

Prediksi Harga Mobil Audi Bekas Menggunakan Model Regresi Linear dengan *Framework Streamlit*

Putri Aulia Azhar¹, Muhammad Arya Pratama², Risma Fitriani³

^{1,2,3}Akuntansi, Universitas Singaperbangsa Karawang, Karawang, Jawa Barat, Indonesia
e-mail: 2210631030128@student.unsika.ac.id¹, 2210631030115@student.unsika.ac.id²,
risma.fitriani@ft.unsika.ac.id³

* Penulis Korespondensi: E-mail: 2210631030128@student.unsika.ac.id

Abstrak: Prediksi harga mobil bekas menjadi aspek krusial dalam pasar otomotif yang terus berkembang, hal ini akan memengaruhi keputusan penjual dan pembeli serta dinamika pasar secara keseluruhan. Meskipun kompleksitas dan variasi kondisi mobil bekas sering menjadi hambatan, adopsi teknologi dan metode yang tepat diharapkan mampu mengatasi tantangan tersebut. Melalui program prediksi harga, peneliti bermaksud untuk memberikan solusi praktis dan efisien bagi penjual dan pembeli mobil bekas Audi, dengan tujuan meningkatkan keputusan yang terinformasi dan tepat. Penelitian ini menggunakan algoritma *Linear Regression* untuk menemukan hubungan linear antara variabel-variabel yang digunakan. Tahapan metode yang digunakan meliputi *problem recognition*, *research plan*, *dataset collection*, *data pre-processing*, *error calculation*, dan *web deployment*. Hasilnya menunjukkan nilai akurasi sebesar 94% dan nilai *Mean Absolute Error* (MAE) sebesar 0,08 atau 8%. Hal ini memberikan indikasi yang kuat bahwa model *linear regression* yang digunakan telah berhasil dalam menghasilkan prediksi yang akurat untuk harga mobil bekas Audi.

Kata Kunci: *Machine Learning*; Prediksi; Regresi Linear; *Streamlit*

Abstract: Used car price prediction is a crucial aspect of the ever-evolving automotive market, affecting the decisions of sellers and buyers as well as the overall market dynamics. While the complexity and varied conditions of used cars are often an obstacle, the adoption of appropriate technologies and methods is expected to overcome these challenges. Through a price prediction program, the researcher intends to provide a practical and efficient solution for sellers and buyers of Audi used cars, with the aim of improving informed and precise decisions. This research utilizes the *Linear Regression* algorithm to find a linear relationship between the variables used. The stages of the method used include *problem recognition*, *research plan*, *dataset collection*, *data pre-processing*, *error calculation*, and *web deployment*. The results show an accuracy value of 94% and a *Mean Absolute Error* (MAE) value of 0.08 or 8%. This gives a strong indication that the *linear regression* model used has been successful in producing accurate predictions for Audi used car prices.

Keywords: *Machine Learning*; Prediction; Linear Regression; *Streamlit*

PENDAHULUAN

Prediksi merupakan analisis data dan informasi untuk meramalkan peristiwa di masa depan. Dalam bidang ilmiah, prediksi berguna untuk memproyeksikan hasil penelitian, perubahan iklim, atau dampak dari aktivitas tertentu. Proyeksi ini menggunakan informasi dari data historis, model matematika, dan analisis statistik. Dalam praktiknya, prediksi harus dilakukan secara terstruktur dan logis, serta mempertimbangkan berbagai faktor yang dapat mempengaruhi hasil akhirnya. Prediksi dapat digunakan sebagai landasan untuk mengambil keputusan strategis yang tepat di berbagai bidang, termasuk riset, kebijakan, dan bisnis [1].

Prediksi harga mobil bekas merupakan suatu aspek penting dalam pasar otomotif yang terus berkembang. Dalam konteks ini, prediksi menjadi sebuah instrumen vital bagi para penjual dan pembeli untuk mengambil keputusan yang tepat. Prediksi harga mobil bekas tidak hanya mempengaruhi keputusan individu, tetapi juga dinamika pasar secara keseluruhan. Kemampuan untuk meramalkan harga dengan akurat memungkinkan penjual untuk menentukan harga yang

kompetitif dan mengoptimalkan keuntungan, sementara pembeli dapat membuat keputusan yang lebih informasional dan tepat waktu. Namun, kendala-kendala yang dihadapi dalam menentukan harga jual maupun beli mobil bekas sering kali menjadi hambatan yang signifikan. Meskipun demikian, kompleksitas pasar dan variasi yang luas dalam kondisi mobil bekas sering kali membuat proses prediksi menjadi rumit dan membutuhkan pendekatan yang cermat. Dengan adopsi teknologi dan metode yang tepat, diharapkan bahwa program prediksi harga ini dapat menjadi solusi yang efektif dalam mengatasi tantangan-tantangan tersebut, menyediakan informasi yang berharga bagi para pemangku kepentingan dalam industri otomotif [2].

Mobil bekas Audi telah menjadi populer di pasar, terutama di kalangan penggemar mobil yang mencari kualitas, teknologi, dan desain yang tinggi namun dengan biaya yang lebih terjangkau. Salah satu alasan popularitas mobil bekas Audi adalah karena kualitas yang sangat baik dan teknologi yang diterapkan pada mobil-mobil ini. Merek ini juga dikenal dengan inovasi teknologi yang terus-menerus dilakukan, seperti pengembangan transmisi *multitronic* yang

memungkinkan pengemudi untuk mengendarai mobil dengan lebih mudah dan efisien. Dalam beberapa tahun terakhir, Audi juga telah meningkatkan penjualan mobil bekas dengan meluncurkan berbagai model yang lebih murah dan lebih efisien. Hal ini membuat mobil bekas Audi menjadi pilihan yang lebih terjangkau bagi banyak penggemar, tanpa mengorbankan kualitas dan teknologi yang tinggi [3].

Dengan menghadapi kompleksitas dan ketidakpastian dalam menentukan harga mobil bekas, kebutuhan akan alat yang dapat memberikan panduan yang jelas dan efektif semakin mendesak. Program prediksi harga ini dikembangkan tidak hanya menghadirkan kemudahan, tetapi juga akurasi dalam proses penentuan harga mobil bekas. Program ini dirancang dengan tujuan utama untuk memberikan solusi yang praktis dan efisien bagi penjual dan pembeli, sehingga mereka dapat mengambil keputusan yang terinformasi dan tepat. Melalui layanan *website* yang *user-friendly*, pengguna dapat dengan mudah mengakses prediksi harga yang disesuaikan dengan kriteria mereka, meningkatkan efisiensi dan keselarasan dalam transaksi jual-beli mobil bekas [4].

Machine Learning (ML) adalah disiplin ilmu yang mempelajari cara mesin belajar dari data untuk meningkatkan kemampuan prediksi tanpa bergantung pada aturan yang telah ditetapkan secara spesifik. Dalam ML, mesin mampu membuat model dari data yang disediakan dan memperbaiki model tersebut melalui iterasi pelatihan. Pendekatan ini menggunakan teknik *Supervised Learning (SL)*, yang terbukti efektif dalam analisis data historis untuk menghasilkan prediksi yang akurat. Dengan menerapkan SL, sistem dapat belajar dari data masa lalu dan mengembangkan model untuk memprediksi hasil yang konsisten dengan data *input*, sehingga meningkatkan akurasi prediksi yang diberikan [5].

Streamlit merupakan sebuah *framework open-source* yang bermanfaat untuk menciptakan aplikasi web yang dapat diakses menggunakan bahasa pemrograman Python. Dengan *framework* ini, pengembang dapat membangun aplikasi dengan mudah menggunakan skrip Python yang sederhana, tanpa perlu memiliki pengetahuan mendalam tentang pengembangan web di *frontend*. Penggunaan Streamlit memungkinkan pembuatan antarmuka pengguna (UI) responsif dengan cepat untuk menampilkan visualisasi data, tabel, grafik, dan elemen interaktif lainnya. Selain itu, Streamlit juga menyediakan fitur-fitur yang mempermudah penyebaran aplikasi sehingga dapat diakses secara daring dengan lancar oleh pengguna lain. Karena alasan tersebut, Streamlit telah menjadi pilihan populer di kalangan pengembang *data science* dan *machine learning* untuk mempercepat proses pembuatan dan distribusi aplikasi berbasis web [6].

Sebuah studi sebelumnya dilakukan oleh Ernianti Hasibuan, dkk (2022), yang menggunakan regresi linear untuk memprediksi harga mobil dengan variabel numerik, mencapai akurasi model sebesar 76% [7]. Selain itu, Amalia (2021) juga mengadakan

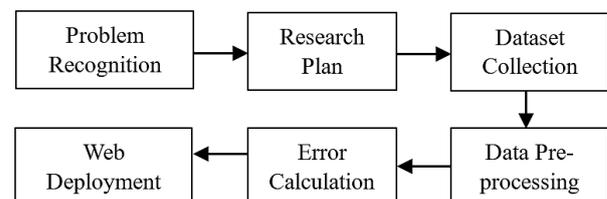
penelitian serupa untuk prediksi harga mobil dengan menggunakan model *Gradient Boost Regression* serta penyetelan *hyper-parameter* untuk mengubah data variabel teks menjadi nilai numerik, dan mendapatkan akurasi prediksi sebesar 67% [8]. Selanjutnya penelitian oleh Dea Miftahul Huda, dkk (2024) menghasilkan prediksi harga mobil bekas menggunakan regresi linear berganda dengan hasil *relative error* sebesar 11,89% yang menunjukkan hasil akurat [9].

Studi terdahulu lain oleh Sulaiman dan Edi Surya Negara (2023) menyatakan tentang prediksi harga mobil bekas menggunakan perbandingan algoritma *K-Nearest* dan *Random Forest* yang menunjukkan hasil dari *Random Forest* lebih baik yaitu 96,38% dibanding *K-Nearest* sebesar 59,17% [10]. Selanjutnya terdapat penelitian oleh Reynaldi, dkk (2021) yang menggunakan metode *fuzzy tsukamoto* dan *sugeno* dalam memprediksi harga mobil. Hasil yang didapat yaitu metode *fuzzy tsukamoto* mempunyai hasil yang lebih akurat yaitu tingkat error 8% [11].

Berdasarkan penelitian relevan sebelumnya, terdapat perbedaan akurasi untuk prediksi dimana tipe variabel yang digunakan hanya tipe data numerik, sedangkan untuk tipe data obyek tidak digunakan serta terdapat perubahan menjadi angka. Selain itu terdapat pula perbedaan algoritma yang digunakan. Penelitian ini akan mengikutsertakan tipe data obyek untuk dijadikan acuan prediksi harga mobil bekas. Oleh karena itu, penulis memutuskan untuk membuat program analisis prediksi harga mobil Audi bekas menggunakan *machine learning* dengan algoritma *linear regression* melalui *framework* streamlit.

METODE

Metode penelitian yang tepat merupakan pondasi utama dalam membangun keandalan dan validitas hasil suatu penelitian ilmiah. Penelitian ini menggunakan metode yang ditunjukkan oleh langkah-langkah pada Gambar 1.



Gambar 1. Langkah-langkah Penelitian

Problem Recognition

Pada langkah pertama ini ditemukan bahwa dalam industri mobil bekas, para penjual dan pembeli sering mengalami kendala dalam menentukan harga yang sesuai. Hal ini disebabkan oleh keterbatasan informasi yang jelas dan terintegrasi mengenai faktor-faktor yang mempengaruhi harga mobil bekas. Saat ini, meskipun informasi tersedia secara luas di *internet*, kebanyakan data yang ditemukan tersebar di berbagai sumber yang terpisah. Hal ini mengakibatkan proses pengambilan keputusan terasa rumit dan memakan banyak waktu.

Kendala utama yang dihadapi adalah ketidakpastian dalam penilaian harga mobil bekas karena informasi yang tidak terstruktur dan seringkali tidak akurat. Para penjual dan pembeli harus menyusun informasi dari berbagai sumber yang berbeda, seperti situs web jual-beli, forum diskusi otomotif, dan ulasan pengguna untuk mencoba memahami faktor-faktor yang mempengaruhi harga mobil bekas. Namun, karena ketidaktersediaan data yang lengkap dan terintegrasi, keputusan yang diambil tidak selalu optimal atau didasarkan pada informasi yang cukup.

Selain itu, proses pengumpulan informasi yang tidak terstruktur juga mengakibatkan pemborosan waktu yang signifikan. Penjual dan pembeli harus menghabiskan waktu mereka untuk mencari dan memverifikasi informasi dari berbagai sumber, yang pada akhirnya bisa saja tidak memberikan gambaran yang jelas atau akurat mengenai harga yang sebenarnya.

Dengan mengidentifikasi permasalahan ini, dapat dipahami bahwa diperlukan solusi yang menyediakan akses terpadu dan akurat terhadap informasi mengenai harga mobil bekas. Hal ini akan membantu meningkatkan kepercayaan dan efisiensi dalam proses jual-beli mobil bekas, serta mengurangi ketidakpastian yang mungkin dialami oleh para pelaku pasar.

Research Plan

Langkah selanjutnya yaitu melakukan rencana penelitian dengan menentukan metode yang tepat terkait prediksi yang dilakukan. Peneliti menggunakan metode *linear regression* atau regresi linier yang digunakan untuk menemukan korelasi linier antara variabel dan untuk memprediksi harga mobil bekas berdasarkan data historis yang tersedia [12]. Tujuan utama metode ini adalah untuk menemukan garis lurus terbaik yang dapat menjelaskan hubungan antara variabel *input* dan *output* dalam data.

Prosesnya dimulai dengan menghitung hubungan antara variabel *input* dan *output* menggunakan teknik seperti metode kuadrat terkecil. Ini melibatkan penghitungan koefisien yang tepat untuk variabel *input* yang dapat meminimalkan kesalahan antara nilai prediksi dan nilai aktual dari data yang diamati. Hasil dari *linear regression* dapat digunakan untuk memprediksi nilai *output* berdasarkan nilai *input* yang diberikan dengan rumus pada persamaan 1 [13].

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

Ket:

Y = variabel terikat

a = konstanta

$b \dots n$ = koefisien regresi

$X \dots n$ = variabel bebas

Dataset Collection

Langkah ini mencakup proses pengumpulan data atau himpunan data yang terstruktur dari berbagai sumber untuk keperluan pengembangan model. Sumber data yang akan digunakan berasal dari dataset publik yang tersedia di Kaggle dengan judul "100,000 UK Used Car Data set" oleh Aditya. Dataset ini disimpan dalam

format *csv* yang terdiri dari 9 kolom dan 10.668 baris yang berisi informasi mengenai model mobil, tahun mobil, harga, transmisi, jarak tempuh, jenis bahan bakar, besaran pajak, konsumsi bahan bakar, dan ukuran mesin [14]. Adanya fitur-fitur kategorial dan kombinasi data akan menambah kompleksitas analisis, memungkinkan model untuk mempertimbangkan berbagai faktor yang berpengaruh terhadap harga mobil bekas secara menyeluruh. Berikut tampilan awal dari dataset yang ditampilkan pada Gambar 2.

	A	B	C	D	E	F	G	H	I
1	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize
2	A1	2017	12500	Manual	15735	Petrol	150	55.4	1.4
3	A6	2016	16500	Automatic	36203	Diesel	20	64.2	2
4	A1	2016	11000	Manual	29946	Petrol	30	55.4	1.4
5	A4	2017	16800	Automatic	25952	Diesel	145	67.3	2
6	A3	2019	17300	Manual	1998	Petrol	145	49.6	1
7	A1	2016	13900	Automatic	32260	Petrol	30	58.9	1.4
8	A6	2016	13250	Automatic	76788	Diesel	30	61.4	2
9	A4	2016	11750	Manual	75185	Diesel	20	70.6	2
10	A3	2015	10200	Manual	46112	Petrol	20	60.1	1.4
11	A1	2016	12000	Manual	22451	Petrol	30	55.4	1.4
12	A3	2017	16100	Manual	28955	Petrol	145	58.9	1.4
13	A6	2016	16500	Automatic	52198	Diesel	125	57.6	2
14	Q3	2016	17000	Manual	44915	Diesel	145	52.3	2
15	A3	2017	16400	Manual	21695	Petrol	30	58.9	1.4

Gambar 2. Tampilan Awal Dataset

Data Pre-processing

Langkah ini merupakan langkah penting dalam *pipeline machine learning* yang melibatkan persiapan data untuk analisis lebih lanjut. Pertama, kita memisahkan fitur (X) dan target (y) dari dataset. Dalam kasus ini, target (y) adalah harga (*price*), sementara fitur (X) mencakup variabel *engine size*, *mpg*, *tax*, *fuel type*, *mileage*, *transmission*, *year*, dan *model*. Selanjutnya membagi data menjadi dua bagian yaitu data pelatihan dan data pengujian. Ini dapat dilakukan dengan menggunakan fungsi *train_test_split*, yang memungkinkan kita untuk memiliki dua set data terpisah yang digunakan untuk melatih model dan menguji performanya secara terpisah. Data pengujian sebesar 15% dan 85% merupakan data pelatihan yang terlampir pada Gambar 3.

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.15, random_state=2)
```

Gambar 3. Pemisahan Set Data

Setelah pemisahan, fitur kategoris memerlukan transformasi menggunakan *ColumnTransformer* dan *OneHotEncoder* agar dapat dipahami oleh model yang dibuat pada step 1. Hal ini dengan tujuan menggabungkan prediksi tipe data numerik dengan objek agar dapat diprediksi bersamaan. Setelah itu dibentuk pelatihan model dengan algoritma *linear regression* menggunakan data pelatihan pada step 2. Proses ini dapat dilihat pada Gambar 4.

```

step1 = ColumnTransformer(transformers=[
    ('col_tnf', OneHotEncoder(sparse=False, drop='first', handle_unknown='ignore')), [0,2,4]
], remainder='passthrough')

step2 = LinearRegression()

pipe = Pipeline([
    ('step1', step1),
    ('step2', step2)
])

pipe.fit(X_train, y_train)

y_pred = pipe.predict(X_test)
    
```

Gambar 4. Transformasi Fitur dan Pelatihan Model

Hasil akhir didapatkan 9.067 data pelatihan dengan 8 kolom berisi variabel *model*, *year*, *transmission*, *mileage*, *fuelType*, *tax*, *mpg*, dan *engineSize* yang akan digunakan dalam pembuatan model seperti yang dapat dilihat pada Gambar 5.

	model	year	transmission	mileage	fuelType	tax	mpg	engineSize	
	5900	Q5	2019	SemiAuto	5937	Diesel	145	38.2	2.0
	2964	A4	2015	SemiAuto	20385	Diesel	125	56.5	2.0
	5388	A4	2018	SemiAuto	31850	Petrol	145	36.7	3.0
	5649	A4	2017	Automatic	38221	Petrol	145	38.7	3.0
	4359	A1	2019	Manual	1308	Petrol	145	47.9	1.0

	1099	A3	2019	Manual	4567	Petrol	145	44.8	1.5
	2514	A4	2014	Manual	73206	Diesel	125	60.1	2.0
	6637	A4	2018	Manual	16217	Petrol	145	51.4	1.4
	2575	Q3	2016	Manual	21421	Petrol	145	49.6	1.4
	7336	A1	2016	Manual	33500	Petrol	30	58.9	1.4

9067 rows x 8 columns

Gambar 5. Tampilan Akhir Data *Pre-processing*

Error Calculation

Penerapan algoritma *linear regression* untuk menghitung *error* atau kesalahan dengan membandingkan harga mobil yang diprediksi dengan harga sebenarnya. Untuk melihat hal tersebut digunakan metrik *R2 score* dan *MAE (Mean Absolute Error)* pada Gambar 6.

```

print('R2 score', r2_score(y_test, y_pred))
print('MAE', mean_absolute_error(y_test, y_pred))
    
```

Gambar 6. Metrik Kesalahan Prediksi

R2 Score atau dikenal sebagai koefisien determinasi, adalah sebuah metrik statistik yang umum digunakan dalam analisis regresi. Rentang nilai *R2 Score* adalah dari 0 hingga 1, yang menggambarkan seberapa baik model regresi sesuai dengan data yang dianalisis. Semakin tinggi nilai *R2 Score*, semakin baik kemampuan model regresi dalam menjelaskan variasi dari variabel dependen dengan menggunakan variabel independen yang tersedia. Oleh karena itu, *R2 Score* memiliki peran penting dalam mengevaluasi presisi prediksi dari model regresi, serta memberikan gambaran tentang seberapa besar variasi data yang dapat dijelaskan oleh model tersebut [15]. Rumus untuk *R2 Score* dapat dilihat pada persamaan 2.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (2)$$

Ket:

SS_{res} = jumlah kuadrat residu

SS_{tot} = jumlah kuadrat total

Sedangkan *Mean Absolute Error (MAE)* adalah teknik yang digunakan untuk menilai presisi model prediksi dengan menghitung rata-rata kesalahan absolut antara nilai prediksi dan nilai sebenarnya. Dengan menggunakan *MAE*, kita dapat menilai seberapa besar kesalahan yang terjadi dalam prediksi. Pendekatan ini bermanfaat untuk mengevaluasi akurasi model prediksi dalam unit yang sama dengan data aslinya. Dalam analisis data, *MAE* digunakan untuk mengukur kinerja model dalam memprediksi nilai aktual, serta membantu dalam memahami variasi nilai dalam kumpulan data untuk meningkatkan kualitas prediksi [16]. Rumus untuk *MAE* dapat dilihat pada persamaan 3.

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

Ket:

n = jumlah total data

y_i = nilai aktual

\hat{y}_i = nilai yang diprediksi oleh model

$|y_i - \hat{y}_i|$ = nilai absolut

Hasil yang didapat adalah nilai *R2 score* sebesar 0,94 atau 94% dan nilai *MAE* sebesar 0,08 atau 8%. Nilai *R2 score* 94% menunjukkan bahwa model *linear regression* yang digunakan sangat baik dalam menjelaskan variasi harga mobil bekas Audi berdasarkan variabel-variabel yang digunakan dalam analisis. Untuk nilai *MAE*, semakin rendah nilainya, maka semakin kecil kesalahan prediksi model. Untuk nilai *MAE* 8% menunjukkan bahwa kesalahan prediksi sangat kecil. Hasil perolehan terlampir pada gambar 7.

R2 score 0.9411715017762609
 MAE 0.08811376199236715

Gambar 7. Hasil Perolehan *Error Calculation*

Web Deployment

Langkah selanjutnya adalah menyimpan model yang telah dilatih serta dataset yang telah dipersiapkan. Hal ini dapat dilakukan dengan menggunakan fungsi *pickle.dump*, yang memungkinkan untuk menyimpan objek Python ke dalam *file*. Dengan menyimpan model, kita dapat menggunakan kembali model tersebut di masa mendatang tanpa perlu melatih ulang. Sedangkan, menyimpan dataset dapat berguna untuk melacak versi data yang digunakan dalam pembangunan model, memungkinkan untuk replikasi hasil atau memperbarui model dengan data baru yang tersedia. Model yang digunakan terlampir pada Gambar 8.

```

import pickle

pickle.dump(df, open('df.pkl', 'wb'))
pickle.dump(pipe, open('pipe.pkl', 'wb'))
    
```

Gambar 8. Penyimpanan Model

Dalam proses *web deployment* untuk aplikasi prediksi harga mobil Audi bekas ini menggunakan *Streamlit* yaitu sebuah *framework* Python yang memudahkan pembuatan antarmuka pengguna web

interaktif. Aplikasi ini dimulai dengan memuat model prediksi dan data yang telah disimpan menggunakan *pickle*. Setelah itu, *interface* pengguna ditampilkan dengan judul sambutan dan gambar mobil Audi untuk memberikan konteks visual. Sebuah paragraf deskriptif mengenai merek Audi diintegrasikan dengan teks yang dibungkus secara rapi agar mudah dibaca. Bagian awal untuk model prediksi ditampilkan pada Gambar 9.

```

1 import streamlit as st
2 import pickle
3 import numpy as np
4 import pandas as pd
5 import textwrap
6
7 pipe = pickle.load(open('pipe.pkl','rb'))
8 df = pickle.load(open('df.pkl','rb'))
9
10 st.title('Selamat Datang!')
11 st.image('audiimage.jpg', width=400)
12
13 text = ""
14 Audi adalah merek mobil yang berasal dari Jerman, dikenal karena kombinasi antara desain elegan, teknologi inovatif, dan performa yang canggih.
15
16 wrapped_text = textwrap.fill(text, width=70)
17
18 st.markdown(f"<div style='text-align: justify;'>{wrapped_text}</div>", unsafe_allow_html=True)
19
20 st.title("Prediksi Harga Mobil Audi Bekas")

```

Gambar 9. Model Prediksi

Bagian inti dari aplikasi ini adalah fitur input di mana pengguna dapat memilih berbagai parameter mobil, seperti model, tahun, transmisi, jenis bahan bakar, jarak tempuh, pajak, konsumsi bahan bakar, dan ukuran mesin. Pengguna memasukkan informasi ini melalui *select box* dan *number input*. Setelah data dimasukkan, aplikasi menggunakan model prediksi untuk menghitung dan menampilkan estimasi harga mobil bekas Audi dalam bentuk yang mudah dibaca, dengan hasil akhir dikonversi ke dalam mata uang Rupiah. Model akhir ditampilkan pada Gambar 10.

```

23 model = st.selectbox('Input Model Mobil', df['model'].unique())
24 year = st.selectbox('Input Tahun Mobil', df['year'].unique())
25 transmission = st.selectbox('Input Transmisi Mobil', df['transmission'].unique())
26 fuelType = st.selectbox('Input Jenis Bahan Bakar Mobil', df['fuelType'].unique())
27 mileage = st.number_input('Input Jarak Tempuh Mobil')
28 tax = st.number_input('Input Pajak Mobil')
29 mpg = st.number_input('Input Konsumsi Bahan Bakar Mobil')
30 engineSize = st.number_input('Input Ukuran Mesin Mobil')
31
32 if st.button('Prediksi Harga Mobil'):
33     query = pd.DataFrame({
34         'model': [model],
35         'year': [year],
36         'transmission': [transmission],
37         'fuelType': [fuelType],
38         'mileage': [mileage],
39         'tax': [tax],
40         'mpg': [mpg],
41         'engineSize': [engineSize]
42     })
43
44     st.write(f"<div style='font-size: 30px; font-weight: bold;'>Prediksi Harga Mobil Audi Bekas Dalam Rupiah yaitu Rp
45         <math>{(int(np.exp(pipe.predict(query)[0]) * 20000))}</math></div>")

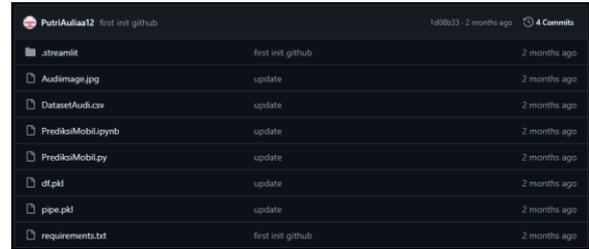
```

Gambar 10. Fitur Model Prediksi

Untuk *deploy* lanjutan aplikasi prediksi harga mobil Audi bekas ke Streamlit Share, langkah pertama adalah memindahkan semua file yang diperlukan ke *repository* GitHub. *Repository* ini akan berisi *file* Python yang mendefinisikan logika aplikasi, model yang disimpan dalam format *pickle*, serta file data yang digunakan untuk pelatihan dan prediksi. Setelah *repository* GitHub siap dan semua *file* terunggah dengan benar, langkah berikutnya adalah menyambungkan *repository* tersebut ke Streamlit Share.

Dalam Streamlit Share, pengguna dapat melakukan integrasi dengan GitHub dengan memasukkan URL *repository* dan melakukan konfigurasi yang diperlukan untuk memastikan aplikasi dapat berjalan dengan baik di *platform* tersebut. Setelah proses ini selesai, aplikasi akan tersedia secara publik, memungkinkan pengguna untuk mengakses dan menggunakan sistem prediksi harga mobil Audi bekas secara *online* melalui antarmuka *web* yang interaktif.

Tampilan GitHub yang digunakan dapat dilihat pada Gambar 11.



Gambar 11. Tampilan GitHub

HASIL DAN PEMBAHASAN

Pengujian Korelasi

Uji korelasi dilakukan untuk mengukur dan memvisualisasikan hubungan linier antara variabel-variabel dalam *dataset*. Nilai korelasi berkisar dari -1 hingga 1, di mana -1 menunjukkan korelasi negatif sempurna, 1 menunjukkan korelasi positif sempurna, dan 0 berarti tidak ada korelasi. Dengan menggunakan *sns.heatmap()* dari Seaborn, matriks korelasi ini divisualisasikan dalam bentuk *heatmap* yang mudah dibaca. Setiap sel pada *heatmap* menunjukkan nilai korelasi antara dua variabel, dengan warna yang mewakili kekuatan korelasi tersebut. Warna yang lebih gelap atau lebih cerah menunjukkan nilai korelasi yang lebih ekstrem, sedangkan warna yang lebih netral menunjukkan korelasi yang lebih lemah. *Heatmap* korelasi antara variabel ditunjukkan pada Gambar 12.



Gambar 12. Heatmap Uji Korelasi

Dapat dilihat bahwa variabel *year* (tahun) yaitu 0.59 menunjukkan korelasi positif yang cukup signifikan antara tahun mobil dan harga. Ini berarti bahwa mobil yang lebih baru cenderung memiliki harga yang lebih tinggi. Korelasi ini dikategorikan sebagai korelasi positif sedang hingga kuat.

Untuk variabel *mileage* (jarak tempuh) yaitu -0.54 menunjukkan korelasi negatif yang moderat antara jarak tempuh dan harga. Ini berarti bahwa mobil dengan jarak tempuh yang lebih tinggi biasanya memiliki harga yang lebih rendah. Korelasi ini dianggap sebagai korelasi negatif sedang.

Selanjutnya variabel *tax* (pajak) yaitu 0.36 menunjukkan korelasi positif yang relatif lemah antara pajak mobil dan harga. Artinya, ada hubungan bahwa pajak yang lebih tinggi akan sedikit berkaitan dengan harga yang lebih tinggi, tetapi pengaruhnya tidak terlalu besar.

Berikutnya variabel *mpg* (konsumsi bahan bakar) yaitu -0.60 menunjukkan korelasi negatif yang kuat antara konsumsi bahan bakar dan harga. Ini berarti bahwa mobil dengan konsumsi bahan bakar yang lebih rendah (lebih efisien) cenderung memiliki harga yang lebih tinggi.

Terakhir untuk variabel *engine size* (ukuran mesin) yaitu 0.59 menunjukkan korelasi positif yang signifikan antara ukuran mesin dan harga. Artinya, mobil dengan ukuran mesin yang lebih besar cenderung memiliki harga yang lebih tinggi. Korelasi ini juga dikategorikan sebagai korelasi positif sedang hingga kuat.

Pengujian Web Deployment dan Hasil Prediksi

Pengujian *web deployment* untuk aplikasi prediksi mobil menunjukkan bahwa aplikasi telah berhasil di-*deploy* pada Streamlit Share dan berfungsi sesuai dengan tujuan yang diharapkan. Tujuannya adalah supaya tidak ditemukan *error* kritis selama pengujian, dan aplikasi siap untuk digunakan oleh pengguna akhir. Digunakan metode *functional testing* untuk melakukan uji tahap ini. Metode *functional testing* yaitu metode yang memastikan bahwa semua fitur dan fungsi aplikasi bekerja sesuai dengan spesifikasi dan kebutuhan pengguna.

Pengujian ini dimulai dengan identifikasi dan verifikasi seluruh fitur utama aplikasi, seperti kolom *input* data dengan *number input*, memilih data dengan *select box*, dan tampilan hasil prediksi. Pengujian fungsional dimulai dengan menguji kemampuan aplikasi dalam menerima dan memproses berbagai jenis *input* dari pengguna, seperti memasukkan angka tahun jarak tempuh, pajak, konsumsi bahan bakar, dan ukuran mesin. Serta dalam memilih data, seperti kolom model mobil, tahun, transmisi, dan jenis bahan bakar. Setiap kombinasi *input* diuji untuk memastikan bahwa aplikasi menghasilkan *output* yang akurat dan relevan sesuai dengan data yang dimasukkan.

Selain itu, pengujian juga meliputi verifikasi alur kerja aplikasi, mulai dari halaman awal hingga tampilan hasil prediksi, untuk memastikan bahwa tidak ada kesalahan navigasi atau fungsi yang terganggu. Pengujian ini bertujuan untuk memastikan bahwa semua fungsi aplikasi berjalan sesuai dengan desain dan spesifikasi yang diharapkan, serta memberikan pengalaman pengguna yang lancar dan bebas dari kesalahan. Pada Gambar 13 dilampirkan tampilan awal aplikasi yang sudah berhasil di-*running*.



Gambar 13. Tampilan Awal Aplikasi

Hasil yang didapat adalah tidak ditemukan *error* pada seluruh rangkaian pada tampilan *web* aplikasi prediksi. Pengguna dapat menggunakan fitur-fitur *input* dengan baik. Tombol “Prediksi Harga Mobil” juga responsif dan sesuai dengan hasil yang sudah dirancang. Model memberikan hasil prediksi yang cocok dengan nilai *input* yang dimasukkan oleh pengguna. Keberhasilan pengujian ini dapat dilihat pada Gambar 14. Aplikasi prediksi harga mobil Audi ini dapat diakses oleh pengguna pada *link* berikut <https://app-projectabcd-2sqk9zqhdahaah26a7ymyz.streamlit.app/>.



Gambar 14. Tampilan Aplikasi Prediksi

KESIMPULAN DAN SARAN

Program ini berhasil mengintegrasikan berbagai tahapan dalam pengembangan aplikasi prediksi harga mobil bekas Audi, mulai dari pengumpulan data, analisis, pemrosesan data, pembangunan model, evaluasi, hingga *deployment* aplikasi *web*. *Web* aplikasi juga dapat diakses oleh pengguna umum untuk memudahkan dalam mencari prediksi harga mobil Audi bekas.

Hasil penelitian menunjukkan nilai *R2 score* sebesar 0,94 atau 94% dan nilai *MAE* sebesar 0,08 atau 8%. Hal ini menunjukkan bahwa model *linear regression*

yang digunakan dalam program memiliki tingkat keakuratan yang tinggi. R^2 score mendekati 1 artinya berhasil menjelaskan variasi dalam data dengan baik, sementara nilai MAE yang rendah menunjukkan bahwa prediksi harga mobil bekas Audi cenderung memiliki kesalahan yang minim dalam rata-rata nilainya.

Saran penelitian adalah ide atau rekomendasi tentang topik dan arah penelitian yang dapat dieksplorasi lebih lanjut oleh peneliti. Rekomendasi ini muncul setelah meninjau literatur yang ada dan mengidentifikasi celah pengetahuan atau aspek yang belum tercakup dengan baik. Penelitian ini hanya menggunakan MAE dan R^2 Score untuk menghitung prediksi, oleh sebab itu penelitian selanjutnya diharapkan menambah komponen prediksi lain untuk memaksimalkan hasil dari prediksi agar lebih akurat.

DAFTAR PUSTAKA

- [1] T. Muhayat, Jayanta, and N. Chamidah, "Prediksi Harga Smartphone Menggunakan Algoritma Multiple Linear Regression," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, Vol. 296, pp. 506-525, 2022.
- [2] M. A. Saputra, Martanto, and U. Hayati, "Estimasi Harga Mobil Bekas Toyota Yaris Menggunakan Algoritma Regresi Linier," *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 8, No. 2, pp. 1696-1701, 2024.
- [3] Tempus, "Why Audi is and always has been a top-quality manufacturer", *Vantage Media Group*, 2021, [Online]. Available: <https://tempusmagazine.co.uk/news/why-audi-is-and-always-has-been-a-top-quality-manufacturer/>. [Diakses: 24 Juni 2024].
- [4] I. N. Simbolon, H. D. S. J. Siburian, and W. A. Manik, "Prediksi Kualitas Air Sungai di Jakarta Menggunakan KNN yang Dioptimalisasi Dengan PSO". *Jurnal Informatika dan Teknik Elektro Terapan*, Vol.12, No. 2, pp. 1193-1203, 2024.
- [5] R. G. Wardhana, G. Wang, and F. Sibuea, "Penerapan Machine Learning dalam Prediksi Tingkat Kasus Penyakit di Indonesia," *Journal of Information System Management (JOISM)*, Vol. 5, No. 1, pp. 40-4, 2023.
- [6] G. A. Syafarina and Zaenuddin, "Implementasi Framework Streamlit Sebagai Prediksi Harga Jual Rumah Dengan Linear Regresi," *Metik Jurnal*, Vol. 7, No. 2, pp. 121-125, 2023.
- [7] E. Hasibuan and A. Karim, "Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web," *Jurnal Ilmiah Komputasi*, Vol. 21, No. 4, pp. 595-602, 2022.
- [8] A. Amalia, M. Radhi, S. H. Sinurat, D. R. H. Sitompul, and E. Indra, "Prediksi Harga Mobil Menggunakan Algoritma Regresi dengan Hyper-Parameter Tuning," *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, Vol. 4, No. 2, pp. 28-32, 2021.
- [9] D. M. Huda, G. Dwilestari, A. R. Rinaldi, and I. Solihin, "Prediksi Harga Mobil Bekas Menggunakan Algoritma Regresi Linear Berganda", *Jurnal Informatika dan Rekayasa Perangkat Lunak*, Vol. 6, No. 1, pp. 150-157, 2024.
- [10] Sulaiman and E. S. Negara, "Komparasi Algoritma K-Nearest Neighbors dan Random Forest Pada Prediksi Harga Mobil Bekas", *Jurnal Penelitian Ilmu dan Teknologi Komputer (JUPITER)*, Vol. 15, No. 1, pp. 337-346, 2023.
- [11] Reynaldi, W. Syafrizal, and M. F. A. Hakim, "Analisis Perbandingan Akurasi Metode Fuzzy Tsukamoto dan Fuzzy Sugeno Dalam Prediksi Penentuan Harga Mobil Bekas", *Indonesian Journal of Mathematics and Natural Sciences*, Vol. 44, No. 2, pp. 73-80, 2021.
- [12] N. Narwanto, Kusri, and H. A. Fatta, "Prediksi Peserta Matakuliah Menggunakan Artificial Neural Network –Fuzzy Inferented System(Ann-Fis) Studi Kasus: Universitas Muhammadiyah Surakarta," *Journal of Technology and Informatic (JoTI)*, Vol. 1, No. 2, pp. 84-88, 2020.
- [13] A. D. Siburian, D. R. H. Sitompul, S. H. Sinurat, A. Situmorang, Ruben, D. J. Ziegel, and E. Indra, "Laptop Price Prediction with Machine Learning Using Regression Algorithm," *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, Vol. 6, No. 1, pp. 87-91, 2022.
- [14] Aditya, "100,000 UK Used Car Data set", *Kaggle*, 2020, [Online]. Available: <https://www.kaggle.com/datasets/adityadesai13/use-d-car-dataset-ford-and-mercedes>. [Diakses: 21 Maret 2024].
- [15] D. A. Rhamadhani and E. E. D. Saputri, "Analisa Model Machine Learning dalam Memprediksi Laju Produksi Sumur Migas 15/9-F-14H," *Journal of Sustainable Energy Development*, Vol. 1, No. 1, pp. 48-55, 2023.
- [16] N. L. Azizah, N. Suarna, W. Prihartono, "Prediksi Tingkat Kemiskinan di Provinsi Jawa Barat Menggunakan Algoritma Regresi Linear," *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 7, No. 6, pp. 3377-3381, 2023.