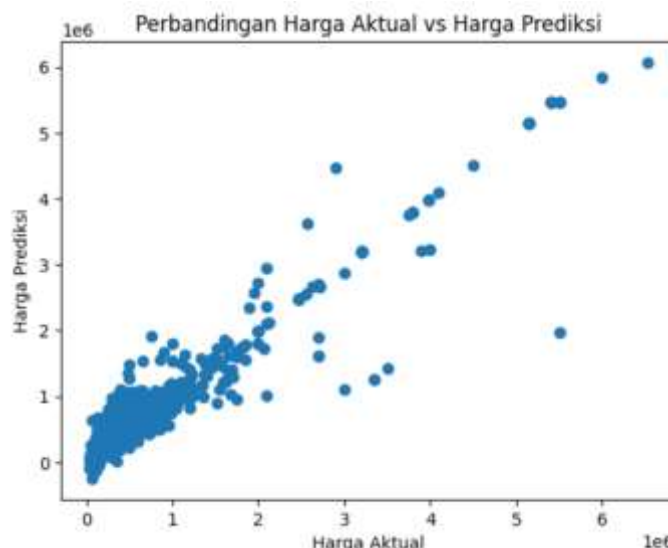
Gambar 7 Preprocessing berhasil

Setelah dilakukan proses *preprocessing*, dataset yang digunakan dalam penelitian ini berada dalam kondisi siap digunakan untuk pemodelan. Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki nilai kosong (missing value), sehingga tidak dilakukan proses penghapusan data. Tahap preprocessing difokuskan pada penyesuaian format data dan transformasi variabel agar seluruh atribut dapat diproses dengan baik oleh algoritma *Linear Regression*. Selain itu, dilakukan proses transformasi data pada atribut yang bersifat kategorikal dengan mengubahnya ke dalam bentuk numerik menggunakan teknik encoding. Seluruh variabel numerik juga telah disesuaikan tipe datanya agar sesuai dengan kebutuhan pemodelan. Dengan kondisi data yang terstruktur dan konsisten, dataset telah memenuhi syarat untuk digunakan pada tahap pelatihan serta uji pada model.

Tabel 3. Hasil pembagian data

Data set	Jumlah data
X_train	6502
X_test	1626
Y_train	6502
Y_test	1626

Pada fase pemodelan, kajian ini memanfaatkan dataset yang memuat 8.128 data mobil bekas. Keseluruhan data kemudian dipisahkan ke dalam dua kategori, yakni data latih serta data uji melalui perbandingan 80% serta 20%. Data latih dipakai untuk membangun serta melatih model Linear Regression agar mampu memahami ikatan antar variabel independen serta variabel dependen. Sementara itu, data uji dimanfaatkan guna menilai kinerja model dalam melakukan prediksi terhadap data yang tidak disertakan pada tahap pelatihan. Hasil pembagian tersebut menghasilkan 6.502 data sebagai data latih yang mencakup X_train serta Y_train. Selanjutnya, data uji terdiri atas 1.626 data X_test serta 1.626 data Y_test. Melalui skema pembagian ini, model yang dikembangkan diupayakan bisa memberikan penghasilan terkait prediksi harga mobil bekas yang lebih akurat serta mempunyai kemampuan generalisasi yang baik kepada data baru.



Gambar 8 Grafik Harga Aktual vs Harga Prediksi

Kedekatan antara nilai prediksi model serta nilai aktual tercermin pada Gambar 8, dimana grafik perbandingan memperlihatkan sebagian besar distribusi data ada pada garis diagonal, hingga memperlihatkan bahwasannya hasil prediksi yang dihasilkan mempunyai kesesuaian yang cukup baik dengan nilai sebenarnya. Meskipun demikian, masih terdapat beberapa titik yang menyimpang, khususnya pada rentang harga mobil yang tinggi, yang menunjukkan adanya kesalahan prediksi pada nilai tertentu.

Tabel hasil prediksi menunjukkan perbedaan antar harga mobil bekas aktual dengan harga yang dihasilkan oleh model *Linear Regression*. Nilai prediksi diperoleh melalui perhitungan matematis menggunakan persamaan *Linear Regression*, di mana setiap variabel input seperti tahun kendaraan, jarak tempuh, kapasitas mesin, dan fitur lainnya dikalikan dengan koefisien masing-masing, kemudian dijumlahkan dengan konstanta model. Hasil perbandingan menunjukkan bahwa sebagian besar nilai prediksi mendekati harga aktual. Hal ini menandakan bahwa model *Linear Regression* yang dibangun mampu mempelajari pola hubungan antara karakteristik kendaraan dengan harga jual mobil bekas secara cukup baik.

Tabel 4 Hasil Evaluasi Model

MAE	MSE	RMSE	R ²
96477.167219665	46546431084.3728	215746.21916588204	0.9289892662354626

Performa model Linear Regression yang dikembangkan tercermin dari hasil evaluasi yang menunjukkan nilai MAE yakni 96.477, MSE yakni 46.546.431.084, serta RMSE yakni 215.746. Nilai MAE tersebut mengindikasikan bahwasannya secara rerata selisih antar hasil prediksi model serta harga aktual kendaraan berada pada kisaran 96 ribu.

Daripada itu, nilai RMSE yang relatif lebih besar memperlihatkan adanya sejumlah prediksi dengan tingkat kesalahan yang cukup tinggi, khususnya pada kendaraan dengan harga jual yang besar. Selain itu, koefisien determinasi (R^2) yakni 0,9289 memperlihatkan bahwasannya 92,89% variasi harga mobil bekas mampu dijelaskan oleh model, hingga bermakna bahwasannya model mempunyai performa yang sangat baik dalam memprediksi harga mobil bekas.

4. KESIMPULAN

Prediksi biaya mobil bekas dapat dilakukan secara cukup akurat melalui penerapan algoritma Linear Regression berbasis machine learning, sebagaimana ditarik kesimpulan dari hasil kajian. Model dibangun menggunakan dataset sebanyak 8.128 data mobil bekas yang telah melalui tahapan preprocessing, pembagian data, serta pelatihan model di lingkungan Google Colab. Hasil evaluasi menunjukkan bahwa model memiliki performa yang baik, ditunjukkan oleh nilai MAE sebesar 96.477, RMSE sebesar 215.746, dan nilai koefisien determinasi (R^2) sebesar 0,9289. Nilai R^2 tersebut mengindikasikan bahwa model mampu menjelaskan sekitar 92,89% variasi harga mobil bekas, sehingga model memiliki kemampuan prediksi yang kuat dan dapat dijadikan sebagai alat bantu dalam menentukan estimasi harga kendaraan secara lebih objektif. Dengan demikian, penelitian ini membuktikan bahwa pendekatan regresi linear dapat membantu meningkatkan akurasi penentuan harga mobil bekas dibandingkan dengan metode subjektif. Model yang dibangun diharapkan dapat memberikan manfaat bagi penjual dan pembeli sebagai bahan pertimbangan dalam pengambilan keputusan. Namun demikian, kajian ini masih mempunyai keterbatasan, sehingga pengembangan lebih lanjut bisa dilaksanakan melalui ditambahkannya variabel yang lebih beragam atau membandingkan dengan algoritma pembelajaran mesin lainnya untuk mendapatkan hasil prediksi yang lebih optimal.

UCAPAN TERIMAKASIH

Dukungan serta bantuan dari berbagai pihak selama tahap pelaksanaan kajian serta menyusun jurnal ini memperoleh apresiasi dari penulis, dengan harapan hasil kajian mampu memberikan manfaat serta menjadi rujukan bagi kajian di masa mendatang.

DAFTAR PUSTAKA

- [1] W. Wilianto, Y. Yuliana, A. Suwandhi, J. Jimmy, and J. Putra, "Penerapan AI dalam Menentukan Harga Mobil Bekas Berdasarkan Tahun Perakitan," *J. Minfo Polgan*, vol. 13, no. 1, pp. 550–560, 2024, doi: 10.33395/jmp.v13i1.13728.
- [2] A. A. Bawa, I. N. Sukajaya, and I. M. A. Wirawan, "Peramalan Jumlah Kendaraan Diuji Baik Menggunakan Metode Regresi Linier Berganda," *Decod. J. Pendidik. Teknol. Inf.*, vol. 5, no. 1, pp. 240–251, 2025, doi: 10.51454/decode.v5i1.1104.
- [3] H. Fadhilah, Nudin, and Salamun Rohman, "Sistem Informasi Prediksi Harga Mobil Bekas Menggunakan Metode KNN," pp. 1–5, 2022.
- [4] F. H. Hamdanah and D. Fitriana, "Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikro, Kecil, dan Menengah," *J. Nas. Pendidik. Tek. Inform.*, vol. 10, no. 1, p. 23, Apr. 2021, doi: 10.23887/janapati.v10i1.31035.
- [5] M. D. H. Kusuma and S. Hidayat, "Penerapan Model Regresi Linier dalam Prediksi Harga Mobil Bekas di India dan Visualisasi dengan Menggunakan Power BI," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 2, pp. 1097–1110, 2024, doi: 10.35870/jimik.v5i2.629.
- [6] Noto Narwanto, Kusri, and Hanif Al Fatta, "Prediksi Peserta Matakuliah Menggunakan Artificial Neural Network - Fuzzy Inferenced System (Ann-Fis) Studi Kasus: Universitas Muhammadiyah Surakarta," *J. Technol. Informatics*, vol. 1, no. 2, pp. 84–88, 2020, doi: 10.37802/joti.v2i1.49.
- [7] Amalia Yunia Rahmawati, "Analisis Prediksi Dengan Metode Regresi Linier di Pt. Eagle Industry Indonesia," vol. 5, no. July, pp. 1–23, 2020.
- [8] P. Aulia Azhar, M. Arya Pratama, and R. Fitriani, "Prediksi Harga Mobil Audi Bekas Menggunakan Model Regresi Linear dengan Framework Streamlit," *J. Technol. Informatics*, vol. 6, no. 1, pp. 22–28, 2024, doi: 10.37802/joti.v6i1.763.
- [9] Sukhmandeep Singh Brar, "Car Price Prediction Dataset," Updated 2 years ago. [Online]. Available: <https://www.kaggle.com/datasets/sukhmandeepsinghbrar/car-price-prediction-dataset?resource=download>
- [10] S. Muti and K. Yildiz, "Using Linear Regression For Used Car Price Prediction," *Int. J. Comput. Exp. Sci. Eng.*, vol. 9, no. 1, pp. 11–16, 2023, doi: 10.22399/ijcesen.1070505.
- [11] N. Ariyani and A. Z. Arifin, "Prediksi Tingkat Pengangguran Terbuka Kabupaten Tuban dengan Metode Regresi Linier Sederhana," *J. UNIROW*, vol. 03, no. 01, pp. 6–13, 2021.
- [12] D. Penerapan and M. Fifo, "Prediksi Penjualan Sepeda Motor Second Menggunakan Algoritma Regresi Linier," *J. Sist. Inf.*, vol. 06, pp. 17–23, 2023.
- [13] W. C. Ferdinan, M. R. Noerfikri, P. A. Panchadri, and F. Ferawati, "Implementasi Algoritma Regresi Linear Berganda untuk Memprediksi Prestasi Siswa," *bit-Tech*, vol. 7, no. 3, pp. 853–864, 2025, doi: 10.32877/bt.v7i3.2228.
- [14] S. Mulyanda, S. Defit, and Sumijan, "Analisis Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Harga Pasar Mobil Bekas," *J. KomtekInfo*, vol. 10, pp. 116–121, 2023, doi: 10.35134/komtekinfo.v10i3.427.
- [15] S. Lestari, "Analisis Algoritma Regresi Linear Sederhana dalam Memprediksi Tingkat Penjualan Album KPOP," *INSOLOGI J. Sains dan Teknol.*, vol. 2, no. 1, pp. 199–209, 2023, doi: 10.55123/insologi.v2i1.1692.
- [16] M. B. S. Qolbi, T. N. Puteh, R. Rivandi, and C. Rozikin, "Prediksi Harga Rumah Di Jakarta Pusat Menggunakan Algoritma Machine Learning," *J. Ilmu Komput. dan Bisnis*, vol. 16, no. 1, pp. 16–24, 2025, doi: 10.47927/jikb.v16i1.840.