

Penerapan Model Regresi Linier dalam Prediksi Harga Mobil Bekas di India dan Visualisasi dengan Menggunakan Power BI

Muhammad Devanda Hendra Kusuma ^{1*}, Syarif Hidayat ²

^{1,2}Program Studi Sistem Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia, Kabupaten Sleman, Daerah Istimewa Yogyakarta, Indonesia.

Email: devandakusuma20@gmail.com ^{1*}, syarif@uii.ac.id ²

Histori Artikel:

Dikirim 19 November 2023; *Diterima dalam bentuk revisi* 5 Desember 2023; *Diterima* 20 Januari 2024; *Diterbitkan* 10 Mei 2024. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Pesatnya pertumbuhan industri otomotif global telah menyebabkan peningkatan penjualan mobil bekas, khususnya di pasar perilaku konsumen yang unik seperti India. Penentuan harga mobil bekas menghadirkan tantangan karena faktor yang tidak konsisten seperti jarak tempuh dan usia. Studi ini bertujuan untuk berkontribusi pada pemahaman pasar mobil bekas di India dan memberikan informasi berharga bagi pemilik bisnis dan pelaku industri untuk membuat keputusan yang tepat. Hasilnya menunjukkan bahwa regresi linier secara akurat dan membentuk model regresi linier berganda $y = -1998.21 + 9.9 X_1 + 1.22 X_2 + 1.3 X_3 - 8.07 X_4 + 1.83 X_5 + 2.89 X_6$ dengan Coefficient of determination 0.6360285441938984, MAE 3.9962839869970264, RMSE 7.015889267114422 dan MSE 49.222702208411334. Selain itu pada proses pengujian model yang dibangun memiliki tingkat akurasi sebesar 71.09%. Penelitian ini bertujuan melakukan prediksi harga mobil bekas dengan mempertimbangkan faktor-faktor yang memengaruhi, dan hasilnya kemudian disajikan secara visual untuk memberikan pemilik bisnis informasi yang lebih interaktif.

Kata Kunci: Mobil bekas; India; machine learning; regresi linier; interaktif.

Abstract

The growing message of the global automotive industry has led to an increase in used car sales, especially in unique consumer behavior markets like India. Pricing used cars presents challenges due to inconsistent factors such as mileage and age. This study aims to contribute to the understanding of the used car market in India and provide valuable information for business owners and industry players to make informed decisions. The results show that linear regression is accurate and forms a multiple linear regression model $y = -1998.21 + 9.9 X_1 + 1.22 X_2 + 1.3 X_3 - 8.07 X_4 + 1.83 X_5 + 2.89 X_6$ and UMK 49.222702208411334. Apart from that, in the testing process the model built had an accuracy level of 71.09%. This research aims to predict used car prices by considering influencing factors, and the results will be presented visually to provide business owners with more interactive information.

Keyword: Used car; India; machine learning; linear regression; interactive.

1. Pendahuluan

Perkembangan industri otomotif global berkembang dengan cepat mengikuti kemajuan zaman dan teknologi. Semakin banyaknya model mobil baru yang diperkenalkan ke pasar tidak hanya memenuhi kebutuhan masyarakat, namun juga berdampak positif pada penjualan mobil bekas. Beragam inovasi pada desain maupun fitur yang dimiliki mobil baru membuat harga mobil bekas menjadi lebih terjangkau dan digemari masyarakat. Dengan adanya kondisi tersebut membuat pasar mobil bekas semakin menarik bagi para pengusaha yang ingin memanfaatkan peluang bisnis di bidang jual beli mobil bekas. Adapun salah satu negara yang menjadi sorotan dalam permasalahan ini adalah India. Dengan menjadi salah satu pasar mobil terbesar di dunia, India menunjukkan keunikan tersendiri dalam perilaku konsumen terhadap mobil bekas. Meningkatnya penjualan mobil bekas di India membuat banyak orang tertarik untuk masuk ke sektor ini, sekaligus menciptakan peluang ekonomi yang terus bertumbuh sejalan dengan pesatnya perkembangan teknologi otomotif di negara tersebut [1].

Menetapkan harga mobil bekas menjadi aspek utama dalam proses jual beli. Informasi yang tidak semua jelas sumbernya di Internet dan faktor yang tidak sepenuhnya pasti dapat menentukan harga mobil bekas. Konsumen beranggapan mengenai produk yang memiliki kualitas yang baik karena dijual ke pasar dengan harga yang relatif tinggi, begitu juga sebaliknya, produk yang dianggap memiliki kualitas yang kurang baik cenderung sulit untuk dijual dengan harga yang tinggi. Penggunaan terminologi "harga" sering dimanfaatkan dalam proses jual beli suatu produk, baik barang maupun jasa. Harga jual ditetapkan oleh penjual dan mendapat keuntungan dari harga tersebut, sedangkan konsumen mencukupi kebutuhannya dengan membayar produk berdasarkan harga yang ditetapkan dan menerima manfaat dari produk maupun jasa yang diterima oleh konsumen [2]. Prediksi harga mobil bekas sangatlah penting untuk menunjang proses jual beli. Prediksi adalah proses sistematis dalam menjabarkan apa yang dapat berlangsung di masa depan dengan mengacu berdasarkan informasi terdahulu dan saat ini. Dalam kasus mobil bekas, prediksi harga menjadi sebuah solusi atas ketimpangan dalam menetapkan harga jual. Berbagai faktor seperti jumlah jarak tempuh hingga usia kendaraan dapat berdampak cukup signifikan terhadap penentuan harga jual [3].

Ethem Alpaydin dalam [4] menyebutkan bahwa salah satu metode komputasi yang dapat digunakan dalam memprediksi harga mobil dengan memanfaatkan *machine learning* (ML). ML sendiri memiliki pengertian sebagai metode yang menjadikan mesin atau komputer dapat belajar melalui pengalaman atau pemrograman sehingga mesin dapat belajar. ML menggunakan data untuk belajar, sehingga dapat juga disebut pembelajaran dengan data [4]. Metode pembelajaran ML dapat diimplementasikan dengan menerapkan berbagai macam algoritma pembelajaran. Secara umum, ada tiga jenis metode pembelajaran yang digunakan, yaitu supervised learning, unsupervised learning, dan reinforcement learning. Supervised learning yaitu metode pembelajaran yang diawasi. Program nantinya dapat menerima sejumlah sampel data dengan tipe/klasifikasi tertentu yang biasa disebut dengan data training dan data testing. Program bisa saja salah mengklasifikasikan suatu objek setelah melakukan pelatihan. Jadi, selain menggunakan data training set, program ini juga memerlukan data testing. Dengan demikian, tingkat keberhasilan akan dapat diukur [4].

Unsupervised learning adalah metode pembelajaran tanpa pengawasan. Metode ini menggunakan proses yang mencoba menemukan partisi dalam sebuah model. Unsupervised learning mempelajari bagaimana sebuah model dapat belajar melakukan representasi pola input dengan cara menggambarkan struktur statistik dari total keseluruhan pola input. Tidak seperti supervised learning, metode ini tidak mempunyai tujuan output yang jelas dan oleh karena itu sering dimanfaatkan untuk tujuan klasifikasi [4]. *Reinforcement learning* adalah salah satu metode pembelajaran yang menerapkan pendekatan untuk mempelajari aturan kontrol dengan cara berinteraksi dengan lingkungan yang tidak dikenal. Program akan diberi hukuman jika membuat keputusan yang salah dan memberi penghargaan jika benar. Seluruh pengalaman interaksi diakumulasikan hingga program dapat menarik kesimpulan di masa depan dengan memakai pola-pola yang telah dipelajari [4]. Dalam hal ini, untuk memecahkan kompleksitas pada penentuan harga mobil bekas, dibutuhkan hubungan antar variabel atau kolom.

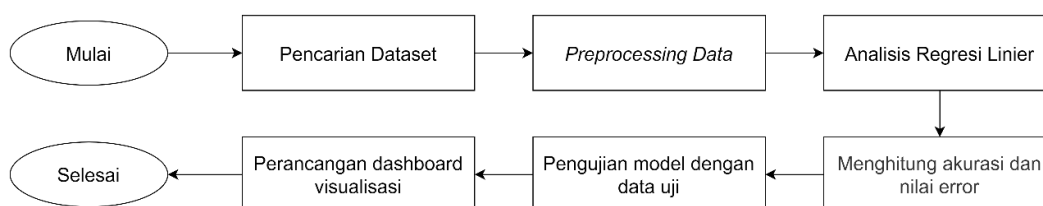
Karenanya, algoritma yang diterapkan untuk memprediksi harga mobil yaitu algoritma regresi linier. Regresi linier telah terbukti mampu memberikan solusi secara efektif. Regresi linier dilakukan untuk menghasilkan model linier dari dua variabel atau lebih untuk mengetahui hubungan antara berbagai variabel tersebut [5]. Evaluasi model regresi linier dapat melalui metrik seperti MAE, RMSE, MSE dan R-Squared, yang menyediakan gambaran secara tepat mengenai seberapa baik kinerja model tersebut.

Penelitian yang dilakukan oleh Leriansyah Putra Nasyuli, Imran Lubis, dan Andi Marwan Elhanafi membahas tentang prediksi harga kendaraan bekas dengan memanfaatkan model machine learning algoritma gradient boosting dan linear regression. Dalam penelitian ini, Proses prediksi melibatkan langkah-langkah membangun sistem, inialisasi hyperparameter, training gradient boosting, dan uji data testing. Selain itu, penelitian ini juga menghasilkan formula matematis yang digunakan dalam proses prediksi harga kendaraan bekas, memberikan kontribusi penting dalam pengembangan metode prediksi harga kendaraan bekas menggunakan model machine learning [6]. Penelitian lain yang membahas tentang Prediksi Harga Mobil Menggunakan Algoritma Regresi dengan *Hyper-Parameter Tuning* yang dilakukan oleh [7] mendapatkan tingkat nilai akurasi model sebesar 97% (setelah *tuning*) untuk melakukan prediksi. Dalam percobaan prediksi, didapat nilai akurasi prediksi sebesar 80% dari 4 percobaan yang telah dilakukan. Untuk mempresentasikan hasil prediksi dan analisis data, visualisasi dengan memakai Power BI menjadi alat bantu yang sangat efektif. Visualisasi data dapat membuat informasi yang rumit dapat disajikan dengan grafis, sehingga akan lebih mudah dalam melihat tren harga mobil bekas. Power BI, dengan keunggulannya dalam menghasilkan visualisasi yang interaktif, yang nantinya digunakan untuk mempermudah penyajian hasil penelitian kepada para pemangku kepentingan, termasuk para pemilik bisnis mobil bekas [8].

Negara India di pilih untuk menjadi data sumber penelitian oleh penelitian dengan tujuan untuk meningkatkan pemahaman mengenai pasar mobil bekas pada negara tersebut. Pertumbuhan ekonomi India yang cukup pesat, didukung dengan pertumbuhan di sektor teknologi informasi, menjadikan pasar mobil bekas sebagai pasar yang menarik. Diharapkan informasi yang dihasilkan dari penelitian ini dapat memudahkan para pemilik bisnis dan pelaku industri otomotif di India dalam mengambil keputusan yang tepat yaitu menetapkan harga mobil bekas.

2. Metode Penelitian

Diagram alir penelitian “Penerapan Model Regresi Linier Dalam Prediksi Harga Mobil Bekas di India dan Visualisasi dengan Menggunakan Power BI”, ditunjukkan melalui diagram berikut ini.



Gambar 1. Diagram Alir Penelitian

2.1 Dataset

Penelitian ini memakai data sekunder, berupa data yang diambil dari situs penyedia data www.kaggle.com. Berikut adalah link dari dataset yang diambil dari situs penyedia data Kaggle.com (<https://www.kaggle.com/code/iabhishekmaurya/used-car-price-prediction>) diakses pada tanggal 26 Februari 2023. Data yang dipakai yaitu data daftar mobil bekas di India yang memuat data pelatihan dan data pengujian. Data pelatihan memiliki 6.019 baris dan 14 kolom, yang mencakup periode waktu dari tahun 1998 hingga 2019. Atribut dari dataset ini mencakup nomor, nama mobil, lokasi, tahun, kilometer mobil, jenis bahan bakar, transmisi, tipe pemilik, konsumsi bbm, kapasitas mesin, tenaga,

jumlah kursi, harga baru, dan harga bekas. Sedangkan untuk data pengujian terdapat 1234 baris dan 13 kolom, yang mencakup periode dari tahun 1996 hingga 2019. Atribut dari dataset terdiri atas nomor, nama mobil, lokasi, tahun, kilometer mobil, jenis bahan bakar, transmisi, tipe pemilik, konsumsi bbm, kapasitas mesin, tenaga, jumlah kursi, dan harga bekas. Di bawah ini adalah beberapa sampel data dari setiap data pelatihan dan data pengujian.

Unnamed: 0	Nama Mobil	Lokasi	Tahun	Kilometer Mobil	Jenis Bahan Bakar	Transmisi	Tipe Pemilik	Konsumsi BBM	Kapasitas Mesin	Tenaga	Jumlah Kursi	Harga Baru
0	Maruti Alto K10 LXI CNG	Delhi	2014	40929	CNG	Manual	First	32.26 km/kg	998 CC	58.2 bhp	4.0	NaN
1	Maruti Alto 800 2016-2019 LXI	Coimbatore	2013	54493	Petrol	Manual	Second	24.7 kmpl	796 CC	47.3 bhp	5.0	NaN
2	Toyota Innova Crysta Touring Sport 2.4 MT	Mumbai	2017	34000	Diesel	Manual	First	13.68 kmpl	2393 CC	147.8 bhp	7.0	25.27 Lakh
3	Toyota Etios Liva GD	Hyderabad	2012	139000	Diesel	Manual	First	23.59 kmpl	1364 CC	null bhp	5.0	NaN
4	Hyundai i20 Magna	Mumbai	2014	29000	Petrol	Manual	First	18.5 kmpl	1197 CC	82.85 bhp	5.0	NaN

Gambar 2. Data Pelatihan

Unnamed: 0	Nama Mobil	Lokasi	Tahun	Kilometer Mobil	Jenis Bahan Bakar	Transmisi	Tipe Pemilik	Konsumsi BBM	Kapasitas Mesin	Tenaga	Jumlah Kursi	Harga Baru
0	Maruti Alto K10 LXI CNG	Delhi	2014	40929	CNG	Manual	First	32.26 km/kg	998 CC	58.2 bhp	4.0	NaN
1	Maruti Alto 800 2016-2019 LXI	Coimbatore	2013	54493	Petrol	Manual	Second	24.7 kmpl	796 CC	47.3 bhp	5.0	NaN
2	Toyota Innova Crysta Touring Sport 2.4 MT	Mumbai	2017	34000	Diesel	Manual	First	13.68 kmpl	2393 CC	147.8 bhp	7.0	25.27 Lakh
3	Toyota Etios Liva GD	Hyderabad	2012	139000	Diesel	Manual	First	23.59 kmpl	1364 CC	null bhp	5.0	NaN
4	Hyundai i20 Magna	Mumbai	2014	29000	Petrol	Manual	First	18.5 kmpl	1197 CC	82.85 bhp	5.0	NaN

Gambar 3. Data Pengujian

2.2 Preprocessing Data Latih

Preprocessing adalah langkah penting pada pemrosesan data dengan tujuan untuk menjaga kualitas data dan menghindari informasi yang tidak sesuai. Dalam penelitian ini, data awal yang dipakai tentunya masih dalam bentuk mentah. Tahapan *preprocessing* tersebut dalam penelitian ini mencakup diantaranya proses penyesuaian format data sesuai pada kebutuhan, misalnya dalam penelitian ini dengan melakukan transformasi pada kumpulan data. Contoh umum dari transformasi adalah mengubah data menjadi tipe numerik. Hal ini dimaksudkan untuk memastikan bahwa semua input yang diberikan kepada model menjadi data numerik, sehingga membuat sistem semakin mudah dioperasikan. Data non-numerik dihapus selama proses ini, untuk memastikan bahwa model tersebut dapat menerima data masukan yang kompatibel dengan bentuk yang diharapkan. Pentingnya *preprocessing* ditunjukkan dalam kemampuannya membersihkan data, menaikkan konsistensi, dan menjamin bahwa data yang dipakai selama proses analisis atau pelatihan model telah memenuhi standar integritas yang baik.

2.3 Analisis Regresi Linier

Dalam proses ini, data yang telah didapatkan oleh peneliti selanjutnya akan dianalisis dengan menerapkan teknik algoritma regresi linier. Proses ini menjadi langkah penerapan algoritma atau metode yang akan dilakukan peneliti dalam menemukan, mengenali, dan menampilkan pola. Penentuan metode dilakukan dengan melihat tipe data, sehingga dapat diketahui bahwa data cocok diaplikasikan untuk melakukan perkiraan, prediksi, klasifikasi, klusterisasi, hingga pengamatan hubungan asosiasi. Dalam penelitian ini, metode yang sesuai untuk diaplikasikan yaitu regresi linier. Sebab akan dicari perkiraan harga mobil bekas pada dataset yang dimiliki. Dalam sebuah model regresi, biasanya didapatkan nilai positif dan negatif pada angka perkiraan yang semestinya memiliki pengaruh signifikan berdasarkan beberapa faktor tertentu. Oleh karena itu, pada tahap ini diperlukan metode machine learning atau statistik agar model dapat diterapkan. Jika telah menetapkan metode, langkah selanjutnya adalah mengaplikasikannya dengan mengacu pada dataset yang telah didapatkan.

Jika nantinya data disesuaikan, langkah ini dapat kembali ke langkah *preprocessing* data.

2.4 Menghitung Akurasi dan Nilai Error

Pada fase ini, performa model dievaluasi dengan mengukur akurasi. Pengujian model regresi menggunakan 1234 data pengujian dengan mengevaluasi metrik kesalahan, yang meliputi perhitungan MSE (*Mean Squared Error*), MAE (*Mean Absolute Error*), RMSE (*Mean Squared Error*), dan R Square. Hasil prediksi yang sesuai dengan data sebenarnya ditandai dengan nilai MSE yang rendah atau mendekati nol. Dalam konteks model regresi linier, evaluasi menggunakan RMSE menjadi penting, dengan nilai yang lebih kecil menunjukkan akurasi yang lebih tinggi. Untuk menghitung RMSE, kuadratkan kesalahan produksi, jumlahkan, dan ambil akar kuadrat. R-squared, yang berkisar dari 0 hingga 1, mencerminkan seberapa cocok model tersebut dengan kumpulan data. Nilai yang lebih tinggi menunjukkan kecocokan yang lebih baik. Di sisi lain, MAE memberikan indikasi performa model dalam prediksi. Nilai kecil menunjukkan kinerja yang baik, dan nilai besar menunjukkan masalah prediksi. Ketika nilai MAE mencapai 0, hal ini mencerminkan bahwa model tersebut merupakan prediktor yang sempurna [9].

2.5 Pengujian Model Data Uji

Pada tahap ini, model diuji untuk mengevaluasi keakuratan dan kinerjanya. Peringkat kinerja tercermin dalam nilai akurasi model, diukur menggunakan data pengujian yang terdiri dari 1234 baris dan 13 kolom, mencakup rentang waktu 1996 hingga 2019. Proses pengujian ini berfungsi sebagai alat validasi terhadap model yang dikembangkan. Perlu diperhatikan bahwa akurasi model dapat ditingkatkan dengan mengoptimalkan rasio data pengujian dan data pelatihan. Akurasi digunakan sebagai cara sederhana untuk mengevaluasi kinerja model. Namun, ada situasi di mana akurasi saja tidak sepenuhnya mencerminkan performa model, terutama ketika keseimbangan antar kelas target tidak merata. Jika model sudah terbukti efektif dan akurat maka proses dapat dilanjutkan ke tahap berikutnya dan model yang dibuat dapat segera digunakan. Regresi linier berganda merupakan teknik analisis regresi yang melibatkan banyak variabel, sehingga memberikan pemahaman yang lebih mendalam mengenai hubungan antar variabel dan memperkaya interpretasi hasil regresi.

2.6 Perancangan Dashboard

Dalam penelitian ini, visualisasi data menggunakan Power BI, platform analisis bisnis dari Microsoft. Power BI bertanggung jawab untuk mengintegrasikan, menganalisis, dan mendeskripsikan data dari berbagai sumber dengan cara yang interaktif dan ramah pengguna. Platform ini menyediakan alat untuk transformasi data, analisis visual, dan berbagi wawasan bisnis. Fitur utama Power BI mencakup beberapa aspek penting. Pertama, pengguna dapat dengan mudah menghubungkan dan menggabungkan data dari berbagai sumber, termasuk database relasional, file Excel, data cloud, dan layanan web. Kedua, alat transformasi data Power BI untuk membersihkan dan mengatur data, termasuk menghapus kolom yang tidak diperlukan, mengganti nilai yang hilang, dan menggabungkan data dari beberapa tabel.

Selain itu, Power BI dapat membangun model data yang kompleks dengan menggabungkan tabel, membuat hubungan antar tabel, dan membuat kolom terhitung menggunakan formula DAX (*Data Analysis Expressions*). Keempat, visualisasi Power BI dapat difilter secara interaktif, sehingga memudahkan eksplorasi data yang lebih dalam sekaligus mendukung perubahan secara real-time melalui interaksi langsung antar visualisasi. Terakhir, Power BI tidak hanya berfungsi sebagai alat visualisasi, namun juga dapat disesuaikan dengan kebutuhan merek dan bisnis Anda serta diintegrasikan ke dalam aplikasi bisnis seperti Microsoft Office 365, SharePoint, dan Azure untuk meningkatkan fungsionalitas dan kolaborasi [10].

3. Hasil dan Pembahasan

3.1 Analisis Regresi Linier

Untuk pemrosesan analisis dengan menggunakan bahasa pemrograman Python, digunakan beragam library seperti Numpy dan Pandas sebagai alat analisis data, serta Matplotlib dan Seaborn sebagai visualisasi data. Scikit-Learn dipilih untuk melakukan machine learning. Perintah untuk memanggil library yang dipakai ditunjukkan pada gambar 4.

```
# Step 2 : import library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_squared_error
sns.set()
```

Gambar 4. Library yang Digunakan

Proses pembuatan model diawali dengan melakukan eksplorasi data yang diperoleh dari dataset yang tersedia di www.kaggle.com, suatu platform yang menyediakan berbagai dataset dalam format CSV. Selain menawarkan lebih dari 6.000 dataset yang dapat dipilih, Kaggle juga sebagai wadah bagi para ilmuwan data terbesar di dunia. Dataset Kaggle banyak digunakan untuk penelitian ilmiah di seluruh dunia dalam mengembangkan model. Penggunaan bahasa pemrograman Python menjadi pilihan utama pada tahap analisis. Menurut Srinath dalam [4], Python memiliki sejumlah karakteristik sebagai berikut: 1) memiliki distribusi Python yang kaya akan pustaka dan modul; 2) memiliki sintaks yang jelas dan mudah dipelajari, 3) Aturan tata letak kode untuk memudahkan peninjauan, 4) Membaca dan menulis ulang kode sumber, 5) Berorientasi pada objek, 6) Dapat ditulis dalam bahasa Python atau C/C++.

Dataset yang didapatkan mengalami sejumlah proses seperti penghilangan nilai missing value (NaN). Selain melakukan proses preprocessing, juga dilakukan proses transformasi data sebagai berikut: Mengganti variabel untuk sejumlah atribut, mengubah satuan untuk variabel Konsumsi BBM, Mengganti variabel untuk sejumlah atribut, mengubah satuan untuk variabel penghematan bahan bakar, menghapus nilai satuan untuk sejumlah variabel, mengganti nama variabel untuk sejumlah atribut, membuat fitur baru dengan menggunakan rekayasa fitur, serta dilakukan penyesuaian tipe data untuk seluruh variabel. Sejumlah langkah tersebut dilakukan bertujuan agar model mampu membaca data dengan kualitas tinggi dan juga memaksimalkan hasil dan akurasi yang diperoleh. Output data yang sudah melalui tahap preprocessing dan transformasi akan dilanjutkan ke tahap selanjutnya yaitu pelabelan data dan penanganan data outlier.

Pelabelan data ini perlu diperhatikan karena terdapat data kategorikal pada variabel seperti nama Mobil, Lokasi, Jenis Bahan Bakar, Transmisi, dan Merk yang perlu dikonversi ke dalam bentuk angka dengan skala nominal. Pengkonversian ini tidak termasuk variabel Nama Mobil dan Merk sehingga variabel tersebut tidak mengalami perubahan. Data kategorikal yang diubah menjadi numerik tidak akan berpengaruh terhadap apapun secara matematis, sebagai permisalan contoh data kategorikal yang dapat diubah menjadi numerik adalah tinggi badan, tanggal lahir, olahraga favorit. Pelabelan ini dibutuhkan dalam machine learning karena mesin hanya sanggup mengolah data numerik, selain itu terdapat juga beberapa manfaat diantaranya dapat meningkatkan kualitas serta akurasi pada model yang akan dibuat. Metode pelabelan data kategorik (*label encoder*) dilaksanakan dengan menggunakan metode *one-hot encoding* yang cocok untuk data yang bersifat nominal dan variabel dengan jumlah kategori yang sedikit. Hasil pelabelan data ditampilkan pada gambar 5.

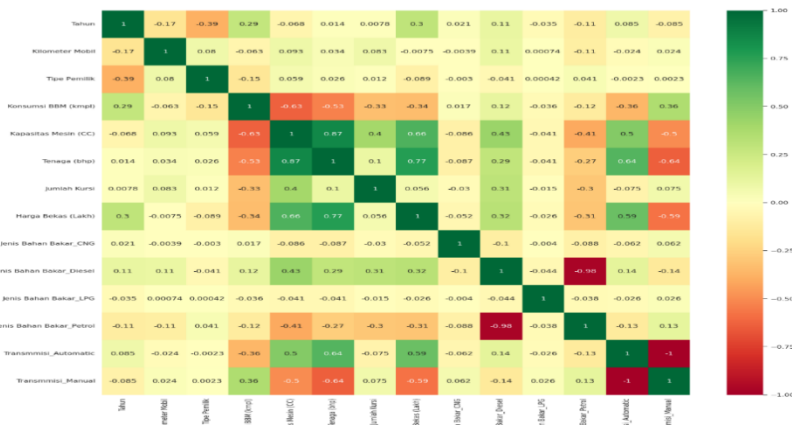
...	Lokasi_Kochi	Lokasi_Kolkata	Lokasi_Mumbai	Lokasi_Pune	Jenis Bahan Bakar_CNG	Jenis Bahan Bakar_Diesel	Jenis Bahan Bakar_LPG	Jenis Bahan Bakar_Petrol	Transmmisi_Automatic	Transmmisi_Manual
...	0	0	1	0	1	0	0	0	0	1
...	0	0	0	1	0	1	0	0	0	1
...	0	0	0	0	0	0	0	1	0	1

Gambar 5. Pelabelan Data Kategorik

Gambar 5 menampilkan output dari pelabelan data. Di dalamnya, masing-masing variabel kategorikal baru terbentuk menjadi sebuah kolom baru. Proses pengkodean one-hot encoding menciptakan kolom baru pada setiap kategori. Nilai 0 berarti kategori tersebut tidak ada, nilai 1 berarti kategori tersebut ada. Sebagai contoh, variabel lokasi dibagi menjadi 10 kategori, variabel jenis dibagi menjadi 4 kategori, dan variabel transmisi dibagi menjadi 2 kategori. Terdapat sebanyak 16 variabel independen baru yaitu Lokasi_Bangalore, Lokasi_Chennai, Lokasi_Coimbatore, Lokasi_Delhi, Lokasi_Hyderabad, Lokasi_Jaipur, Lokasi_Kochi, Lokasi_Kolkata, Lokasi_Mumbai, Lokasi_Pune, Jenis Bahan Bakar_CNG, Jenis Bahan Bakar_Diesel, Jenis Bahan Bakar_LPG, Jenis Bahan Bakar_Petrol, Transmmisi_Automatic, dan Transmmisi_Manual. Data outlier adalah data yang menyimpang dari pola umum data lainnya. Data outlier dapat menyebabkan distorsi dalam analisis dan mengurangi akurasi model. Oleh karena itu, data outlier perlu diidentifikasi dan dihapus sebelum dilakukan analisis. Dalam kasus ini, data outlier diidentifikasi berdasarkan data dengan rentang harga yang berada di bawah 25% dan di atas 75%. Data-data tersebut kemudian dihapus untuk memastikan kejelasan dan akurasi analisis.

Data outlier merupakan data yang berbeda secara signifikan dari kisaran normal dalam rangkaian data. Menghapus data outlier menjadikan analisis data lebih jelas dan meningkatkan persentase akurasi data. Dalam proses ini, data dengan rentang harga di bawah 25% dan di atas 75% akan dihilangkan karena akan berpengaruh terhadap tingkat akurasi data pada saat pengolahan data. Langkah selanjutnya adalah memilih variabel dengan tingkat korelasi yang tinggi seperti yang terlihat pada gambar 6, yang menunjukkan korelasi masing-masing variabel. Variabel Tahun, Kapasitas Mesin (cc), Tenaga (bhp), Jumlah Kursi, Jenis Bahan Bakar Diesel, dan Transmisi_Automatic menampilkan tingkat akurasi yang tinggi berdasarkan variabel harga dependen dengan nilai 30%, 66%, 77%, 5,6%, 32%, dan 59%. Variabel Kilometer Mobil, Tipe Pemilik, Konsumsi BBM (kmpl), Jenis Bahan Bakar_CNG, Jenis Bahan Bakar_LPG, Jenis Bahan Bakar_Petrol, dan Transmisi_Manual memiliki tingkat akurasi yang rendah.

Pemilihan variabel didasarkan pada tingkat korelasi terhadap variabel harga dependen. Variabel dengan korelasi tinggi dianggap signifikan dan relevan untuk dimasukkan dalam model. Oleh karena itu, variabel seperti Tahun, Kapasitas Mesin (cc), Tenaga (bhp), Jumlah Kursi, Jenis Bahan Bakar Diesel, dan Transmisi_Automatic dipilih dan akan digunakan dalam tahap implementasi algoritma. Dalam konteks ini, matriks korelasi digunakan untuk mengevaluasi sejauh mana variabel-variabel lain berkaitan dengan variabel harga dependen. Korelasi yang tinggi dapat menunjukkan hubungan yang kuat, meskipun tidak selalu berarti sebab akibat. Menyertakan variabel-variabel ini dalam model diharapkan meningkatkan akurasi prediksi harga mobil, berdasarkan hubungan yang kuat dan signifikan dengan variabel dependen. Hasil ini menjadi dasar kuat untuk pengambilan keputusan dalam analisis regresi, menjamin bahwa model yang dikembangkan dapat memberikan hasil prediksi yang andal.



Gambar 6. Matriks Korelasi

Selanjutnya, setelah proses pelabelan data dan pengolahan *data outlier* selesai dilakukan, proses terakhir dalam analisis regresi linier adalah validasi data. Proses tersebut memvalidasi tingkat akurasi data dan menegaskan kembali bahwa data tersebut bersih dan dapat digunakan untuk pemrosesan serta analisis lebih lanjut guna memperoleh *insight* dan wawasan. Tahap ini juga berfungsi sebagai cara untuk memverifikasi kualitas data itu sendiri, menghindari *overfitting*, menjamin konsistensi, mengatasi *missing value*, dan banyak tujuan lainnya setelah tahap pembersihan data selesai.

Metode penelitian yang digunakan oleh peneliti adalah metode regresi linier berganda. Penerapan metode regresi linier berganda dalam analisis data memiliki beberapa kelebihan yang signifikan. Pertama, metode ini memungkinkan penelitian untuk memahami hubungan kompleks antara variabel terikat dan dua atau lebih variabel bebas. Dengan melibatkan multiple variabel bebas, regresi linier berganda memfasilitasi analisis yang lebih komprehensif, mengidentifikasi pengaruh relatif dari setiap variabel terhadap variabel terikat. Keakuratan prediksi juga menjadi salah satu kelebihan, karena model ini dapat mengakomodasi variasi dari beberapa variabel bebas, meningkatkan ketepatan dalam memprediksi nilai variabel terikat. Selain menggunakan metode tersebut dapat juga dipertimbangkan untuk menggunakan metode *random forest* sebagai metode alternatif yang sangat baik, terutama dalam situasi tertentu seperti menangani non-linearitas, yang memungkinkan penanganan hubungan non-linear antara variabel tanpa memerlukan asumsi bahwa hubungan tersebut bersifat linier. Selain itu, regresi linier berganda dapat mengatasi keterbatasan regresi sederhana, khususnya ketika hubungan antar variabel tidak dapat dijelaskan hanya dengan satu variabel independen. Dengan demikian, metode ini memberikan fleksibilitas yang lebih besar dalam pemodelan hubungan yang kompleks. Analisis multivariabel yang holistik juga menjadi keunggulan, memberikan pemahaman yang lebih mendalam terhadap faktor-faktor yang berkontribusi pada variabel terikat.

Variabel yang digunakan dalam penelitian ini adalah Tahun, Kapasitas Mesin (cc), Tenaga (bhp), Jumlah Kursi, Jenis Bahan Bakar_Diesel, dan Transmisi_Automatic sebagai variabel penentu. Oleh karena itu, model regresi linier pada penelitian ini bisa dinyatakan sebagai persamaan 1:1 dimana y merupakan variabel respon, yaitu Harga Bekas (Lakh). Adapun b0 sebagai intercept, b1 sebagai slope dari variabel x1 yaitu Tahun, b2 sebagai slope dari variabel x2 yaitu Kapasitas Mesin (cc), b3 sebagai slope dari variabel x3 yaitu jenis Tenaga (bhp), b4 sebagai slope dari variabel x4 yaitu Jumlah Kursi, b5 sebagai slope dari variabel x5 yaitu Jenis Bahan Bakar_Diesel, dan b6 sebagai slope dari variabel x6 yaitu Transmisi_Automatic. Nilai b0(intercept), b1 – b6 (slope) dihasilkan melalui perhitungan kuadrat terkecil dengan memanggil fungsi model intercept dari library regresi linier dari modul Sklearn di Python. Seperti yang tampak pada gambar 7 nilai intercept dan nilai slope di bawah ini.

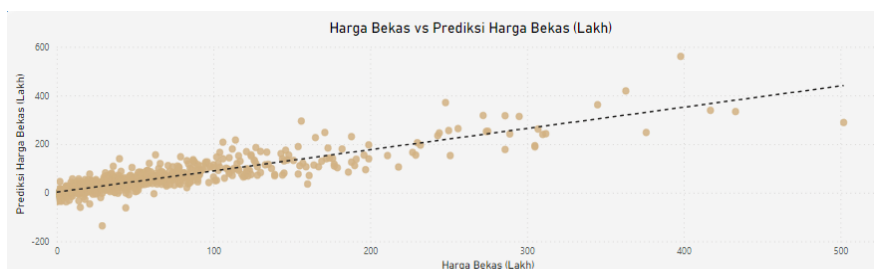
```
[ ] linear_reg.intercept_
print("Intercept : ",linear_reg.intercept_)

Intercept : [-1998.21043556]

[ ] linear_reg.coef_
print("Slope : ",linear_reg.coef_)

Slope : [[ 9.89886845e-01  1.22296690e-03  1.29803456e-01 -8.06157151e-01
 1.83068285e+00  2.88598646e+00]]
```

Gambar 7. Nilai Intercept serta Slope yang Didapatkan



Gambar 8. Grafik Harga Prediksi serta Harga Sebenarnya

Gambar 8 menampilkan grafik komparasi dari data tes yang diprediksi dan aktual. Dari gambar ini, terlihat bahwa nilai prediksi yang dihasilkan oleh model kurang lebih sama dengan nilai aktual, tetapi nilai prediksi yang dihasilkan bisa saja sama atau lebih rendah dari nilai aktual. Hal ini membuat data dapat diperiksa, termasuk menentukan pengaruh variabel independen terhadap variabel dependen menggunakan regresi linier dengan algoritme yang tepat. Keakuratan algoritma ini memberikan model yang lebih baik dengan akurasi yang lebih tinggi, dan nilai prediksi yang lebih mendekati nilai aktual.

Tabel 1. Nilai Prediksi serta Nilai Sebenarnya

Nama Mobil	Harga Bekas Aktual (Lakh)	Harga Prediksi (Lakh)
Maruti Wagon R LXI CNG	1	-3
Hyundai Creta 1.6 CRDi SX Option	12	12
Honda Jazz V	4	1
Maruti Ertiga VDI	6	2
Audi A4 New 2.0 TDI Multitronic	17	15
...
Chevrolet Beat Diesel	2	-1

Tabel 1 menunjukkan nilai prediksi harga mobil bekas atas beberapa variabel independen yang mempengaruhi. Prediksi harga mobil bekas yang ditampilkan dalam model hanya sedikit mengalami deviasi dari nilai aslinya. Seperti contoh data ke 5, prediksi harga mobil yang diperoleh yaitu 15 Lakh, namun harga asli dari mobil tersebut adalah 17 Lakh, dan hal yang sama juga berlaku untuk mobil-mobil lainnya.

Tabel 2. Pengukuran Metrik Kesalahan

MAE	MSE	RMSE	R2
3.9962839869970264	49.222702208411334	7.015889267114422	0.6360285441938984

Setelah memperoleh model, evaluasi model kemudian dilakukan. Tahap ini menentukan keakuratan dari model. Ukuran-ukuran yang digunakan untuk mengevaluasi hasil prediksi suatu model adalah ukuran R squared, Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Square Error (RMSE). Dari Tabel 2, didapatkan nilai RMSE sebesar 7.015889267114422

(dibulatkan menjadi 7,02). Nilai RMSE berguna untuk merepresentasikan tingkat kesalahan pada data model yang digunakan. Semakin rendah nilai RMSE, maka semakin tinggi tingkat akurasi sistem. Kemudian didapatkan nilai MAE sebesar 3.9962839869970264 (dibulatkan menjadi 4). Nilai MAE menghitung rata-rata selisih absolut antara nilai aktual dengan nilai prediksi. Semakin rendah nilai MAE, semakin tinggi kemampuan model untuk memprediksi nilai. Selanjutnya, terdapat nilai MSE sebesar 49.222702208411334 (dibulatkan menjadi 49,2). Nilai MSE ini mengukur rata-rata kesalahan kuadrat antara nilai aktual dan nilai prediksi. Nilai MSE yang rendah atau mendekati nol menunjukkan bahwa hasil prediksi sesuai dengan data aktual. Terakhir, diketahui bahwa nilai R-squared yang dihasilkan oleh model adalah 0.6360285441938984 (dibulatkan menjadi 0.64 atau 64%). Semakin dekat nilai R-squared dengan 1, semakin akurat model yang dihasilkan.

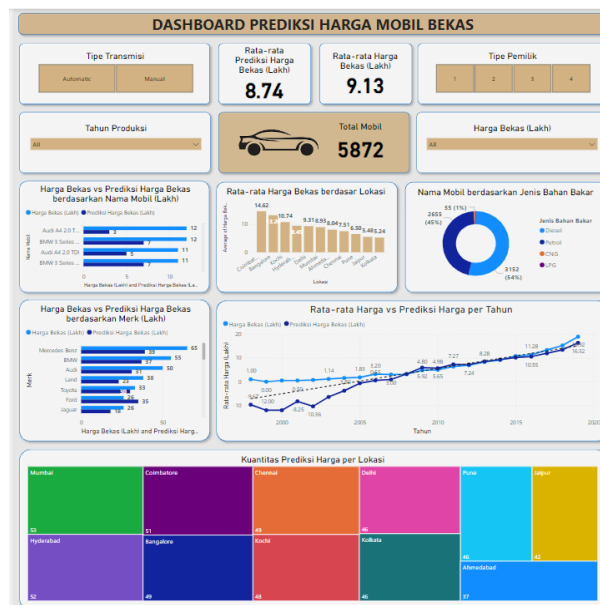
Tabel 3. Pengujian Hasil Prediksi

No	Rasio Data	Akurasi
1	60:40	0.7019326529589567
2	70:30	0.7092304064473374
3	80:20	0.7109418721278501
4	90:10	0.7089342854702261

Seperti yang terlihat pada tabel 3 berdasarkan hasil pengujian prediksi dengan rasio data latih dan data uji yang berbeda, akurasi prediksi tertinggi dihasilkan ketika rasio data uji dan data latih sebesar 80%:20% dengan tingkat akurasi (71.09%).

3.2 Implementasi dengan Visualisasi Dashboard

Bentuk visualisasi dashboard yang terbentuk dalam penelitian ini ditunjukkan dalam Gambar 9.



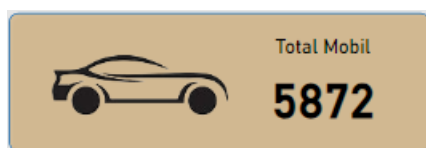
Gambar 9. Tampilan Dashboard

Dashboard yang telah dibuat tersebut digunakan untuk menampilkan beberapa visualisasi yang berisi informasi, yaitu data aktual harga mobil bekas (1998-2019) dan hasil prediksi harga yang telah didapatkan, pemilihan transmisi, tipe pemilik, tahun pembuatan, serta daftar harga bekas yang dapat dimaksimalkan untuk menampilkan informasi yang lebih detail dan interaktif. Selain pilihan-pilihan tersebut, ditampilkan juga beragam informasi seperti rata-rata prediksi harga bekas dan rata-rata harga bekas yang terlihat dalam gambar 10 berikut.



Gambar 10. Rata-rata Harga Prediksi dengan Sebenarnya

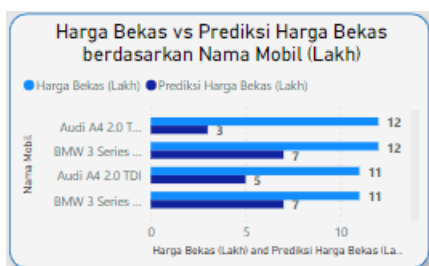
Dalam gambar 10, terlihat tampilan data rata-rata harga mobil bekas prediksi dan aktual secara detail dengan memilih harga dan tahun pembuatan yang ingin ditampilkan datanya. Dengan menampilkan rata-rata harga prediksi dan harga aktual dalam format *card*, pemilik bisnis bisa melihat dengan lebih detail rata-rata harga prediksi dan harga aktual berdasarkan harga serta tahun pembuatan yang diinginkan.



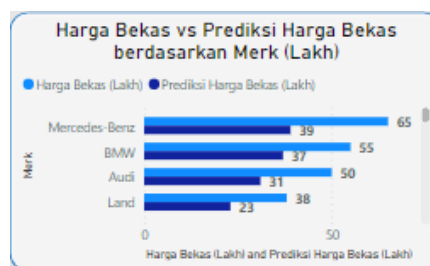
Gambar 11. Total Mobil

Dari gambar 11, menampilkan data total kendaraan yang menunjukkan jumlah total mobil serta dapat melihat secara detail dengan memilih tipe transmisi dan tipe pemilik yang akan disajikan. Dengan menampilkan data jumlah total mobil dalam *card*, pemilik bisnis bisa memeriksa data jumlah total kendaraan yang lebih rinci berdasarkan jenis transmisi dan jenis kepemilikan yang dipilih.

Pada gambar 12, diagram batang komparasi yang disusun secara horizontal adalah alat visual yang efektif untuk menganalisis perbedaan untuk setiap nama mobil antara harga prediksi dan harga sebenarnya. Dengan menggunakan kumpulan data yang telah dikelompokkan berdasarkan nama mobil, bagan ini menggambarkan dua batang untuk setiap nama mobil: Satu bilah mewakili harga prediksi dan bilah lainnya mewakili harga sebenarnya. Warna yang berbeda pada setiap batang membuatnya lebih mudah dipahami, dan label yang jelas pada sumbu dan batang memberikan informasi yang mudah dicerna. Melalui analisis grafik, dimungkinkan untuk melihat pola atau tren perbedaan untuk setiap nama mobil, dengan fokus pada nama mobil yang menunjukkan perbedaan yang signifikan. Grafik ini memberikan wawasan yang mendalam tentang sejauh mana model prediksi berhasil atau tidak untuk setiap nama mobil, sehingga memudahkan identifikasi peningkatan atau penyempurnaan model. Secara keseluruhan, visualisasi ini memberikan gambaran yang jelas dan mudah tentang bagaimana prediksi harga mobil sesuai dengan harga aktual berdasarkan nama mobil. Adapun hasil visual dari komparasi prediksi dan harga aktual mobil bekas untuk setiap merek dapat ditampilkan melalui grafik seperti pada gambar 12.

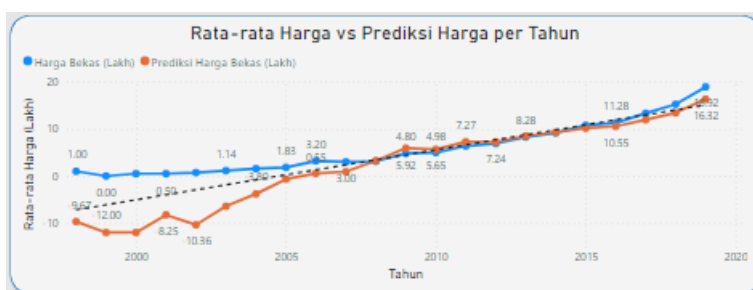


Gambar 12. Komparasi Harga Prediksi dan Harga Aktual setiap Nama Mobil



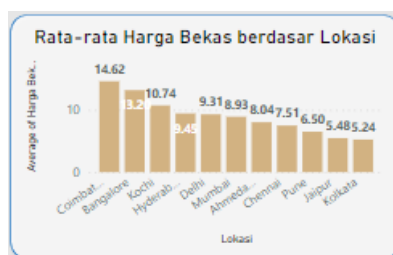
Gambar 13. Komparasi Harga Prediksi dan Harga Aktual setiap Merk

Pada gambar 13, diagram batang komparasi berbentuk horizontal merupakan alat visual yang efektif untuk menganalisis perbedaan antara harga prediksi dan harga aktual dari setiap merek mobil. Dengan menggunakan dataset yang telah dikelompokkan berdasarkan merek, grafik ini menggambarkan dua batang untuk setiap merek: satu bilah mewakili harga prediksi dan bilah lainnya mewakili harga sebenarnya. Warna yang berbeda pada setiap bilah membuatnya lebih mudah untuk dipahami, dan label yang jelas pada sumbu dan bilah memberikan informasi yang mudah dicerna. Melalui analisis grafik, dimungkinkan untuk melihat pola atau tren dalam perbedaan untuk setiap merek, dengan fokus pada merek yang menunjukkan perbedaan yang signifikan. Grafik ini memberi wawasan mendalam tentang sejauh mana model prediksi berhasil atau tidak untuk setiap merek, sehingga membantu dalam identifikasi area perbaikan atau penyempurnaan model. Secara keseluruhan, visualisasi ini membantu menggambarkan dengan jelas dan mudah bagaimana prediksi harga mobil sesuai dengan harga sebenarnya. Untuk melihat harga rata-rata dengan harga prediksi tiap tahunnya, dapat dilihat pada gambar 14.



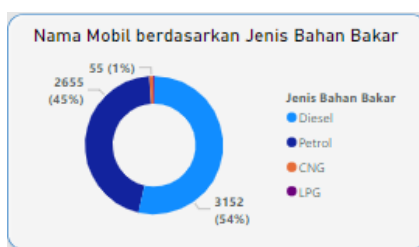
Gambar 14. Komparasi Rata-rata Harga dan Prediksi Harga setiap Tahun

Gambar 14 menunjukkan sumbu horizontal mewakili tahun, sementara sumbu vertikal mewakili harga rata-rata mobil. Ada dua garis yang menampilkan tren harga aktual dan harga prediksi. Garis biru menunjukkan harga aktual rata-rata mobil bekas setiap tahunnya, sedangkan garis orange memperlihatkan perkiraan harga yang dihasilkan dari model regresi yang sudah terbentuk sebelumnya. Tren garis biru akan menyajikan informasi visual tentang bagaimana harga mobil bekas berfluktuasi dari tahun ke tahun. Sementara itu, perbandingan dengan garis orange memperjelas keakuratan model prediksi harga dan perbandingannya dengan harga aktual mobil bekas yang dimiliki. Grafik tersebut menunjukkan bahwa dari tahun 1998 hingga 2004, perbandingan antara harga aktual dan harga prediksi cukup signifikan. Namun berbeda jika diamati pada tahun 2005 hingga 2019, dimana perbandingan antara harga aktual dan harga prediksi cenderung kurang signifikan atau masih stabil. Hasil visualisasi ini dapat memberikan informasi yang berharga bagi para stakeholder, diantaranya para pemilik bisnis mobil bekas. Menganalisis tren ini dapat digunakan sebagai bahan pertimbangan dalam perencanaan bisnis dan juga menentukan strategi pemasaran bagi pemilik bisnis selanjutnya. Adapun informasi terkait perbandingan harga di atas, terdapat juga informasi lain seperti informasi visual mengenai harga rata-rata yang mobil bekas setiap lokasi seperti yang ditunjukkan dalam gambar 15.



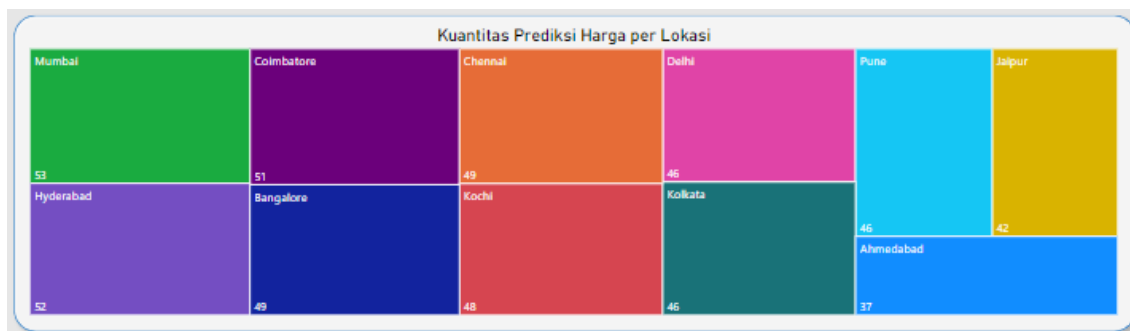
Gambar 15. Rata-rata Harga Bekas setiap Lokasi

Informasi pada gambar 15 ini dapat digunakan sebagai dasar untuk mengambil sebuah strategi penetapan harga, pemasaran atau alokasi sumber daya di berbagai lokasi. Dengan demikian, para pelaku bisnis atau analis dapat mengambil kebijakan yang lebih tepat dan akurat berdasarkan pemahaman yang didapat dari hasil visualisasi data. Dapat dilihat, untuk grafik yang menunjukkan harga rata-rata mobil bekas di area atau lokasi Coimbatore masih lebih tinggi dibandingkan area lain yang notabene masih berada dibawah wilayah tersebut. Hal ini bisa jadi disebabkan oleh perbedaan geografis dan tingkat pendapatan, faktor spesifik wilayah seperti regulasi yang mengaturnya, atau karena lokasi Coimbatore sendiri mengalami permintaan yang tinggi untuk mobil bekas dibandingkan dengan wilayah lain namun ketersediaannya terbatas. Jadi, pemilik bisnis di wilayah tersebut tentu akan menaikkan harga untuk memerangi hal ini dan secara otomatis akan mempengaruhi harga rata-rata mobil bekas di lokasi atau wilayahnya.



Gambar 16. Klasifikasi Nama Mobil setiap Jenis Bahan Bakar

Pada gambar 16, visualisasi ini menyajikan gambaran yang jelas mengenai penyebaran mobil berdasarkan jenis bahan bakar. Sebagai contoh, terlihat mobil dengan bahan bakar diesel lebih mendominasi daripada bahan bakar lain seperti Petrol, CNG, dan LNG. Hal ini dikarenakan di India, bahan bakar diesel sering ditemukan dengan harga yang lebih murah dibandingkan dengan bahan bakar lainnya, sehingga lebih ekonomis untuk digunakan. Selama beberapa tahun terakhir, rendahnya harga bahan bakar diesel telah menjadi faktor yang mempengaruhi keputusan pembeli di pasar otomotif India. Secara keseluruhan, efisiensi bahan bakar yang tinggi, torsi yang besar serta harga yang lebih rendah membuat mesin diesel menjadi pilihan yang menarik di India, terutama untuk kendaraan komersial dan mereka yang mencari efisiensi bahan bakar yang baik.



Gambar 17. Kuantitas Prediksi Harga Setiap Lokasi

Pada gambar 17, visualisasi ditampilkan dalam bentuk *treemap* di mana setiap blok atau kotak diwakilkan dengan sebuah lokasi dan ukuran dari blok tersebut merefleksikan jumlah prediksi harga mobil bekas di lokasi tersebut. Semakin besar kotaknya, semakin tinggi harga prediksi untuk lokasi tersebut. Terlihat bahwa Mumbai adalah wilayah dengan nilai prediksi tertinggi dari tahun 1998 hingga 2019 dengan nilai 53 Lakh, diikuti oleh wilayah Hyderabad, Coimbatore, Bangalore, Chennai, Kochi, Delhi, Kolkata, Pune, Jaipur, serta Ahmedabad. Hasil visualisasi dengan *treemap* tersebut mampu menjadi alat yang berguna dalam memahami kontribusi relatif setiap lokasi terhadap kuantitas prediksi harga.

4. Kesimpulan

Dari hasil penelitian ini, dapat disimpulkan bahwa metode regresi linier memiliki kemampuan untuk memprediksi harga mobil bekas, khususnya di wilayah India. Proses pengujian terhadap model yang dikembangkan memberikan tingkat akurasi sebesar 71.09%. Namun demikian, masih terdapat ruang untuk meningkatkan akurasi dengan menambah jumlah data training yang digunakan, mengganti parameter yang dipakai, atau dengan menerapkan arsitektur yang lebih kompleks. Penelitian ini diharapkan dapat memberikan kontribusi terhadap pemahaman pasar mobil bekas di India dan memberikan informasi yang berharga bagi pemilik bisnis dan pelaku industri untuk mengambil keputusan yang tepat.

5. Daftar Pustaka

- [1] Susanti, P., & Sussolaikah, K. (2022). Penerapan metode regresi linear untuk memprediksi harga jual mobil bekas yaris dan jazz pada wilayah DKI Jakarta. *Network Engineering Research Operation*, 7(2), 133-144.
- [2] Sukimin, S., & Indriastuty, N. (2021). Pengaruh Kualitas Produk, Harga dan Promosi terhadap Keputusan Pembelian Mobil Merek Toyota di Kota Balikpapan. *Jurnal GeoEkonomi*, 12(2), 194-204. DOI: <https://doi.org/10.36277/geoekonomi.v12i2.162>.
- [3] Kriswantara, B., & Sadikin, R. (2022). Machine Learning Used Car Price Prediction with Random Forest Regressor Model. *JISICOM (Journal of Information System, Informatics and Computing)*, 6(1), 40-49. DOI: <https://doi.org/10.52362/jisicom.v6i1.752>.
- [4] Hasibuan, E., & Karim, A. (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>.
- [5] Fadhilah, M., Martanto, M., & Ali, I. (2023). Prediksi Jumlah Produksi Sablon Tahun Menggunakan Algoritma Regresi Linear di Nolbas SVNR. *INTERNAL (Information System Journal)*, 6(1), 22-32.
- [6] Nasyuli, L. P., Lubis, I., & Elhanafi, A. M. (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.
- [7] Amalia, A., Radhi, M., Sinurat, S. H., Sitompul, D. R. H., & Indra, E. (2021). Prediksi harga mobil menggunakan algoritma regresi dengan hyper-parameter tuning. *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 4(2), 28-32. DOI: <https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v4i2.2479>.
- [8] Beltran, D. J., Kangleon, Y., Balan, A. K., & De Goma, J. (2021). Credit card sales performance dashboard. In *Proceedings of the International Conference on Industrial Engineering and Operations Management* (pp. 1-12).
- [9] Faiqoh, H. *Penerapan Ensemble Feature Selection Untuk Mengurangi Dimensionalitas dalam Prediksi data Time Series* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).
- [10] Jusuf, D. N. B. (2024). Implementasi Intelijen Bisnis Untuk Visualisasi Pola Pembelian Dengan Algoritma Fp-Growth. *EDUSAINTEK: Jurnal Pendidikan, Sains dan Teknologi*, 11(2), 483-507.