

**PENERAPAN REGRESI LINIER DALAM PREDIKSI HARGA
MOBIL BEKAS DAN VISUALISASI DENGAN POWER BI
(Studi Kasus: India)**



Disusun Oleh:

N a m a : Muhammad Devanda Hendra K
NIM : 20523098

**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA**

2024

HALAMAN PENGESAHAN DOSEN PEMBIMBING

**PENERAPAN REGRESI LINIER DALAM PREDIKSI HARGA
MOBIL BEKAS DAN VISUALISASI DENGAN POWER BI**

(Studi Kasus: India)

TUGAS AKHIR



Disusun Oleh:

Nama : Muhammad Devanda Hendra K
NIM : 20523098

الجامعة الإسلامية
الاستدراكية

Yogyakarta, 27 Oktober 2024

Pembimbing,

(Dr. Syarif Hidayat, S.Kom., M.I.T)

**PENERAPAN REGRESI LINIER DALAM PREDIKSI HARGA
MOBIL BEKAS DAN VISUALISASI DENGAN POWER BI
(Studi Kasus: India)**

TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 12 November 2024

Tim Penguji

Dr. Syarif Hidayat, S.Kom., M.I.T.

Anggota 1

Kholid Haryono, S.T., M.Kom.

Anggota 2

Sheila Nurul Huda, S.Kom., M.Cs.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



(Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama :Muhammad Devanda Hendra Kusuma

NIM : 20523098

Tugas akhir dengan judul:

PENERAPAN REGRESI LINIER DALAM PREDIKSI HARGA MOBIL BEKAS DAN VISUALISASI DENGAN POWER BI (Studi Kasus: India)

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 27 Oktober 2024



(Muhammad Devanda Hendra Kusuma)

HALAMAN PERSEMBAHAN

1. Pertama-tama, saya ingin mengucapkan puji syukur kehadirat Allah Subhanahu wa ta'ala, Tuhan Yang Maha Esa, yang telah memberikan rahmat, karunia dan kekuatan kepada saya sehingga saya dapat menyelesaikan skripsi saya ini.
2. Selanjutnya skripsi ini saya persembahkan secara khusus kepada ayah dan ibu yang sangat saya sayangi dan hormati. Ada banyak hal yang ingin saya katakan, tetapi saya tidak bisa menuliskannya satu per satu. Semoga hasil dan perjuangan saya selama ini membuahkan hasil yang baik. Semangat terus membara dalam diri saya untuk bisa menghadapi dunia luar yang sebenarnya nanti, lebih dari 17 tahun setelah menekuni dunia pendidikan tinggi, saatnya saya membuktikan kepada Anda dan dunia bahwa saya siap melangkah lebih jauh di dunia. masa depan, dan saya berharap saya bisa menjadi orang yang bertanggung jawab dan bekerja untuk kepentingan banyak orang.
3. Bapak dan Ibu Dosen khususnya Fakultas Teknologi Industri Program Studi Informatika Universitas Islam Indonesia yang telah mendidik dan membina saya, terimakasih telah menjadikan saya pribadi yang lebih baik dan berkembang dari sebelumnya. Semoga ilmu yang bapak ibu berikan kepada saya dapat bermanfaat dan menjadi berkah bagi banyak orang.
4. Kakak tercinta yang selalu membantu dan mendukung dalam pengerjaan.
5. Teman-temanku yang ada di Pekalongan dan Yogyakarta yang tidak pernah lelah memberikan dukungan moral dan moril kepadaku. Semoga kebaikan selalu kembali padamu dan dibalas oleh Allah Subhanahu Wa Ta'ala.
6. Untuk seseorang yang saya cintai, terima kasih telah menjadi salah satu penyemangat, pendengar keluh kesah dalam penulisan skripsi ini, penasehat yang baik dan senantiasa memberikan cinta dan kasih sayang. Terima kasih juga telah menjadi sosok rumah yang selalu ada buat saya serta menyemangati setiap proses yang saya lakukan.
7. Terima kasih, Muhammad Devanda Hendra Kusuma. Terimakasih bertahan sampai saat ini. Untuk kedepannya, ayo berkembang lebih jauh, lebih indah, lebih ceria. *You've been great for getting this far, don't let irresponsible people hurt you! Someday you'll find that you are brighter than the star. Just be strong, just be brave, and be sure. Yes you can!!.*

HALAMAN MOTO

“Bersabarlah kamu dan kuatkanlah kesabaranmu dan tetaplah bersiap siaga dan bertakwalah kepada Allah supaya kamu menang.” (QS Ali Imran: 200)

"Barang siapa yang bersungguh-sungguh, sesungguhnya kesungguhan tersebut untuk kebaikan dirinya sendiri." (Qs. Al-Ankabut: 6)

"Sesungguhnya pertolongan datang bersama kesabaran." (HR. Ahmad)

“Jika kita ingin sukses, kita harus lelah. Jika kita tidak lelah, kita tidak berhasil.” (Raffi Ahmad)

“Di hidup ini kita butuh orang lain, kita butuh sesama. Tapi untuk kesuksesan dan keberhasilan diri kita sendiri, jangan pernah mengandalkan orang lain” (Windah Basudara)

“Kalau kita capek kita boleh istirahat sejenak, tapi setelah itu kita kembali lagi ke *track* kita untuk mengejar mimpi kita.” (Raffi Ahmad)

KATA PENGANTAR

Assalamualaikum Wr. Wb.

Dengan segala rasa syukur dan pujian, penulis mengucapkan terima kasih kepada Allah SWT yang telah memberikan rahmat, petunjuk, dan kesempatan bagi penulis dal penyelesaian penulisan skripsi ini sebagai salah satu persyaratan meraih gelar Sarjana Komputer dari Fakultas Teknologi Industri Universitas Islam Indonesia. Saya menyadari bahwa penyelesaian skripsi ini tidak terlepas dari bantuan, dukungan, dan dorongan dari berbagai pihak. Oleh karena itu, pada kesempatan ini, dengan penuh rasa tulus dan ikhlas, saya ingin mengucapkan terima kasih sebesar-besarnya kepada

1. Yang terhormat Bapak Prof. Fathul Wahid, S.T., M.Sc., Ph.D., sebagai Rektor Universitas Islam Indonesia Yogyakarta.
2. Bapak Hari Purnomo, Prof., Dr., Ir., M.T., sebagai Dekan Fakultas Teknologi Industri Universitas Islam Indonesia Yogyakarta..
3. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. sebagai Ketua Program Studi Informatika Program Sarjana, Fakultas Teknologi Industri, Universitas Islam Indonesia
4. Bapak Dr. Syarif Hidayat, S.Kom., M.I.T., sebagai Dosen Pembimbing, dimana dengan penuh kesabaran dan perhatian dalam memberikan pengarahan, bimbingan, serta petunjuk kepada penulis.
5. Kedua Orangtua papa dan mama yang selalu mendoakan penulis agar apa yang dikerjakannya dapat berjalan dengan lancar sampai selesai.
6. Kakak yang selalu membantu dan mendukung dalam pengerjaan.
7. Teman-teman yang berada di Yogyakarta maupun di Pekalongan yang telah memberikan dukungan serta *support*.
8. Seseorang yang kucintai dan mencintaiku.
9. Dan semua pihak yang telah membantu penulis yang tidak dapat disebutkan satu persatu.

Semoga amal kebaikan senantiasa mendapatkan balasan dan rahmat dari Allah SWT.

Tentu saja penulis menyadari bahwa penyusunan skripsi ini masih jauh dari kata sempurna karena keterbatasan penulis. Namun demikian, penulis telah berusaha maksimal agar hasil dan pembahasan skripsi ini dapat bermanfaat bagi penulis dan pembaca.

Wassalamualaikum Wr. Wb.

Yogyakarta, 29 April 2024

A handwritten signature in black ink, featuring a large, stylized 'D' followed by a series of loops and a long horizontal stroke.

(Muhammad Devanda Hendra Kusuma)

SARI

Pasar mobil bekas India berkembang pesat karena minat pada mobil bekas yang semakin tinggi. Namun, sulitnya menentukan harga dalam transaksi menyebabkan mobil bisa dijual dengan harga terlalu tinggi atau terlalu rendah, memungkinkan penjual memperoleh keuntungan dari menjual mobil dengan kualitas rendah namun disisi lain pembeli bisa memperoleh harga rendah dari mobil bekas kualitas tinggi di karenakan kurangnya kurang informasi.

Dalam penelitian ini mencoba mencari fakta bahwa dengan menggunakan algoritma regresi linier dapat terbukti efektif mengatasi masalah seperti pada penelitian sebelumnya. Analisis data dilakukan untuk memprediksi harga mobil bekas di India yang bertujuan untuk mengurangi potensi terjadinya penentuan harga mobil yang *overprice* atau *underprice*. Visualisasi data dilakukan melalui platform Power BI, meningkatkan efektivitas dalam mempresentasikan hasil prediksi harga mobil bekas, memudahkan pemangku kepentingan dalam mengoptimalkan bisnis.

Hasil analisis menunjukkan bahwa sistem prediksi harga mobil bekas dengan menggunakan algoritma regresi linier telah berhasil dilakukan dengan tingkat akurasi 73.18%. Selain itu, pada penelitian ini juga berhasil meningkatkan efektivitas presentasi hasil prediksi harga melalui visualisasi yang baik dengan dilakukan pengujian *dashboard* dengan pengujian *System Usability Scale (SUS)* yang memperoleh nilai *excellent*. Dengan demikian mampu menegaskan bahwa hasil prediksi disajikan secara efektif dalam merepresentasikan hasil dari prediksi harga mobil bekas.

Kata kunci: India, Regresi Linier, Prediksi Harga, Viualisasi Data, *System Usability Scale*.

GLOSARIUM

Dataset	kumpulan data yang digunakan untuk melatih dan menguji model Machine Learning.
Evaluasi Model	tahap dimana digunakan untuk mengidentifikasi keakuratan suatu model.
Fitur (Features)	variabel independen atau atribut dalam dataset yang digunakan sebagai input untuk model prediksi
Machine Learning	bidang ilmu komputer yang fokus pada pengembangan algoritma dan model statistik yang membuat sistem komputer untuk belajar dari data dan membuat prediksi atau keputusan mun pemrograman eksplisit.
MAE	Mean Absolute Error (MAE) merupakan matrik evaluasi yang mengukur selisih absolut rata-rata antara nilai prediksi dan nilai sebenarnya.
MSE	MSE (Mean Square Error) adalah nilai rerata dari selisih kuadrat antara nilai prediksi dan nilai aktual yang sedang diamati.
Power BI	perangkat lunak Business Intelligence (BI) yang digunakan untuk menggabungkan, menganalisis, dan memvisualisasikan data untuk membuat laporan dan dashboard interaktif.
R-squared (R^2)	metrik evaluasi yang mengukur seberapa baik model sesuai dengan data.
Regresi Linier	metode statistik yang digunakan untuk memodelkan hubungan linier antara variabel independen (fitur) dan variabel dependen (target). Ini adalah metode dasar dalam analisis regresi.
RMSE	RMSE (Root Mean Squared Error) merupakan hasil dari akar kuadrat rata-rata kuadrat antara nilai aktual dan nilai prediksi.
Seleksi Fitur	proses pemilihan fitur yang paling relevan atau penting dari dataset untuk meningkatkan kinerja model.
Visualisasi Data	proses mewakili data dalam bentuk grafik, diagram, atau tampilan visual lainnya untuk memahami pola dan tren dalam data.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING.....	ii
HALAMAN PENGESAHAN DOSEN PENGUJI.....	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR.....	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO	vi
KATA PENGANTAR	vii
SARI	ix
GLOSARIUM.....	x
DAFTAR ISI	xi
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5
BAB II LANDASAN TEORI.....	6
2.1 Dasar Teori	6
BAB III METODOLOGI PENELITIAN	30
3.1 Studi Literatur.....	30
3.2 Populasi dan Sampel	34
3.3 Pengumpulan Data	34
3.4 Variabel Penelitian	35
3.5 Perangkat Penelitian	36
3.6 Metode Analisis.....	37
3.7 Tahapan Penelitian	37
BAB IV HASIL DAN PEMBAHASAN	47
4.1 Implementasi Regresi Linier	47
4.2 ETL dan Perancangan <i>Data Warehouse</i>	63
4.3 Visualisasi Dashboard	67
4.4 Pengujian Dasbor	77
BAB V KESIMPULAN DAN SARAN	80
5.1 Kesimpulan.....	80
5.2 Saran.....	80
DAFTAR PUSTAKA	81
LAMPIRAN	89

DAFTAR TABEL

Tabel 3.1 Analisis Perbandingan Literatur.....	32
Tabel 3.2 Definisi Operasional Variabel (DOV).....	35
Tabel 3.3 Daftar Perangkat Lunak	36
Tabel 3.4 Daftar Perangkat Keras.....	37
Tabel 3.5 Daftar Kuesioner	46
Tabel 4.1 Feature Missing.....	49
Tabel 4.2 Sampel Fitur baru yang dihasilkan.....	51
Tabel 4.3 Nilai Prediksi dan Nilai Aktual.....	59
Tabel 4.4 Hasil pengukuran Metrik Kesalahan.....	61
Tabel 4.5 Pengujian Prediksi.....	62
Tabel 4.6 Pencatatan Informasi secara Acak.....	77
Tabel 4.7 Perbandingan Prediksi Harga dengan Harga Rill.....	78
Tabel 4.8 Hasil Pengujian SUS.....	79
Tabel 4.9 Daftar Perangkat Keras.....	79

DAFTAR GAMBAR

Gambar 1.1 <i>Wholesales</i> Mobil Baru vs Bekas	2
Gambar 2.1 Komponen Utama Data Warehouse	19
Gambar 2.2 Arsitektur Data Warehouse	20
Gambar 2.3 Skor SUS	29
Gambar 3.1 Sample dataset uji yang digunakan	35
Gambar 3.2 Sample dataset latih yang digunakan	35
Gambar 3.3 Tahapan Penelitian	38
Gambar 3.4 Kode Program untuk Menghitung Metrik Kesalahan	43
Gambar 4.1 Hasil Analisis Korelasi	47
Gambar 4.2 Kode Program untuk Penyesuaian Tipe Data.....	49
Gambar 4.3 Kode Program untuk Mengubah Satuan	50
Gambar 4.4 Kode Program untuk Menghapus Satuan Variabel	50
Gambar 4.5 Kode Program untuk Penggantian Nama Beberapa Variabel	50
Gambar 4.6 Kode Program Penerapan One-Hot Encoding.....	52
Gambar 4.7 Label Encoding.....	52
Gambar 4.8 Pengecekan Korelasi dengan <i>Heatmap</i>	53
Gambar 4.9 Kode Program Identifikasi Data Outliers	54
Gambar 4.10 Hasil Identifikasi Data Outliers	55
Gambar 4.11 Analisis Matriks Korelasi	57
Gambar 4.12 <i>Data Verification</i>	58
Gambar 4.13 Fungsi Regresi Linier Berganda	59
Gambar 4.14 Grafik Harga Prediksi dan Harga Aktual	60
Gambar 4.15 Save Model.....	62
Gambar 4.16 Pengecekan Dataset.....	63
Gambar 4.17 Validitas Dataset yang sudah digabungkan	64
Gambar 4.18 Pengaturan Relasi Tabel	64
Gambar 4.19 Relasi Tabel Data yang Terbentuk	65
Gambar 4.20 <i>Dimensional Modelling</i>	66
Gambar 4.21 Rancangan Dashboard Power BI.....	68
Gambar 4.22 Distribusi Rerata Harga Bekas (IDR) setiap Lokasi	69
Gambar 4.23 Analisis Tren Harga.....	70
Gambar 4.24 Rata-rata Tenaga (bhp) tiap Merk	71
Gambar 4.25 Rata-rata Kapasitas Mesin (cc) tiap Merk.....	71

Gambar 4.26 Rata-rata Konsumsi BBM (kmpl) tiap Merk.....	72
Gambar 4.27 Komparasi Harga Bekas vs Prediksi Harga tiap Mobil.....	73
Gambar 4.28 Grafik Proporsi Jumlah Mobil tiap Jumlah Kursi	73
Gambar 4.29 Komparasi Harga Bekas vs Prediksi Harga tiap Mobil.....	74
Gambar 4.30 Slicer Tipe Transmisi dan Tipe Pemilik.....	75
Gambar 4.31 Slicer Tahun Produksi dan Harga Bekas (IDR)	76
Gambar 4.32 <i>Card</i>	76

BAB I

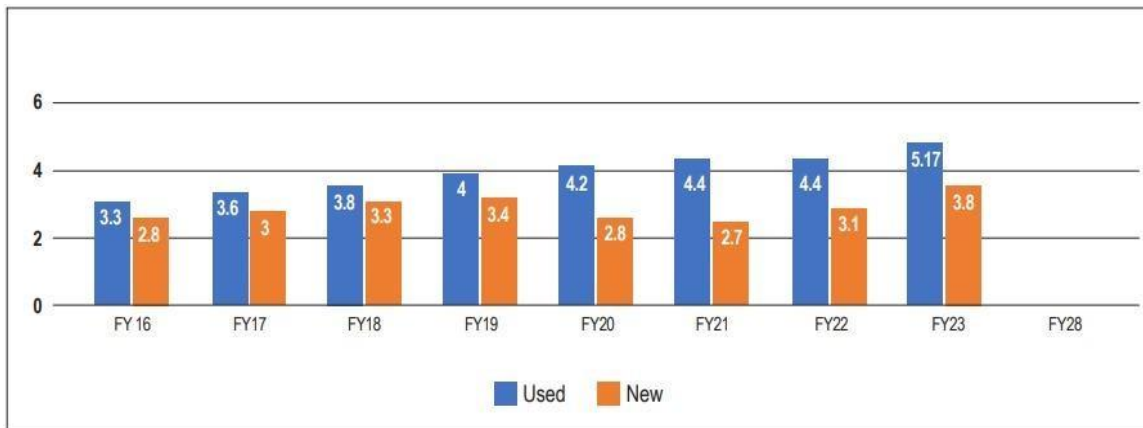
PENDAHULUAN

1.1 Latar Belakang

Bursa mobil bekas di India terus berkembang dengan pesat sebagai sebuah bisnis yang dinamis dan terus berkembang. Hal ini disebabkan oleh adanya perbedaan preferensi konsumen, faktor ekonomi dan kemajuan teknis. Urbanisasi yang cepat, peningkatan infrastruktur transportasi dan peningkatan penghasilan berkontribusi pada peningkatan permintaan mobil bekas. Beragamnya konsumen yang sensitif dengan harga juga menjadi penyebab semakin meningkatnya pangsa pasar mobil bekas. Calon konsumen mulai berpaling ke mobil bekas sebagai alternatif yang lebih ekonomis daripada membeli mobil baru. Di samping itu, mobil bekas turut berperan dalam menjaga kelestarian lingkungan melalui perpanjangan masa pakai kendaraan serta mengurangi dampak karbon secara menyeluruh dari industri otomotif (Kumar, Manish, & Vatsalya, 2023).

Sulitnya menentukan harga dalam transaksi penjualan mobil bekas di India dapat menyebabkan harga mobil yang terlalu mahal atau terlalu murah. Faktor-faktor seperti kondisi kendaraan, harga, spesifikasi, dan riwayat pemilik sebelumnya menyulitkan pemilik bisnis untuk menentukan harga yang tepat, sehingga memungkinkan pemilik bisnis tidak dapat mendapat keuntungan secara maksimal (Chen, Gu, Deng, & Huang, 2022). Selain itu, harga mobil bekas juga dipengaruhi oleh permintaan pasar, pemanfaatan kapasitas, target area distribusi, dan program promosi antara pelanggan dan perusahaan (Wandi & Abaharis, 2017).

Sektor otomotif di India telah menjadi salah satu industri mobil yang paling cepat berkembang. Berdasarkan data dari Society of Indian Automotive Manufacturers (SIAM), penjualan tahunan kendaraan penumpang mencapai sekitar 3,8 juta unit, mencatat peningkatan rekor sebesar 26%. Pertumbuhan penjualan mobil baru ini didorong oleh perkembangan signifikan di industri mobil bekas, yang memberikan berbagai opsi bagi pembeli untuk memiliki kendaraan dengan harga yang lebih terjangkau. Penjualan unit mobil bekas telah meningkat secara signifikan, dan industri ini diprediksi akan terus mengalami pertumbuhan. Pada tahun fiskal 2023, penjualan mobil bekas di India mencapai hampir 5,2 juta unit (BlueBook, n.d.).



Gambar 1.1 *Wholesales Mobil Baru vs Bekas*

Sumber: (BlueBook, n.d.)

Grafik yang disajikan menunjukkan bahwa pasar mobil baru di India mengalami pertumbuhan yang kuat, namun terdapat juga potensi besar yang belum dimanfaatkan di pasar mobil bekas. Menariknya, jika kita menganalisis secara terpisah industri mobil bekas dan baru dalam periode empat hingga lima tahun terakhir, Tingkat Pertumbuhan Tahunan Gabungan (CAGR) untuk kendaraan bekas telah jauh melampaui pertumbuhan mobil baru. Seperti yang terlihat pada grafik di bawah ini, industri mobil baru telah kembali ke angka sebelum pandemi, sementara industri mobil bekas terus mengalami pertumbuhan yang pesat.

Prediksi tidak selalu memberikan jawaban yang pasti tentang kejadian apa yang akan terjadi, tetapi mencoba untuk mendapatkan hasil yang mendekati apa yang sebenarnya terjadi. Data prediksi memegang peranan penting dalam sebuah bisnis, seperti jual beli mobil bekas. Saat memilih mobil bekas tentu ada banyak faktor yang harus dipertimbangkan dan tentunya akan menentukan harga mobil yang ditawarkan. Semakin tinggi angka kilometer yang ditempuh, artinya mobil tersebut telah digunakan untuk jarak jauh. Dari segi harga jual, hal ini tentu saja akan berpengaruh pada harga jual. Konsumen meyakini bahwa mobil dengan jarak tempuh yang tinggi maka performanya akan menurun seiring berjalannya waktu. Sama halnya dengan mobil yang sudah tua, maka harga jualnya pun akan semakin menurun (Kriswantara & Sadikin, 2022). Salah satu cara yang dapat digunakan untuk memecahkan permasalahan prediksi ini adalah dengan menerapkan machine learning dan metode yang bisa diadaptasi yaitu metode regresi linier.

Dalam penelitian ini metode regresi linier yang digunakan adalah metode regresi linier berganda. Metode ini digunakan untuk membuat model linier dua variabel atau lebih, dengan

tujuan untuk mencari tahu hubungan antara variabel yang dimodelkan tersebut (Fadhilah & Ali, 2023). Keunggulan yang dimiliki dari metode regresi linier berganda sendiri antara lain adalah dapat melakukan generalisasi sekaligus ekstraksi terhadap pola data tertentu, dapat memperoleh informasi meskipun belum ada kepastian, dan dapat melakukan perhitungan secara bersamaan sehingga prosesnya lebih singkat serta sangat cocok digunakan untuk memprediksi data harga mobil bekas. Ketika melakukan pemodelan prediksi, penting untuk menghitung kesalahan prediksi (Ihzaniah, Setiawan, & Wijaya, 2023), ini dapat dilakukan dengan mengevaluasi model menggunakan beberapa metrik diantaranya MAE (Error mean absolute), RMSE (Root Mean Squared Error), MSE (Mean Squared Error), dan R-Squared (R^2).

Setelah dilakukan pemodelan dan evaluasi model, nantinya akan divisualisasikan untuk memudahkan dalam pembacaannya. Visualisasi data dalam penelitian ini menggunakan aplikasi Power BI. Power BI adalah platform visualisasi data yang memungkinkan pengguna untuk mengubah data yang kompleks menjadi visualisasi yang mudah dipahami (Becker & Gould, 2019). Visualisasi ini dapat membantu untuk menemukan tren harga mobil bekas dengan prediksi harga untuk setiap tahun, kemudian dapat memilah berdasarkan jenis kepemilikan, jenis transmisi, tahun pembuatan, dan harga mobil bekas yang diinginkan. Oleh karena itu, penelitian ini tidak hanya berfokus pada prediksi harga, tetapi juga pada kemampuan untuk mengkomunikasikan informasi ini secara visual kepada para pemangku kepentingan di industri perdagangan mobil bekas. Visualisasi data sangat penting untuk dilakukan karena memungkinkan pengambilan keputusan yang lebih cepat serta akurat. Tanpa visualisasi, analisis data yang kompleks akan sulit dan rentan terhadap kesalahan.

Dengan demikian, penelitian ini dilakukan dengan judul ***“Penerapan Regresi Linier Dalam Prediksi Harga Mobil Bekas dan Visualisasi dengan Power BI”***

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang diatas, maka rumusan masalah dalam penelitian ini sebagai berikut:

- 1 Sulitnya menentukan harga yang akurat dalam transaksi penjualan mobil bekas yang menyebabkan terjadinya potensi *overpriced* atau *underpriced*.
- 2 Kurangnya visualisasi data yang efektif dalam mempresentasikan hasil dari prediksi harga mobil bekas.

1.3 Batasan Masalah

Agar pembahasan masalah tidak menyimpang dari tujuan sebenarnya, serta sesuai dengan permasalahan, maka peneliti membatasi topik yang dibahas, yaitu:

- 1 Dataset used car price prediction dari situs web Kaggle adalah satu-satunya sumber data mobil bekas yang digunakan.
- 2 Tujuan utama dari model yang dibuat adalah memberikan hasil prediksi harga mobil bekas untuk mengatasi tantangan penentuan harga mobil bekas serta menghasilkan sebuah keputusan kepada pihak yang terlibat terkhusus pemilik bisnis.
- 3 Metode yang digunakan pada penelitian ini adalah menggunakan Regresi Linier dengan Bahasa pemrograman Python menggunakan tools Google Colaboratory.
- 4 Tools yang digunakan sepenuhnya untuk melakukan visualisasi data adalah dengan menggunakan Microsoft Power BI.

1.4 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang telah diuraikan, maka tujuan dilakukannya penelitian ini adalah:

- 1 Menerapkan model regresi linier untuk memprediksi harga mobil bekas di India yang bertujuan untuk mengurangi potensi terjadinya penentuan harga mobil yang terlalu mahal atau terlalu rendah.
- 2 Meningkatkan efektivitas dalam mempresentasikan hasil prediksi harga mobil bekas, sehingga memudahkan pemangku kepentingan dalam mengoptimalkan bisnis.

1.5 Manfaat Penelitian

- 1 Bagi Pemilik Bisnis

Dengan menerapkan model regresi linier untuk memprediksi harga mobil bekas, pemilik bisnis dapat menentukan harga yang lebih akurat, mengurangi potensi untuk menjual mobil dengan harga terlalu tinggi atau terlalu rendah. Hal ini membantu mereka mengoptimalkan keuntungan serta menghindari kerugian. Selain itu, dengan meningkatkan efektivitas dalam mempresentasikan hasil prediksi harga mobil bekas, pemilik bisnis dapat mengoptimalkan bisnis mereka.

2 Bagi Peneliti Lain

Penelitian ini dapat memberikan wawasan serta pengetahuan kepada peneliti lain tentang bagaimana menggunakan regresi linier untuk memprediksi harga mobil bekas. Hal ini akan membantu dalam memperluas pemahaman dan penggunaan model regresi linier dalam konteks harga mobil bekas.

3 Bagi Akademik

Penelitian ini bertujuan untuk menambah literatur dalam bidang *data science* dan dapat digunakan sebagai bahan bacaan untuk meningkatkan pemahaman maupun pengetahuan serta sebagai referensi untuk penelitian-penelitian selanjutnya.

1.6 Sistematika Penulisan

Dalam melakukan penyusunan laporan penelitian, pembahasan dibagi menjadi lima bab, yang pertama meliputi pendahuluan, landasan teori, metodologi penelitian, hasil dan pembahasan, dan penutup. Penjabaran selengkapnya sebagai berikut:

BAB I PENDAHULUAN

Bab ini menyajikan gambaran umum dari keseluruhan penelitian secara ringkas seperti latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan tentang beberapa penelitian serupa yang pernah dilakukan sebelumnya serta teori-teori yang mendukung permasalahan yang dibahas dalam penelitian ini.

BAB III METODE PENELITIAN

Bab ini menjelaskan tahapan yang akan dilakukan dalam proses pelaksanaan penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dan pembahasan untuk menjawab permasalahan yang ada dalam penelitian ini.

BAB V KESIMPULAN DAN SARAN

Bab ini memuat kesimpulan serta saran yang ditujukan untuk meningkatkan kelengkapan penyempurnaan penelitian.

BAB II

LANDASAN TEORI

2.1 Dasar Teori

Memuat teori-teori utama yang menjadi dasar dari berbagai teori lain yang terdapat dalam skripsi ini.

2.1.1 Prediksi

Prediksi atau peramalan diartikan sebuah kegiatan yang bertujuan untuk memprediksikan apa yang akan terjadi di masa yang akan mendatang (Prasetyowati, 2018). Dalam bisnis, prediksi diperlukan untuk menentukan dasar perencanaan. Tanpa dokumen acuan (dasar), mustahil untuk membangun rencana bisnis karena akan berfokus pada jumlah bahan yang dibutuhkan, peralatan apa yang akan digunakan, dimana melakukannya, siapa yang melakukannya dan berapa biayanya. Semuanya akan menjadi sulit tanpa adanya referensi (Utama, 2019). Fungsi prediksi sendiri adalah untuk membuat rencana tentang kebutuhan (permintaan) apa yang harus dicapai dan dinyatakan dalam jumlah (kuantitas) sebagai fungsi waktu. Prediksi dibuat dalam jangka waktu yang panjang. Prediksi berkenaan dengan pertanyaan (1) apa yang akan dibutuhkan, (2) berapa banyak, dan (3) kapan harus dipasok. Prediksi sangat dibutuhkan untuk membandingkan prediksi dengan permintaan aktual (Fengki, 2018).

Secara ringkas, pada tahap prediksi atau peramalan, ada tiga langkah yang harus dilakukan dalam merancang metode prediksi, yaitu (Indarwati, Irawati, & Rimawati, 2019):

- 1 Menganalisis data masa lalu. Langkah ini ditujukan untuk mendapatkan gambaran tentang data historis.
- 2 Menentukan metode yang digunakan. Terdapat berbagai macam metode yang tersedia sesuai dengan kebutuhan. Penentuan metode dapat mempengaruhi hasil prediksi. Hasil prediksi diukur dengan melakukan perhitungan kesalahan terkecil. Oleh karena itu, tidak ada metode prediksi yang benar-benar cocok untuk semua jenis data.
- 3 Proses transformasi data masa lalu dengan menerapkan metode yang dipilih. Jika perlu, perubahan dilakukan sesuai dengan kebutuhan.

Prediksi dapat bersifat kualitatif (bukan dalam bentuk angka) maupun kuantitatif (dalam bentuk angka). Prediksi kualitatif sulit mencapai hasil yang baik karena variabel-variabelnya bersifat sangat relatif. Prediksi kuantitatif dibagi menjadi dua, yaitu: prediksi tunggal (point prediction) dan prediksi interval (interval prediction) (Syafuruddin, Hakim, & Despa, 2014). Terdapat sejumlah teknik prediksi yang telah dikembangkan untuk melakukan prediksi yang diantaranya dapat dikategorikan menjadi dua metode, yaitu:

- 1 Metode prediksi kualitatif menggabungkan elemen-elemen seperti keputusan intuisi pengambilan keputusan, emosi, dan pengalaman pribadi.
- 2 Metode prediksi kuantitatif memakai satu atau lebih model matematis dengan data historis dan variabel penyebab untuk memprediksi permintaan (Tindriyani, 2017).

(Fengki, 2018) mengatakan bahwa metode prediksi kualitatif didasarkan pada pemikiran, penilaian, atau perspektif intuitif, pengetahuan, dan pengalaman penyusunnya, yang memengaruhi hasilnya. Prediksi ini sangat bergantung pada siapa yang menyusunnya. Biasanya, prediksi kualitatif ini didasarkan pada hasil survei. Situasi ini berbeda dengan metode prediksi kuantitatif yang didasarkan pada data historis dan variabel penyebab. Hasil prediksi yang diberikan sangat bergantung pada metode yang digunakan dalam prediksi tersebut. Dengan metode yang berbeda, maka akan diperoleh hasil prediksi yang berbeda pula. Yang perlu diperhatikan dalam penggunaan metode-metode tersebut adalah baik tidaknya metode yang digunakan sangat ditentukan oleh selisih atau deviasi antara hasil prediksi dengan hasil aktual yang terjadi. Metode yang baik akan memberikan hasil prediksi yang lebih akurat dan tepat waktu. Prediksi kuantitatif hanya dapat digunakan jika tiga kondisi berikut terpenuhi:

- 1 Informasi yang tersedia tentang data masa lampau (data historis).
- 2 Informasi ini dapat dinyatakan dalam bentuk angka.
- 3 Dapat diukur berdasarkan beberapa aspek pola masa lampau akan terus berkembang seiring waktu.

Berdasarkan periode waktunya, prediksi dapat digolongkan menjadi tiga, yaitu (Utama, 2019):

- 1 Prediksi jangka pendek menghasilkan hasil prediksi untuk satu tahun ke depan atau kurang.
- 2 Prediksi jangka menengah untuk memprediksi kondisi hingga 5 tahun ke depan.
- 3 Prediksi jangka panjang dipakai untuk mengambil keputusan yang berhubungan dengan perencanaan produk dan perencanaan pasar, pengeluaran perusahaan, studi

kelayakan pabrik, anggaran, pesanan, penjadwalan tenaga kerja dan perencanaan kapasitas kerja serta pengambilan keputusan yang berkaitan dengan kejadian yang berlangsung lebih dari lima tahun yang akan datang.

2.1.2 Python

Python adalah bahasa pemrograman interpretatif multiguna. Tidak seperti bahasa lain yang susah untuk dibaca dan dipahami, Python lebih menekankan pada keterbacaan kode agar lebih mudah untuk memahami sintaks. Hal ini membuat Python sangat mudah dipelajari baik untuk pemula maupun untuk yang sudah menguasai bahasa pemrograman lain.

Bahasa pemrograman Python muncul pertama kali pada tahun 1991, yang dirancang oleh seseorang bernama Guido van Rossum. Sampai saat ini Python masih dikembangkan oleh Python Software Foundation. Bahasa Python mendukung hampir semua sistem operasi, bahkan untuk sistem operasi Linux, dan hampir semua distronya sudah menyertakan Python di dalamnya.

2.1.3 Google Colaboratory

Sebelum adanya Google Colaboratory, Jupyter Notebook sudah menjadi dasar sebagai Colaboratory ini sendiri. Jupyter adalah alat sumber terbuka dan berbasis browser yang mengintegrasikan bahasa interpretasi, perpustakaan, dan alat untuk visualisasi (Carneiro et al., 2018). Notebook Jupyter dapat bekerja secara lokal atau di cloud. Setiap dokumen terdiri dari beberapa sel, di mana setiap sel berisi bahasa skrip atau kode penurunan harga, dan outputnya tertanam dalam dokumen. Output yang umum mencakup teks, tabel, bagan, dan grafik. Penggunaan teknologi ini memudahkan untuk berbagi dan mereplikasi karya ilmiah karena eksperimen dan hasilnya disajikan secara mandiri (Randles, Pasquetto, Golshan, & Borgman, 2017)

Google Colaboratory (alias Colab) adalah proyek yang memiliki tujuan untuk menyebarkan pendidikan dan penelitian pembelajaran mesin. Buku catatan kolaborasi didasarkan pada Jupyter dan berfungsi sebagai objek Google Dokumen: dapat dibagikan dan pengguna dapat berkolaborasi pada buku catatan yang sama. Colaboratory menyediakan runtime Python 2 dan 3 yang telah dikonfigurasi sebelumnya dengan pustaka pembelajaran mesin dan kecerdasan buatan yang penting, seperti TensorFlow, Matplotlib, dan Keras. Mesin virtual yang sedang berjalan (VM) dinonaktifkan setelah jangka waktu tertentu, dan semua data serta konfigurasi pengguna akan hilang. Namun demikian, notebook tetap dipertahankan, dan

juga dapat mentransfer file dari hard disk VM ke akun Google Drive pengguna. Terakhir, layanan Google ini menyediakan runtime yang dipercepat GPU, juga dikonfigurasi sepenuhnya dengan perangkat lunak yang diuraikan sebelumnya. Infrastruktur Google Colaboratory dihosting di platform Google Cloud (Carneiro et al., 2018).

2.1.4 Pandas

Pandas merupakan sebuah library pada Python yang berlisensi BSD dan open source yang menyediakan struktur data dan analisis data yang mudah untuk digunakan. Pandas melakukan tugas penting seperti menyelaraskan data untuk perbandingan dan penggabungan dataset, penanganan data yang hilang, dan lain sebagainya.

Struktur data Pandas dinamakan dataframe, yaitu sebuah koleksi kolom yang berurutan dengan nama dan jenis. Dengan adanya fitur dataframe memudahkan untuk membaca sebuah file dan menjadikannya sebuah table. Dataframe juga dapat mengolah suatu data dengan menggunakan operasi seperti join, distinct, group by, agregasi, dan fitur lainnya yang terdapat pada SQL. Format file yang didukung Pandas meliputi file dengan ekstensi .txt, .csv, .tsv dan lainnya (Wijoyo, 2021).

2.1.5 Scikit-learn

Scikit-Learn atau SKlearn Machine Learning in Python merupakan Alat sederhana dan efisien untuk penambahan data dan analisis data. Dapat diakses oleh semua orang, dan dapat digunakan kembali dalam berbagai konteks, dibangun di NumPy, SciPy, dan matplotlib, dan Sumber terbuka, dapat digunakan secara komersial - lisensi BSD.

Dalam SKlearn terdapat banyak modul seperti classification, regression, clustering, preprocessing, model selection. Metode Support Vector Classification dapat diperluas untuk memecahkan masalah regresi. Metode ini disebut Support Vector Regression. Model yang dihasilkan oleh klasifikasi vektor pendukung (seperti dijelaskan di atas) hanya bergantung pada subset dari data pelatihan karena fungsi biaya untuk membangun model tidak peduli dengan titik pelatihan yang berada di luar margin. Secara analog, model yang dihasilkan oleh Support Vector Regression hanya bergantung pada subset dari data pelatihan, karena fungsi biaya untuk membangun model mengabaikan data pelatihan yang dekat dengan prediksi model.

2.1.6 Numpy

Numpy merupakan salah satu library dalam Python. Library ini dapat digunakan untuk banyak case dalam data science. Dengan adanya library ini, maka tidak diperlukan lagi baris kode yang panjang untuk menjalankan program machine learning. Numpy sendiri merupakan singkatan dari Numerical Python. Pada umumnya penggunaan library ini untuk menghitung operasi matematika pada array (Wijoyo, 2021).

2.1.7 Matplotlib

Matplotlib adalah library Python 2 Dimensi yang dapat menghasilkan plot dengan kualitas tinggi dalam berbagai format dan dapat digunakan di banyak platform. Matplotlib dapat digunakan sebagai pembuat grafik dalam berbagai platform, seperti Python dan Jupyter Notebook. Grafik yang dapat dibuat oleh matplotlib cukup beragam, seperti grafik garis, batang, lingkaran, histogram dan berbagai macam lainnya (Wijoyo, 2021).

2.1.8 Seaborn

Seaborn adalah library yang bertujuan untuk membuat grafik dan statistik dengan menggunakan Python. Seaborn dibangun berdasarkan library matplotlib serta terintegrasi dengan struktur data pada Pandas (Wijoyo, 2021).

2.1.9 Machine Learning

Machine learning (ML) adalah salah satu bagian dari artificial intelligence (AI) yang dapat menghasilkan keputusan, salah satunya prediksi harga mobil bekas berdasarkan data historis (Alzubi, Nayyar, & Kumar, 2018). ML adalah metode yang membuat mesin maupun komputer dapat belajar dari pengalaman atau bagaimana caranya agar mesin dapat belajar. ML memerlukan data untuk belajar sehingga dapat juga disebut learn from data. Pembelajaran ML sendiri dapat diimplementasikan dengan menggunakan banyak algoritma pembelajaran yang berbeda. Secara umum, ada tiga jenis metode pembelajaran yang digunakan, yaitu supervised learning, unsupervised learning, dan reinforcement learning.

Supervised learning merupakan metode pembelajaran yang terawasi. Program menerima beberapa sampel data dari jenis atau klasifikasi yang disebut materi pembelajaran atau pelatihan. Program ini masih dapat salah mengklasifikasikan objek setelah pelatihan. Oleh

karena itu, selain menggunakan data pelatihan, penggunaan data uji juga diperlukan. Tingkat kelulusan program dapat diukur berdasarkan data uji ini.

Sedangkan unsupervised learning merupakan metode pembelajaran tanpa pengawasan. Metode ini digunakan sebagai proses yang mencoba menemukan partisi dari sebuah model. Unsupervised learning mempelajari bagaimana sebuah sistem dapat belajar memodelkan pola input dengan cara yang dapat menggambarkan struktur statistik dari pola input secara keseluruhan. Metode ini berbeda dengan pembelajaran terawasi, karena tidak memiliki tujuan keluaran yang jelas dan oleh karena itu sering digunakan untuk tujuan pengelompokan. (clustering).

Reinforcement learning sendiri merupakan metode pembelajaran untuk mempelajari aturan pengendalian dengan cara berinteraksi dengan lingkungan yang masih belum dikenal. Program akan menerima hukuman jika keputusannya salah dan hadiah apabila keputusannya benar. Pengalaman interaksi diakumulasikan sehingga program dapat membuat kesimpulan di masa depan dengan menggunakan pola yang telah dipelajari (E. Hasibuan & Karim, 2022).

2.1.10 Metode Regresi Linier

Analisis regresi terdiri dari satu variabel dependen (variabel terikat, variabel respon atau terikat) dan satu variabel independen (variabel bebas, variabel bebas atau prediktor), namun pada analisis regresi tidak menyatakan hubungan sebab akibat (kausalitas) (Rinaldo Fernandes & Solimun, 2022). Persamaan yang terdiri dari hanya satu variabel independen dan satu variabel dependen disebut persamaan regresi sederhana, sementara persamaan yang terdiri dari lebih dari satu variabel independen dan variabel dependen dinamakan persamaan regresi berganda. Regresi dapat dibedakan dalam dua jenis, yaitu regresi linier dan regresi non linier.

Regresi adalah teknik penyesuaian kurva untuk data dengan presisi rendah. Contoh data dengan presisi rendah adalah data dari observasi, eksperimen laboratorium, dan data statistik. Data seperti ini disebut data observasi dari objek yang kita amati. Kesalahan yang terdapat pada data timbul akibat ketidakteelitian alat ukur yang digunakan, kesalahan dalam membaca alat ukur (paralaks), atau akibat pengoperasian sistem pengukuran. Dalam regresi, penyesuaian kurva (curve fitting) dilakukan dengan membuat kurva fungsi yang mendekati titik-titik data (L. H. Hasibuan & Musthofa, 2022).

Regresi dapat dijadikan sebagai metode pengambilan keputusan yang biasa digunakan dalam mengembangkan model matematika dan statistic (Muriyatmoko, 2018). (Munte, 2022) menyebutkan bahwa regresi ini sangat erat kaitannya dengan korelasi, sehingga setiap regresi

pasti memiliki korelasi tetapi setiap korelasi belum tentu berlanjut selama regresi berlangsung.

Regresi mengukur pengaruh suatu variabel terhadap variabel lainnya, sehingga dapat digunakan untuk memprediksi nilai suatu variabel berdasarkan variabel lainnya (Fengki, 2018).

Regresi linier adalah alat statistik yang digunakan untuk mengetahui pengaruh satu atau lebih variabel terhadap suatu variabel (Fengki, 2018). Variabel yang berpengaruh biasanya disebut variabel bebas, variabel independen, atau variabel penjelas. Sedangkan variabel yang dipengaruhi disebut variabel terikat atau variabel dependen. Regresi linier hanya dapat digunakan pada skala interval dan rasio (Kusumawati, Marisa, & Wijaya, 2017). Secara umum regresi linier terdiri dari dua, yaitu regresi linier sederhana yaitu dengan satu buah variabel bebas dan satu buah variabel terikat dan regresi linier berganda dengan beberapa variabel bebas dan satu buah variabel terikat (Munte, 2022). Dikarenakan regresi linier dapat menghitung hubungan dan korelasi antar variabel yang dibutuhkan sehingga hasil prediksi lebih akurat, maka pemanfaatan modul phyton terkait dengan memanfaatkan sklearn (scikit-learn) yang cukup mudah digunakan karena menggunakan bahasa tingkat tinggi.

Definisi regresi linier menurut (Marbun, Sihotang, & Nababan, 2018) adalah salah satu metode yang digunakan di bidang manufaktur untuk meramalkan atau memprediksi karakteristik kualitas dan kuantitas. Dengan memperkirakan kombinasi produk yang berbeda, perusahaan diharapkan dapat memaksimalkan keuntungan dan mengestimasi kuantitas produksi yang sesuai, namun untuk menerapkannya diperlukan pengetahuan tentang bauran produk yang tepat, target pasar, dan tingkat permintaan terhadap produk tersebut (Indarwati et al., 2019)

Tujuan dari analisis regresi pada dasarnya adalah untuk menjelaskan variabel dependen dalam kaitannya dengan variabel independen. Kesesuaian dan keakuratan kesimpulan bergantung pada data yang digunakan. Oleh karena itu, data yang tidak representatif atau tidak teragregasi dengan baik akan menghasilkan kesimpulan yang tidak tepat.

Maka, untuk menggunakan analisis regresi secara efektif, seseorang harus mengikuti langkah-langkah berikut:

- 1 Menginvestigasi proses pengumpulan data
- 2 Menentukan batas-batas dari data yang dikumpulkan
- 3 Membatasi batasan kesimpulan yang sesuai

2.1.11 Regresi Linier Berganda

Model regresi linier berganda (Multi-Linear Regression, MLR) yaitu model yang menerangkan hubungan satu variabel dependen (dependent variable) dengan dua atau lebih variabel independen predictor (predictor variables) (Farizal, Rachman, & Al Rasyid, 2014). Regresi linier berganda sebagai perluasan dari regresi linier sederhana. Jika regresi linier sederhana mempersoalkan hubungan antara variabel dependen atau variabel kriteria (respon) dengan satu variabel independen prediktor (deterministik), maka regresi linier berganda mempersoalkan hubungan linier antara satu variabel dependen dengan beberapa variabel independen.

Variabel dependen dapat berupa ukuran atau kriteria keberhasilan, sedangkan variabel independen dapat berupa faktor-faktor yang menentukan keberhasilan. Untuk dapat melakukan prediksi melalui regresi, data untuk setiap variabel harus tersedia. Selain itu, berdasarkan data tersebut, peneliti dapat memperoleh persamaan regresi melalui perhitungan (Fengki, 2018).

MLR bisa digunakan sebagai alat bantu untuk berbagai macam tujuan, antara lain:

- 1 Mengestimasi nilai rerata dari variabel terikat berdasarkan nilai estimasi variabel yang digunakan,
- 2 Menguji sebuah hipotesis mengenai sifat ketergantungan antar variabel,
- 3 Memprediksi atau meramalkan nilai rerata variabel dependen berdasarkan nilai variabel independen yang berada di luar rentang sampel,
- 4 Mengetahui tingkat kontribusi (tingkat pengaruh) masing-masing variabel bebas dengan variabel yang diestimasi.

Langkah-langkah dalam melakukan pengimplementasian dengan regresi linier berganda adalah sebagai berikut:

- 1 Menentukan variabel bebas dan tidak bebas.
- 2 Menentukan nilai konstanta dan koefisien regresi.
- 3 Menginterpretasi koefisien regresi.
- 4 Menghitung koefisien determinasi
- 5 Menghitung koefisien korelasi berganda
- 6 Menghitung nilai standar Error Estimate
- 7 Menghitung nilai korelasi parsial, korelasi antara X_1 dan Y' , sehingga X_2 dianggap tetap. Korelasi antara X_2 dan Y' , X_1 dianggap tetap, korelasi antara X_1 dan X_2 dimana Y' dianggap tetap.

2.1.12 Pengujian Keakurasian

Akurasi dari suatu prediksi ditentukan oleh tingkat penyimpangan antara data yang diramalkan dengan data aktual (Susiana, 2015). Kesalahan dalam melakukan pemodelan prediksi tidak hanya disebabkan oleh faktor kesalahan saja tetapi juga disebabkan oleh ketidakmampuan model prediksi dalam mengenali unsur-unsur lain dalam deret data yang memberikan pengaruh terhadap tingkat kesalahan dalam prediksi. Kesalahan atau penyimpangan yang besar ini dapat disebabkan oleh banyaknya faktor yang tidak terduga (outlier) sehingga menyebabkan tidak ada metode prediksi yang mampu memberikan ramalan yang akurat, atau dapat juga disebabkan karena metode prediksi yang digunakan tidak dapat memprediksi secara akurat keadaan komponen tren, unsur musiman, komponen siklus yang dapat terjadi atau komponen yang terkandung dalam deret data (L. H. Hasibuan & Musthofa, 2022).

Ketika melakukan prediksi tentunya harus menghitung kesalahan prediksi (Ihzaniah et al., 2023). Beberapa metode yang digunakan yaitu mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), serta R-Squared (R^2). Berikut penjelasannya:

1 MSE

MSE adalah nilai rerata dari selisih kuadrat antara nilai prediksi dan nilai aktual yang sedang diamati. MSE memakai metode berbasis gradien, dimana semakin rendah nilainya, maka semakin baik prediksi yang dihasilkan (L. H. Hasibuan & Musthofa, 2022).

2 RMSE

RMSE atau Root Mean Square Error adalah hasil dari akar kuadrat rata-rata kuadrat antara nilai aktual dan nilai prediksi (Tatachar, 2021). RMSE sama seperti MSE, yaitu semakin rendah nilainya, maka semakin baik prediksi yang dihasilkan (L. H. Hasibuan & Musthofa, 2022).

3 MAE

Mean Absolute Error (MAE) merupakan rata-rata antara nilai aktual dengan nilai prediksi yang nilainya adalah mutlak positif (Azmi, Hadi, & Soraya, 2020).

4 R-Squared

R-Squared (R^2) disebut juga dengan koefisien determinasi, memiliki efek untuk mengetahui secara simultan hubungan antara variabel independen dengan

variabel dependen, atau dengan kata lain koefisien determinasi dapat mengetahui tingkat pengaruhnya terhadap variabel dependen. R-Square berkisar antara 0 sampai dengan 1 (Budilaksana, Sukarsa, & Wiranatha, 2021). Jika nilainya mendekati 1 dan tidak kurang dari 0,5 maka dapat dikatakan tingkat pengaruh variabel independen memiliki pengaruh yang kuat terhadap nilai variabel dependen dan sebaliknya. R-Squared menentukan proporsi varians dalam variabel dependen yang dapat dijelaskan oleh variabel independen. R squared memberikan kita goodness of fit (seberapa baik nilai yang diamati sesuai dengan nilai yang diprediksi) untuk model regresi (Tatachar, 2021).

2.1.13 Harga

Harga memegang peranan penting dalam menentukan pilihan pembeli dalam memutuskan untuk membeli barang atau jasa. Kotler dalam (Hidayat & Dewi Lubis, 2017) menerangkan bahwa harga merupakan ukuran uang yang diberikan pada barang atau jasa atau ukuran nilai uang yang dibeli untuk beberapa penggunaan sejak pembelian barang atau jasa. (Nasution, Limbong, & Ramadhan, 2020) juga menerangkan bahwa hakikat harga yang dibebankan oleh pelanggan atau pembeli adalah biaya yang dikeluarkan untuk memperoleh sesuatu. Harga menjadi salah satu faktor yang menentukan keberhasilan suatu bisnis karena menentukan keuntungan yang akan diperoleh bisnis dari penjualan produk, baik berupa barang maupun jasa. Menentukan harga yang terlalu tinggi juga akan mengakibatkan berkurangnya pendapatan, tetapi menentukan harga yang terlalu rendah juga akan berdampak terhadap keuntungan yang akan diperoleh pemilik bisnis. Harga memiliki pengaruh yang besar dalam menggambarkan kualitas merek suatu produk. Konsumen percaya bahwa harga yang tinggi sering kali sama dengan kualitas yang baik. Secara umum, harga memiliki pengaruh positif terhadap kualitas: semakin tinggi harga, semakin tinggi kualitas dan sebaliknya (Saputra, 2020). (Ehsani & Ehsani, 2015) menarik simpulan bahwa harga dapat dijadikan sumber daya untuk meningkatkan keuntungan dan kepuasan pelanggan.

Dijelaskan oleh Kotler dan Keller dalam (Tan, 2020), penetapan harga memiliki enam tujuan utama:

- 1 Perusahaan secara hati-hati membangun tujuan pemasarannya, seperti mempertahankan kelangsungan hidup, meningkatkan laba saat ini, keinginan untuk mendapatkan pangsa pasar atau kualitas produk.
- 2 Perusahaan menentukan kurva permintaan yang menunjukkan kuantitas produk

yang dapat dijual setiap periode, pada harga alternatif. Semakin tidak responsif permintaan, semakin tinggi harga yang dapat dibebankan perusahaan.

- 3 Perusahaan menghitung bagaimana biaya akan bervariasi pada berbagai tingkat produksi.
- 4 Perusahaan menggunakan harga pesaing untuk menetapkan harga mereka sendiri.
- 5 Perusahaan memilih salah satu metode penentuan harga termasuk penentuan harga cost- plus, analisis peluang utama dan penentuan target laba, penentuan harga nilai yang diperoleh, penentuan harga tingkat pertumbuhan, dan harga menurut siklus tertutup.
- 6 Perusahaan memilih harga akhir, menyatakannya dengan cara yang paling efektif secara psikologis, dan memeriksanya untuk memastikan bahwa harga tersebut konsisten dengan kebijakan penetapan harga perusahaan dan sesuai untuk para dealer, pedagang besar, dan pengecer. penjual, pesaing, pemasok, dan otoritas perusahaan.

Secara umum, ada empat metode penentuan harga diantaranya penetapan harga berbasis permintaan, berbasis biaya, berbasis laba, dan berbasis persaingan. Di bawah ini adalah beberapa penjelasan terkait metode penilaian (Saputra, 2020):

- 1 Berbasis permintaan

Pendekatan yang berfokus pada berbagai faktor yang mempengaruhi selera dan preferensi pelanggan berdasarkan kemampuan dan kemauan mereka untuk membeli, manfaat yang ditawarkan produk, dan perilaku konsumen secara umum.

- 2 Berbasis biaya

Faktor penentuan harga cenderung dipengaruhi oleh aspek penawaran atau biaya, bukan aspek permintaan. Harga akan ditetapkan sesuai dengan biaya produksi serta pemasaran produk, ditambah dengan jumlah tertentu untuk menutupi biaya langsung, operasional, dan laba/rugi

- 3 Berbasis laba

Penentuan harga dilakukan berdasarkan keseimbangan antara biaya dan pendapatan. Metode ini terdiri dari 3 pendekatan, yaitu *profit target pricing* (penetapan harga berdasarkan target laba), *sales profit target pricing* (penetapan harga berdasarkan penjualan), dan penetapan harga berdasarkan tujuan pengembalian investasi suatu bisnis.

4 Berbasis persaingan

Penentuan harga dilakukan dengan mengamati apa yang dilakukan oleh kompetitor. Metode ini memiliki tiga pendekatan, melalui sistem penjualan di bawah harga normal kompetitor untuk menarik konsumen, dengan menyamakan harga agar daya saing tidak terlalu kuat, atau dengan menetapkan harga yang lebih tinggi dari kompetitor. Pesaing percaya bahwa produk yang mereka sediakan memiliki kualitas yang lebih baik. Untuk proses penentuan harga, sebuah bisnis membutuhkan wawasan tentang keadaan produk, keuangan, dan hasil akhir dalam hal keuntungan atau margin yang akan diperoleh.

Penaksiran harga juga mempunyai tujuan tersendiri yang dijelaskan oleh Rahman dalam (Nasution et al., 2020) dibagi menjadi tiga arah, yaitu: (1) Pendapatan. Meskipun sebagian besar bisnis berfokus pada pendapatan, hanya bisnis nirlaba atau bisnis layanan publik yang biasanya berfokus pada titik impas; (2) Kapasitas. Bidang kegiatan tertentu sering kali menyesuaikan penawaran dan permintaan serta memanfaatkan kapasitas produksi mereka secara maksimal; (3) Pelanggan. Biasanya, harga yang ditawarkan cukup representatif dan dapat disesuaikan dengan semua jenis pelanggan, segmen pasar, dan perbedaan daya beli. Sistem yang diterapkan dapat berupa diskon, bonus, dan lainnya.

Adapun beberapa instrumen yang digunakan untuk melakukan pengukuran harga menurut perspektif Stanton dalam (Ofela & Agustin, 2016) adalah sebagai berikut:

1 Harga yang terjangkau

Harga yang terjangkau adalah harga aktual produk yang tertera dalam suatu produk yang harus dibayar oleh pelanggan. Hal ini berarti pelanggan cenderung melihat harga akhir dan memutuskan apakah mereka ingin mendapatkan nilai yang baik yang mereka harapkan.

2 Kesepakatan antara harga dan kualitas produk

Pertama-tama yaitu harga yang tertera pada produk, karena sebelum membeli, konsumen harus memikirkan sistem penghematan yang sesuai. Selain itu, konsumen mempertanyakan apakah harga yang ditawarkan sesuai dengan produk yang akan mereka beli

3 Persaingan harga

Harga yang ditawarkan oleh produsen atau penjual bervariasi dan bersaing

dengan harga yang ditawarkan oleh produsen lain untuk produk yang sama

4 Harga yang menguntungkan

Aspek penetapan harga ditentukan pihak produsen atau penjual berdasarkan manfaat yang dapat diperoleh konsumen dari produk yang dibeli.

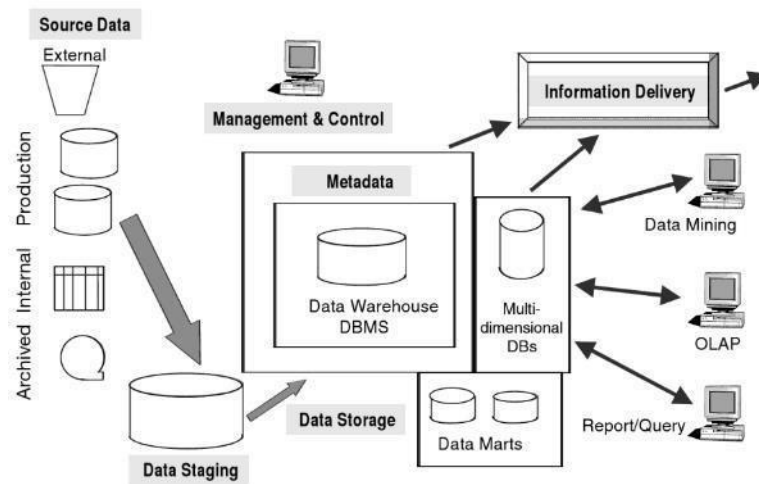
2.1.14 Data Warehouse

Data warehouse (DW) merupakan sekumpulan data yang berorientasi pada topik, terintegrasi, bervariasi dalam waktu, dan tidak berubah-ubah yang dimaksudkan untuk memberikan informasi yang dibutuhkan untuk pengambilan keputusan (Umam, Agung Wicaksono, & Purnomo, 2019).

Selain itu (Maryanto, Putranto, Kartadie, Guntara, & Saptoto, 2023) menyatakan bahwa data warehouse (DW) merupakan data warehouse yang mengekstrak, membersihkan, menyesuaikan, dan menyebarkan sumber data ke dalam simpanan data dimensional dan menggunakannya untuk dukungan kueri dan sistem penegakan hukum yang dirancang untuk menggunakan analisis dalam pengambilan keputusan.

Data warehouse sendiri memiliki karakteristik yang berbeda, antara lain (Muhaemin, 2019):

- 1 Berorientasi pada subjek yang artinya data warehouse didesain sedemikian rupa untuk dianalisis berdasarkan subjek yang spesifik di dalam organisasi, bukan didasarkan pada proses atau fungsi tertentu.
- 2 Terintegrasi secara eksplisit yang mampu melakukan penyimpanan data secara dekat dari sumber-sumber yang terpisah ke dalam sebuah format yang konsisten serta saling berhubungan.
- 3 *Time-variance* atau kurun waktu, terutama kurun waktu data yang berkaitan dengan akurasi data.
- 4 Data yang bersifat non volatile dalam *data warehouse* tidak dapat di *update* secara *real time* tetapi di-*refresh* dari operasi sistem secara berkala.



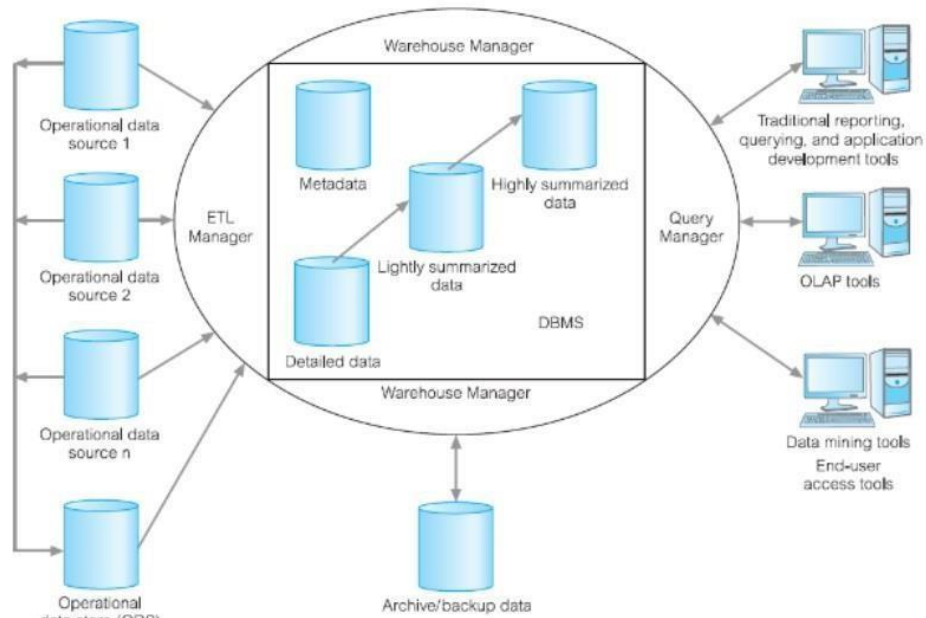
Gambar 2.1 Komponen Utama *Data Warehouse*

Sumber: (Ponniah, 2004)

Gambar 2.1 menunjukkan komponen utama dari data warehouse. Berikut adalah penjelasan terkait gambar di atas:

- 1 Berdasarkan jenisnya, sumber data dibagi menjadi beberapa jenis, yaitu data produksi, data internal, data arsip, serta data eksternal,
- 2 *Data Staging* memegang peranan dalam mengolah serta mengelola data mulai dari tahap *preparation* hingga tahap penyimpanan di data warehouse,
- 3 Penyimpanan data disimpan secara terpisah dari penyimpanan data untuk sistem operasi,
- 4 *Data Mart* memiliki kesamaan peran dengan data warehouse namun memiliki ruang lingkup yang terbatas,
- 5 Menyediakan informasi,
- 6 Metadata bertindak sebagai kamus data atau katalog data,
- 7 Manajemen dan kontrol adalah faktor yang paling penting.

2.1.15 Arsitektur Data Warehouse



Gambar 2.2 Arsitektur *Data Warehouse*

Sumber: (Connolly & Begg, 2005)

Arsitektur data warehouse sendiri terdiri dari data operasional, penyimpanan data operasional, pengelola ETL, pengelola gudang, pengelola kueri, data detail, ringkasan data, data arsip/cadangan, metadata, dan alat akses pengguna. Menurut Inmon dalam (Ganesha, 2018) menjelaskan bahwa data mengalir dari operasi ke data warehouse dengan transformasi tertentu, diikuti dengan alat akses sisi pengguna yang akan digunakan untuk analisis dan evaluasi aktivitas bisnis. Keunikan setiap organisasi dalam mencapai tujuannya akan menimbulkan strategi yang berbeda-beda. Ini juga menciptakan tipe data, tipe data, arsitektur, dan proses bisnis yang berbeda. Oleh karena itu, ketika merancang data warehouse, harus ditentukan arsitektur yang sesuai (Ganesha, 2018).

2.1.16 Manajemen *Data Warehouse*

Menurut (Puspita Dewi & Indra Sensuse, 2013) ada 3 (tiga) komponen utama dalam mengelola sebuah data warehouse, yaitu:

1 *Source System*

Sumber data atau objek yang merupakan sumber data dari keseluruhan proses. Bisa

berupa basis data, file, layanan direktori, XML atau lainnya.

2 ETL (*Extracting, Transformation, Loading*)

Tugas ETL adalah melakukan ekstraksi dari sumber data, mentransformasi data, kemudian mengirimkannya ke data store yang dituju.

a. *Extracting*

Ini adalah proses mendapatkan data dari berbagai sumber yang digunakan untuk memasukkannya ke dalam *data warehouse*. Tidak semua data operasional dimasukkan, hanya data penting saja. Pada langkah ini, pembersihan data juga dilakukan sebelum dilakukan transformasi.

b. *Transformation*

Ini adalah proses mengubah data dari format operasional ke format data warehouse. Ada 2 (dua) cara transformasi, yaitu fungsionalitas level record, fungsi ini bekerja terhadap serangkaian record seperti file maupun tabel melalui perintah *selection* dan *joining*. Yang kedua normalisasi dan agregasi melibatkan fungsionalitas level *field*, yang mengubah format data dari record sumber ke format yang berbeda pada record tujuan. Fungsi ini memiliki dua jenis yaitu: *Single Field* dan *Multi Field*.

c. *Loading*

Ini adalah langkah terakhir dalam mengisi *data warehouse*. Proses ini mengisikan data yang dipilih dan membuat indeks yang diperlukan. Ada dua cara untuk memuat data dari gudang, yaitu dengan menyegarkan atau menulis kembali semua data di gudang data secara berkala dan memperbarui atau menulis data yang diubah ke data sumber. Tidak ada proses penghapusan di sini. Mode refresh biasanya dipakai ketika memuat data untuk pertama kalinya ke dalam data warehouse, sedangkan mode update dipakai ketika memelihara dan menjalankan data warehouse.

3 Dimensional Data Store

Merupakan sumber dari hasil ETL yang akan digunakan sebagai dasar pengolahan data analitik dalam bentuk *business intelligence*, *reporting*, dan *query*.

2.1.17 *Extract, Transform, and Load (ETL)*

ETL adalah suatu alat integrasi data yang berguna untuk mengekstrak data dari sumber data eksternal (sistem operasi), memodifikasi data sesuai kebutuhan bisnis, dan terakhir memindahkannya ke data warehouse. ETL tidak hanya dapat memilih data tetapi juga menggabungkan kembali data terkait sebagai informasi berguna dalam data warehouse (Oslan & Kristanto, 2019). Kriteria dari sistem ETL yang baik adalah kemampuan untuk mengekstrak data dari sumber yang berbeda, menerapkan standar kualitas dan konsistensi yang sama, dan menganalisis data yang diambil dengan benar.

Di bawah ini adalah penjabaran singkat dari proses extract, transform, loading itu sendiri (Zikra & Idris, 2022):

1 *Extract*

Extract merupakan proses yang dilakukan untuk mengambil data. Proses ini dapat dilakukan dengan berbagai cara misalnya melalui *database*, *file*, *cloud* hingga dapat melalui *script*.

2 *Transform*

Transformasi data adalah proses dimana data dapat dipanggil ketika proses ekstraksi sudah selesai. Proses ini ditujukan untuk mengolah *data raw* yang berisi data dengan format yang berbeda sehingga membuat proses visualisasi menjadi kurang optimal.

3 *Loading*

Load yaitu tahapan dimana data yang sudah mengalami transformasi data pada tahap sebelumnya dimuat ke dalam penyimpanan data yang telah ditentukan. Data inilah yang bisa dimuat dalam Power BI melalui proses *apply* supaya data dapat tersimpan dalam *data warehouse* Power BI sehingga nantinya dapat dilakukan pembuatan *dashboard* visualisasi.

2.1.18 *Star Schema*

Star schema atau skema bintang merupakan jenis model dimensional dimana tabel fakta terletak di pusat sedangkan tabel-tabel dimensi berada di sekelilingnya. Pada umumnya, dalam skema bintang, tabel fakta adalah bentuk normalisasi ketiga dari database sementara tabel dimensi adalah bentuk normalisasi kedua dari database. Skema bintang ini termasuk bentuk

yang paling sederhana dari model dimensi. Ada beberapa keuntungan dari penggunaan skema bintang ini, antara lain mudah dimengerti serta rendah dalam hal pemeliharaan. Namun, ada pula kekurangan dari skema bintang ini, yakni jumlah tabel dimensi dapat menjadi sangat banyak serta performanya yang kurang baik jika diukur dengan skema lainnya (Tjung, Hugeng, & Sutrisno, 2023). Selain itu, (Connolly & Begg, 2010) menambahkan untuk skema bintang memiliki keuntungan dalam mempercepat proses *query* karena tabel dimensi telah didenormalisasi, sehingga mengurangi waktu yang dibutuhkan untuk proses *join*.

Skema Bintang terdiri dari satu atau lebih tabel fakta yang saling terhubung dengan sejumlah tabel dimensi. Definisi tabel fakta serta tabel dimensi dijelaskan sebagai berikut:

- 1 Tabel fakta memuat tentang informasi data primer yang digunakan dalam proses analisis, juga memuat tentang primary key yang berhubungan dengan tabel dimensi. Data dalam tabel fakta harus berupa angka agar mudah dikelompokkan.
- 2 Tabel dimensi berisi data yang bersifat deskriptif dan digunakan sebagai parameter dalam query (Connolly & Begg, 2010).

Menurut (Ponniah, 2001), skema bintang terdiri dua jenis tabel, yaitu dimensi dan tabel fakta, dan kedua tabel tersebut mempunyai karakteristik sebagai berikut:

- 1 Tabel Dimensi
 - a. Kunci tabel dimensi, adalah kunci utama dari tabel dimensi yang secara unik mengidentifikasi setiap baris dalam tabel.
 - b. Merupakan tabel dengan ukuran yang lebar. Tabel dimensi mempunyai banyak kolom atau atribut, oleh karena itu tabel dimensi berbentuk lebar.
 - c. Atribut berbentuk teks. Pada tabel dimensi, jarang sekali ditemukan nilai angka sebagai perhitungan, atribut biasanya berisi teks yang menyajikan deskripsi secara tekstual mengenai berbagai komponen dalam dimensi bisnis.
 - d. Atribut tidak berkaitan satu sama lain secara langsung.
 - e. Tidak dinormalisasikan. Untuk performa *query* yang lebih efektif, akan lebih baik jika *query* tersebut diambil dari tabel dimensi kemudian langsung menuju ke tabel fakta tanpa melalui perantara tabel yang terbentuk apabila tabel dimensi dinormalisasikan.
 - f. Kemampuan *drill-down* dan *roll-up*. Atribut pada tabel dimensi memberikan kemudahan untuk memperoleh detail dari tingkat agregat yang tinggi ke tingkat detail yang rendah.
 - g. Hirarki yang beragam. Bagian yang berbeda dari perusahaan dapat menyusun

dimensi dengan cara yang berbeda, sehingga membentuk satu hirarki.

- h. Jumlah catatan yang lebih sedikit. Tabel dimensi biasanya mempunyai jumlah *record* atau baris yang cenderung lebih sedikit dibandingkan dengan tabel fakta.

2 Tabel Fakta

- a. Kunci gabungan. Baris pada tabel fakta dikenali dengan memakai kunci primer dari tabel dimensi, sehingga kunci primer dari tabel fakta adalah kunci primer gabungan dari seluruh tabel dimensi.
- b. Data *Grin*, adalah level detail untuk sebuah pengukuran. Contohnya, kuantitas pesanan berkaitan erat dengan kuantitas produk tertentu atas suatu pesanan, tanggal tertentu, bagi pelanggan tertentu, dan didapatkan melalui perwakilan penjualan spesifik. Jika jumlah pesanan dipandang sebagai kuantitas produk per bulan, berarti data *grin* berbeda dan berada di level yang lebih tinggi.
- c. *Fully additive measure*. Penjumlahan dari *fully additive measure* dilakukan melalui agregasi sederhana dari nilai atribut.
- d. Ukuran semiaditif (*semiadditive measure*). *Semiadditive measure* adalah nilai yang tidak dapat secara langsung dijumlahkan, contohnya adalah persentase keuntungan.
- e. Tabel besar, bukan berarti lebar. Tabel fakta biasanya memiliki atribut yang lebih sedikit dibandingkan tabel dimensi, tetapi mempunyai catatan dalam jumlah yang lebih banyak.
- f. *Sparse data*. Tabel fakta tidak harus menyimpan *record* dengan nilai nol. Jadi tabel fakta dapat memiliki celah.
- g. *Degenerate dimensions*. Ada elemen data dalam sistem operasional yang bukan termasuk fakta atau dimensi, misalnya nomor pesanan, nomor faktur, dan sebagainya. Akan tetapi, atribut ini dapat berguna dalam beberapa jenis analisis. Contohnya, untuk menemukan jumlah rata-rata produk per pesanan, produk harus dikaitkan dengan nomor pesanan guna memperoleh nilai rata-rata. Atribut-atribut ini disebut dimensi yang dihasilkan oleh sistem serta disimpan dalam tabel fakta.

2.1.19 Visualisasi

Disampaikan oleh Sedigheh dan Ziaei dalam (Malik, 2023) visualisasi adalah proses menciptakan representasi grafis atau gambaran mental dari data atau informasi yang kompleks atau abstrak. Visualisasi sering digunakan untuk membantu memperjelas konsep, membantu pemahaman, atau mendukung pengambilan keputusan dengan menampilkan data secara visual. Teknik visualisasi dapat digunakan di berbagai bidang, seperti bisnis, sains, teknologi, arsitektur, dan seni. Contoh teknik visualisasi termasuk diagram, grafik, peta, gambar, dan animasi. Visualisasi biasanya digunakan di berbagai bidang, seperti bisnis, sains, teknologi, arsitektur, dan seni. Teknik visualisasi dapat digunakan untuk memvisualisasikan berbagai jenis data, mulai dari data numerik, kualitatif, hingga data abstrak. Tujuan dari visualisasi adalah untuk membuat informasi yang kompleks atau membingungkan menjadi lebih mudah dipahami melalui representasi visual yang mudah dimengerti. Visualisasi dapat membantu mengambil keputusan, menjelaskan konsep, atau menyajikan informasi dengan cara yang lebih menarik dan interaktif.

Nugroho dan Fatchurrohman dalam (Malik, 2023) menambahkan bahwa visualisasi data biasanya dimulai dengan mengimpor data mentah atau file data, kemudian dilakukan pembersihan data dan pra-pemrosesan untuk mendapatkan olah lebih lanjut. Pengguna juga dapat memilih jenis tampilan yang sesuai dengan data yang dimiliki seperti diagram batang, diagram garis, peta panas, dan lain-lain. Pengguna dapat menambahkan filter dan parameter untuk mendukung analisis data dan membuat visualisasi data untuk menyajikan hasil analisis secara interaktif. Pengguna dapat menyajikan beberapa visualisasi data secara bersamaan dan mengatur interaksi antara satu visualisasi dengan visualisasi lainnya (Malik, 2023).

Visualisasi sendiri menurut (Darman, 2018) ialah suatu teknik yang melibatkan pembuatan gambar, diagram, atau animasi untuk menampilkan suatu informasi. Terdapat dua konsep visualisasi, yaitu visualisasi ilmiah dan visualisasi informasi. Kedua konsep tersebut membuat model grafis dan representasi visual dari data yang secara langsung berinteraksi dengan pengguna untuk mengeksplorasi dan mendapatkan wawasan yang terkandung dalam data visualisasi ilmiah. Pada umumnya, model grafis dibangun dari data pengukuran atau simulasi yang merepresentasikan objek atau konsep yang berkaitan dengan fenomena nyata.

Tujuan utama dilakukannya visualisasi data menurut (Siswanto & Asmunin, 2016) adalah untuk menginformasikan secara jelas serta efektif kepada pengguna melalui grafik informasi yang dipilih, seperti tabel dan bagan. Visualisasi yang efektif membantu pengguna

menganalisis dan bernalar tentang data dan bukti. Hal ini membuat data yang kompleks dapat diakses, dimengerti, dan berguna. Pengguna dapat melakukan tugas analisis tertentu, seperti membandingkan atau memahami hubungan sebab-akibat dan prinsip-prinsip desain diagram (misalnya, menunjukkan perbandingan atau hubungan sebab-akibat) dari misi ini. Tabel sering digunakan ketika pengguna ingin melihat pengukuran spesifik dari sebuah variabel, sementara grafik dari berbagai jenis digunakan untuk memvisualisasikan pola atau hubungan dalam data satu atau lebih variabel.

2.1.20 Power BI

Power BI dikembangkan dan dirilis oleh Microsoft pada bulan September 2013 sebagai Power BI untuk Office 365. Pada tanggal 24 Juli 2015 Power BI pertama kali dirilis ke publik. Power BI adalah alat analisis data berbasis cloud yang dapat digunakan untuk menganalisis dan melaporkan data dari berbagai sumber data. Power user dan analis bisnis dapat menggunakan Power BI karena kemudahan penggunaannya. Untuk memodelkan data yang kompleks dan skenario agregat dalam sistem bisnis, pengembang BI dapat menggunakan Power BI yang canggih dan kuat (Gowthami & Kumar, 2017).

(Darman, 2018) memaparkan bahwa Microsoft Power BI merupakan perangkat lunak business intelligent atau sekumpulan alat analisis bisnis yang dapat meningkatkan wawasan ke dalam suatu instansi atau organisasi. Microsoft Power BI dapat terhubung ke ratusan sumber data, menyederhanakan persiapan data dan mendorong analisis adhoc. Laporan yang dihasilkan dapat ditampilkan di web atau perangkat seluler dan dapat membuat dasbor yang dipersonalisasi dengan tampilan 360 derajat. Terdapat berbagai macam bentuk visualisasi grafik yang dapat digunakan pada Microsoft Power BI diantaranya adalah *stacked bar chart*, *stacked column chart*, *clustered bar chart*, *clustered column chart*, *line chart*, *area chart*, *stacked area chart*, *ribbon chart*, *pie chart*, *donut chart*, *treemaps*, dan lain-lain. Microsoft Power BI terintegrasi dengan Bing Maps untuk menyediakan koordinat peta default (proses yang digunakan *geo-coding*) sehingga dapat membuat tampilan peta yang dilengkapi dengan algoritma untuk mengidentifikasi lokasi yang benar. Microsoft Power BI dapat menentukan bidang geografis berdasarkan bidang yang telah dikodekan secara geografis dengan menetapkan data kategori pada bidang data, dengan cara: memilih tabel yang diinginkan, kemudian masuk ke *Advanced Ribbon* dan mengatur data kategori menjadi alamat, kota, benua, negara, kode pos, negara bagian, atau provinsi. Kategori akan membantu Bing mengodekan tanggal dengan benar.

Power BI dapat dilakukan dengan mudah dan cepat, memberikan wawasan dari database lokal maupun file Excel. Selain itu, Power BI juga dapat diandalkan, dengan kualitas terbaik, yang dilengkapi dengan model terperinci, pemrograman khusus, dan analitik waktu nyata. Dengan pendekatan ini, Power BI dapat berfungsi untuk menjadi alat analisis dan pengambilan keputusan untuk tim, departemen, atau proyek perusahaan, selain menjadi alat pelaporan dan visualisasi individu. Perangkat lunak, aplikasi, dan layanan konektivitas yang membentuk Power BI. Power BI merupakan sumber daya bagi siapa saja yang tertarik dengan analitik dan pengambilan keputusan serta alat pelaporan dan visualisasi. Dengan menggunakan Power BI, data simulasi dari berbagai sumber diubah menjadi format data yang menarik dan dapat dibagikan (Bororing, 2022)

Fungsi utama Power BI dalam visualisasi data adalah:

- 1 Power BI memberikan kemudahan bagi pengguna untuk menghubungkan sekaligus menggabungkan data dari berbagai sumber, termasuk *database relational*, file Excel, *data cloud*, layanan web, dan berbagai sumber data lainnya.
- 2 Alat transformasi data Power BI berfungsi untuk membersihkan dan mengatur data. Operasi seperti menghapus kolom yang tidak perlu, mengganti nilai yang hilang atau kosong, atau menggabungkan data dari beberapa tabel menjadi satu.
- 3 Power BI mampu membuat model data yang kompleks dengan menggabungkan tabel, mendefinisikan hubungan antar tabel, hingga membuat kolom kalkulasi menggunakan formula DAX (*Data Analysis Expressions*).
- 4 Visualisasi yang dibuat di Power BI dapat di filter secara interaktif langsung pada visualisasi tersebut untuk penelitian data yang lebih mendalam.
- 5 Interaksi antar visualisasi dapat dilakukan secara langsung di dalam visualisasi, di mana perubahan pada satu visualisasi mempengaruhi visualisasi lainnya secara *real time*.
- 6 Power BI dapat dimanfaatkan sebagai alat untuk menyesuaikan visualisasi dan deskripsi data berdasarkan kebutuhan brand atau bisnis. Selain itu, Power BI dapat diintegrasikan dengan berbagai aplikasi bisnis seperti Microsoft Office 365, SharePoint, dan Azure untuk meningkatkan fungsionalitas dan kolaborasi (Jusuf, 2023).

Microsoft Power BI menyediakan 3 konsep operasional yang akan sangat berguna untuk menganalisis data seperti dashboard, laporan, dan dataset. Data set adalah kumpulan data yang diimpor atau terhubung ke Power BI. Sementara itu, laporan terdiri dari satu atau lebih halaman visual. Laporan bisa dalam bentuk tabel atau bagan, dan dasbor itu sendiri adalah tampilan terintegrasi yang menampilkan kumpulan laporan dari sekumpulan set data. Dashboard memberikan wawasan data, analitik, dan memberikan gambaran umum dalam satu visualisasi dashboard (Febriansyah, 2022).

Menurut (Gowthami & Kumar, 2017), komponen Power BI terdiri dari:

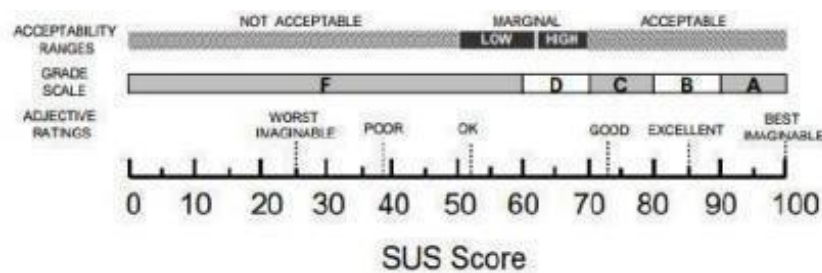
- 1 *Power Query*: *Power Query* adalah alat ETL (*Extract, Transform, Load*) *self-service* yang menjalankan *add-in* Excel. Mengambil data dari berbagai sumber, memanipulasinya, mengubahnya menjadi bentuk, dan memuatnya ke dalam Excel.
- 2 *Power Pivot*: Komponen pemodelan data dalam memori yang memudahkan dalam melakukan agregasi data dan komputasi penyimpanan yang sangat cepat.
- 3 *Power View*: *Power View* adalah alat visualisasi interaktif berupa antarmuka seret dan lepas yang berfungsi untuk membuat visualisasi data di buku kerja Excel dengan cepat dan mudah.
- 4 *Power Map*: *Power Map* untuk Excel adalah alat visualisasi data tiga dimensi (3D) yang memudahkan pengguna untuk mendapatkan gambaran besar. Sajikan informasi dengan cara baru. Peta sumber memberi Anda informasi yang tidak dapat diberikan oleh tabel dan bagan dua dimensi (2D) tradisional.
- 5 *Power Q&A*: Mengidentifikasi kata-kata yang telah diketikkan dan menentukan di mana letak jawabannya. Q&A membantu menyusun pertanyaan menggunakan parafrase, pelengkapan otomatis, dan teks serta alat bantu visual lainnya
- 6 *Power BI Desktop*: Membantu melakukan analisis dengan laporan visual. Seret dan letakkan konten tepat di tempat yang diinginkan pada kanvas yang fleksibel dan fleksibel. Jelajahi model dengan cepat sekaligus menemukan tampilan terpadu dari elemen interaktif yang terlibat dalam visualisasi.

Power BI mempunyai beberapa fitur seperti *dashboard*, visualisasi, Connector for Saas service, Live connectivity to SSAS service, dan Power BI Designer (Bororing, 2022).

2.1.21 System Usability Scale (SUS)

System Usability Scale (SUS) adalah kuesioner yang praktis digunakan untuk mengukur hasil pengujian *usability* sistem dari sudut pandang masing-masing pengguna. Kuesioner ini menggunakan skala Likert dengan lima poin: “Sangat tidak setuju (STS)”, “Tidak setuju (TS)”, “Netral (N)”, “Setuju (S)”, dan “Sangat setuju (ST)”. Ada beberapa karakteristik yang membuat SUS menarik dan berbeda dari kuesioner lainnya. Pertama, SUS terdiri dari sepuluh pertanyaan, sehingga relatif cepat dan mudah untuk diisi oleh responden. Kedua, SUS bersifat teknologi agnostik, yang berarti dapat diterapkan secara luas untuk mengevaluasi hampir semua jenis antarmuka, termasuk situs web, aplikasi smartphone, sistem respon suara interaktif (IVR), sistem (*touch-tone dan speech*), televisi, dan lain-lain. Ketiga, hasil dari kuesioner ini adalah nilai tunggal yang berkisar antara 0 hingga 100, dan mudah dipahami oleh berbagai disiplin ilmu, baik oleh individu maupun kelompok (Sidik, 2018).

Perhitungan SUS menggunakan metode kuantitatif dengan rumus yang telah ditentukan. Untuk pertanyaan dengan nomor ganjil, skor dikurangi 1, sedangkan untuk pertanyaan dengan nomor genap, nilai 5 dikurangi dengan jawaban yang diberikan oleh responden. Total skor dihitung dengan menjumlahkan nilai akhir dari setiap pertanyaan, lalu hasilnya dikalikan dengan 2,5 (Herawati & Azahra, 2024).



Gambar 2.3 Skor SUS

BAB III

METODOLOGI PENELITIAN

3.1 Studi Literatur

Adapun beberapa penelitian terdahulu yang dapat digunakan sebagai referensi dalam penelitian ini. Penelitian terdahulu juga sebagai upaya untuk mendapatkan aspek pembandingan yang kemudian dapat dijadikan langkah untuk menemukan inspirasi untuk melakukan penelitian ini.

Penelitian pertama dilakukan oleh (Baihaqi & Dianingrum, 2019) yang berjudul "Regresi Linier Sederhana untuk Memprediksi Kunjungan Pasien di Rumah Sakit Berdasarkan Jenis Layanan dan Umur Pasien" bertujuan untuk memprediksi jumlah kunjungan pasien RSUD Cilacap menggunakan regresi linier sederhana. Penelitian ini dilakukan dengan melakukan pengumpulan data kunjungan rawat jalan, rawat inap, dan rawat darurat selama 7 tahun dari tahun 2011 hingga 2017. Data yang dipakai dalam penelitian ini adalah data primer yang diambil dari bagian SIM-RS dan rekam medis RSUD Cilacap. Regresi linier sederhana digunakan untuk memprediksi kunjungan pasien berdasarkan jenis layanan dan usia pasien. Hasil prediksi kunjungan pasien dapat membantu manajemen rumah sakit dalam perencanaan dan pengembangan kebijakan.

Penelitian kedua dilakukan oleh (Bengnga & Ishak, 2018) yang berjudul "Prediksi Jumlah Mahasiswa Registrasi Per Semester Menggunakan Linier Regresi Pada Universitas Ichsan Gorontalo". Penelitian ini bertujuan untuk meningkatkan kualitas pendidikan dengan memberikan solusi untuk memprediksi jumlah pendaftar sesuai program studi. Metode Data Mining digunakan dengan data dari gudang database besar. Hasilnya menunjukkan bahwa dari 2 prodi yang dipilih yakni prodi Teknik Informatika didapatkan hasil tingkat error 4.24% atau tingkat akurasi 95.76%, dan untuk prodi Ilmu Hukum didapatkan tingkat error 7.69% atau tingkat akurasi 92.31%, penelitian juga membahas faktor-faktor yang mempengaruhi pembayaran SPP dan teknik prediksi menggunakan regresi linier untuk mengatasi masalah pendaftaran mahasiswa dan tren penurunan.

Penelitian ketiga dilakukan oleh (Sri Mulyani et al., 2020) yang berjudul "Estimasi Harga Jual Mobil Bekas Menggunakan Metode Regresi Linier Berganda". Penelitian ini bertujuan untuk melakukan pengestimasi harga jual mobil bekas menggunakan metode regresi linier berganda dengan tujuan untuk membangun sebuah sistem untuk mengestimasi harga jual mobil

bekas pada Showroom Gulur Motor. Pada penelitian ini dijabarkan langkah- langkah yang dilakukan dalam penelitian, mulai dari pengumpulan data, pengolahan data hingga pengujian sistem. Variabel yang dipakai dalam metode regresi linier berganda untuk mengestimasi harga jual mobil bekas adalah tahun perakitan, harga beli dan harga jual baru. Dengan demikian, hasil penelitian menunjukkan bahwa sistem estimasi harga jual mobil bekas dengan menggunakan metode regresi linier berganda dapat mengestimasi harga jual mobil bekas secara akurat, serta dapat menghemat waktu dan tenaga dalam proses akuntansi secara manual. Namun, penelitian ini masih memiliki kekurangan dan belum sempurna seperti penggunaan variabel yang terbatas dan ukuran sampel yang kecil.

Penelitian keempat dilakukan oleh tentang (Rahmawati, Kristanto, Setya Pratama, & Abiansa, 2022) yang berjudul “Prediksi Pelaku Perjalanan Luar Negeri Di Masa Pandemi COVID-19 Menggunakan Metode Regresi Linier Sederhana”. Penelitian ini bertujuan untuk memprediksi jumlah pengunjung asing ke Indonesia selama pandemi COVID-19 dengan menggunakan metode regresi linier sederhana. Penelitian ini menggunakan data primer dan data sekunder dengan jumlah sampel yang representatif. Hasil analisis memperlihatkan bahwa terdapat hubungan yang moderat antara jumlah kunjungan wisatawan mancanegara dan jumlah kasus Covid-19. Namun, dampak dari jumlah kasus Covid-19 terhadap jumlah kunjungan wisatawan mancanegara dapat diabaikan

Penelitian kelima dilakukan oleh (Pohan & Dar, 2022) yang berjudul “Penerapan Data Mining untuk Prediksi Penjualan Produk Sepatu Terlaris Menggunakan Metode Regresi Linier Sederhana”. Penelitian ini bertujuan untuk memprediksi penjualan produk sepatu terlaris. Penelitian ini dilakukan di PT. Bata Shoes yang bertujuan untuk mengoptimalkan penggunaan data transaksi penjualan yang ada. Pada penelitian ini digunakan algoritma regresi linier sederhana untuk menganalisis hubungan antara persediaan produk (variabel independen) dan penjualan (variabel dependen). Hasil dari prediksi penjualan dapat digunakan sebagai informasi yang berharga untuk mengambil keputusan tentang berapa banyak stok produk sepatu yang akan dijual pada bulan yang akan datang.

Penelitian keenam dilakukan oleh (Farizal et al., 2014) yang berjudul “Model Peramalan Konsumsi Bahan Bakar Jenis Premium di Indonesia dengan Regresi Linier Berganda”. Penelitian ini bertujuan untuk memprediksi konsumsi bahan bakar premium di Indonesia. Penelitian ini memakai regresi linier berganda (MLR) dengan 10 variabel prediktor. Variabel-variabel tersebut antara lain inflasi, selisih harga antara Pertamina dan Premium, pertumbuhan ekonomi, jumlah mobil, jumlah sepeda motor, jumlah angkutan umum, jumlah penduduk,

infrastruktur (jarak tempuh), jumlah subsidi, dan PDB. Hasilnya didapatkan model MLR berhasil memprediksi konsumsi premium tahun 2013 dengan tingkat kesalahan hanya 1,05%, mendekati konsumsi aktual 29,26 juta liter. MAPE sebesar 5,18%, menunjukkan keandalan model dalam ramalan.

Perbandingan penelitian terdahulu dengan penelitian yang akan di teliti adalah penelitian ini mencoba untuk mengadaptasi dengan metode yang sama yaitu metode regresi linier namun dengan domain yang berbeda yaitu prediksi harga mobil bekas. Tujuan utama dari penelitian ini adalah untuk memperoleh model prediksi harga mobil bekas yang akurat serta membuktikan efektivitas dari metode regresi linier itu sendiri dalam konteks tersebut. Prediksi ini bermanfaat untuk mengatasi kendala dalam menentukan harga pada proses jual beli, serta memberikan solusi akurat untuk menghasilkan sebuah keputusan kepada pihak yang terlibat terkhusus pemilik bisnis.

Tabel 3.1 Analisis Perbandingan Literatur

No	Judul	Metode	Hasil
1.	Regresi Linier Sederhana untuk Memprediksi Kunjungan Pasien di Rumah Sakit Berdasarkan Jenis Layanan dan Umur Pasien (Baihaqi & Dianingrum, 2019)	Prediksi dengan Algoritma Regresi Linier Sederhana	Hasil prediksi kunjungan pasien dapat membantu manajemen rumah sakit dalam perencanaan dan pengembangan kebijakan
2.	Prediksi Jumlah Mahasiswa Registrasi Per Semester Menggunakan Linier Regresi Pada Universitas Ichsan Gorontalo (Bengnga & Ishak, 2018)	Prediksi dengan Algoritma Regresi Linier	Hasil penelitian dari 2 prodi yang dipilih yakni prodi Teknik Informatika didapatkan hasil tingkat error 4.24% atau tingkat akurasi 95.76%, dan untuk prodi Ilmu Hukum didapatkan tingkat error 7.69% atau tingkat akurasi 92.31%
3.	Estimasi Harga Jual Mobil Bekas Menggunakan Metode Regresi Linier Berganda	Prediksi dengan Algoritma Regresi Linier Berganda	Menunjukkan bahwa sistem estimasi harga jual mobil bekas dengan

No	Judul	Metode	Hasil
	(Sri Mulyani et al., 2020)		menggunakan metode regresi linier berganda dapat mengestimasi harga jual mobil bekas secara akurat dan menghemat waktu dan tenaga dalam proses akuntansi secara manual. Namun, penelitian ini masih memiliki kekurangan dan belum sempurna seperti penggunaan variabel yang terbatas dan ukuran sampel yang kecil
4.	Prediksi Pelaku Perjalanan Luar Negeri Di Masa Pandemi COVID-19 Menggunakan Metode Regresi Linier Sederhana (Rahmawati et al., 2022)	Prediksi dengan Algoritma Regresi Linier Sederhana	Hasil analisis memperlihatkan bahwa terdapat hubungan yang moderat antara jumlah kunjungan wisatawan mancanegara dan jumlah kasus Covid-19.
5.	Penerapan Data Mining untuk Prediksi Penjualan Produk Sepatu Terlaris Menggunakan Metode Regresi Linier Sederhana (Pohan & Dar, 2022)	Prediksi dengan Algoritma Regresi Linier Sederhana	Hasil dari prediksi penjualan dapat digunakan sebagai informasi yang berharga untuk mengambil keputusan tentang berapa banyak stok produk sepatu yang akan dijual pada bulan yang akan datang.
6.	Model Peramalan Konsumsi Bahan Bakar Jenis Premium di Indonesia dengan Regresi Linier Berganda	Prediksi dengan Algoritma Regresi Linier Berganda	Model MLR (Model Regresi Linier Berganda) berhasil memprediksi konsumsi premium tahun

No	Judul	Metode	Hasil
	(Farizal et al., 2014)		2013 dengan tingkat kesalahan hanya 1,05% mendekati konsumsi aktual 29,26 juta liter MAPE sebesar 5,18%, menunjukkan keandalan model dalam ramalan

3.2 Populasi dan Sampel

Penelitian ini menggunakan data sekunder yang diperoleh dari situs web Kaggle. 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. Populasi yang digunakan pada penelitian ini hanya mencakup mobil bekas di India. Sedangkan sampel yang digunakan adalah data mobil bekas di seluruh India yang diambil secara acak dari Kaggle.com berdasarkan kriteria tertentu (variabel yg dipakai).

3.3 Pengumpulan Data

Data yang digunakan adalah data daftar mobil bekas yang ada di wilayah India. Berikut adalah linknya (<https://www.kaggle.com/code/iabhishekmaurya/used-car-price-prediction>) diakses pada tanggal 26 Februari 2023. Jenis data yang digunakan dalam penelitian ini sendiri merupakan data sekunder, dimana data atau informasi yang diambil dari website Kaggle. Data berisikan data latih serta data uji. Data latih memiliki 6019 baris dan 14 kolom, dengan rentang waktu dari 1998 hingga 2019. Atribut dataset meliputi 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. Sementara itu, data uji memiliki 1234 baris dan 13 kolom, dengan rentang waktu dari 1996 hingga 2019. Atribut datasetnya yaitu nomor, nama mobil, lokasi, tahun, kilometer mobil, jenis bahan bakar, transmisi, tipe pemilik, konsumsi bbm, kapasitas mesin, tenaga, jumlah kursi, dan harga bekas. Gambar 3.1 adalah tampilan beberapa sampel dari dataset data latih serta gambar 3.2 adalah tampilan beberapa sampel dari dataset data uji.

	Unnamed: 0	Nama Mobil	Lokasi	Tahun	Kilometer Mobil	Jenis Bahan Bakar	Transmisi	Tipe Pemilik	Konsumsi BBM	Kapasitas Mesin	Tenaga	Jumlah Kursi	Harga Baru	Harga Bekas
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.75
1	1	Hyundai Creta 1.6 CRDI SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	NaN	6.00
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17.74

Gambar 3.1 Sample dataset uji yang digunakan

	Unnamed: 0	Nama Mobil	Lokasi	Tahun	Kilometer Mobil	Jenis Bahan Bakar	Transmisi	Tipe Pemilik	Konsumsi BBM	Kapasitas Mesin	Tenaga	Jumlah Kursi	Harga Baru
0	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	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	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	3	Toyota Etios Liva GD	Hyderabad	2012	139000	Diesel	Manual	First	23.59 kmpl	1364 CC	null bhp	5.0	NaN
4	4	Hyundai i20 Magna	Mumbai	2014	29000	Petrol	Manual	First	18.5 kmpl	1197 CC	82.85 bhp	5.0	NaN

Gambar 3.2 Sample dataset latih yang digunakan

3.4 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah:

- 1 Variabel dependen merupakan variabel terikat yang dipengaruhi oleh variabel independen. Variabel dependen dalam penelitian ini yaitu Harga Bekas dengan satuan skala numerik
- 2 Variabel independen merupakan variabel yang memberikan pengaruh terhadap variabel dependen. Adapun variabel independent dalam penelitian ini adalah Nama Mobil, Lokasi, Tahun, Kilometer Mobil, Jenis Bahan Bakar, Transmisi, Tipe Pemilik, Konsumsi BBM, Kapasitas Mesin, Tenaga, Jumlah Kursi, Harga Baru

Tabel 3.2 Definisi Operasional Variabel (DOV)

No	Nama Variabel	DOV (Definisi Operasional Variabel)	Satuan atau skala	Contoh
1.	Unnamed: 0	Penomoran seperti pada umumnya	Numerik	0
2.	Nama Mobil	Nama-nama mobil bekas yang dijual	Kategorikal	Toyota Fortuner
3.	Lokasi	Lokasi atau tempat mobil itu dijual	Kategorikal	Pune
4.	Tahun	Tahun mobil bekas itu dijual	Numerik	2015

5.	Kilometer Mobil	Jarak yang telah ditempuh mobil yang akan dijual	Numerik	59000
6.	Jenis Bahan Bakar	Jenis bahan bakar mobil itu sendiri	Kategorikal	Diesel
7.	Transmisi	Transmisi yang digunakan mobil bekas tersebut	Kategorikal	Automatic
8.	Tipe Pemilik	Kepemilikan mobil bekas yang akan dijual tersebut milik tangan beberapa	Numerik	First
9.	Konsumsi BBM	Konsumsi bahan bakar itu mobil bekas itu sendiri	Kategorikal	12.55 kmpl
10.	Kapasitas Mesin	Kapasitas mesin dari mobil bekas yang akan dijual	Kategorikal	2982 CC
11.	Tenaga	Tenaga yang terdapat pada mobil bekas tersebut	Kategorikal	168.7 bhp
12.	Jumlah Kursi	Jumlah banyaknya kursi yang tersedia pada mobil bekas tersebut	Numerik	5.0
13.	Harga Baru	Harga baru ketika kondisi mobil dengan kondisi benar-benar baru (tidak semua mobil memiliki data tersebut)	Kategorikal	21 Lakh
14.	Harga Bekas	Harga bekas yang ditawarkan oleh mobil-mobil bekas yang akan dijual	Numerik	17.5

3.5 Perangkat Penelitian

Perangkat lunak dan perangkat keras yang digunakan peneliti dalam penelitian ini yaitu:

Tabel 3.3 Daftar Perangkat Lunak

No	Perangkat Lunak	Kegunaan
1.	Python	Bahasa Pemrograman untuk membuat model prediksi
2	Google Colaboratory	Menulis, meng-edit code, serta melakukan pre- processing data olah
3	Microsoft Excel	Penyimpanan data latih dan data uji

4	Microsoft Power BI	Visualisasi data hasil prediksi dan data analisis deskriptif
5	Google Chrome	Mencari, mengobservasi, menemukan data yang nantinya akan digunakan

Tabel 3.4 Daftar Perangkat Keras

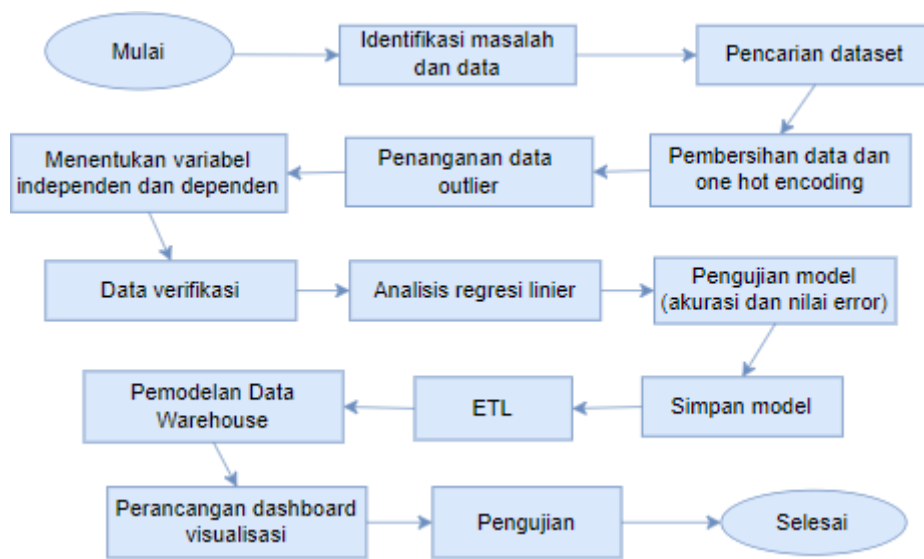
No	Komponen	Spesifikasi
1.	<i>Processor</i>	Intel(R) Core (TM) i3-1115G4 @ 3.00GHz 3.00 GHz
2.	RAM	8,00 GB
3.	<i>OS Type</i>	Windows 11 Home Single Language
4.	GPU	Intel UHD Graphics

3.6 Metode Analisis

- 1 Analisis menggunakan metode Regresi Linier untuk mengetahui pemodelan dalam prediksi harga mobil bekas serta keakuratan model yang terbentuk dengan Bahasa pemrograman Python menggunakan tools Google Colaboratory.
- 2 Model prediksi yang telah dibuat dengan machine learning kemudian di visualkan menjadi sebuah visualisasi data dengan tools Microsoft Power BI yang untuk menghasilkan solusi akurat guna menghasilkan sebuah keputusan kepada pihak yang terlibat terkhusus pemilik bisnis.

3.7 Tahapan Penelitian

Pada bab tahapan ini memaparkan terkait tahap-tahap yang digunakan untuk prediksi harga mobil bekas telah terbentuk pada gambar 3.3.



Gambar 3.3 Tahapan Penelitian

3.7.1 Identifikasi Masalah dan Data

Penelitian ini berfokus pada masalah prediksi harga mobil bekas, yang seringkali dihadapi oleh pemilik bisnis penjualan mobil bekas. Mereka kesulitan menentukan harga jual yang tepat, sehingga dapat mengakibatkan overprice atau underprice, yang berdampak pada keuntungan yang diperoleh dan kesulitan dalam tentunya akan kesulitan dalam mengoptimalkan stok. Oleh karena itu, penelitian ini berusaha untuk mengatasi masalah tersebut dengan menggunakan analisis regresi linier. Data yang diperlukan untuk penelitian ini terdiri dari berbagai fitur mobil bekas seperti tahun produksi, kilometer yang sudah ditempuh mobil itu sendiri, jenis bahan bakar, dan lain sebagainya.

3.7.2 Pencarian Dataset

Dataset yang digunakan dalam penelitian ini adalah data "Used Car Price Prediction" yang diambil dari situs web Kaggle. Data ini diperoleh melalui platform Kaggle yang kemudian disimpan dalam format Comma Separated Value (CSV). Dataset ini didalamnya mencakup informasi mengenai data historis mobil bekas di India dari tahun ke tahun. Dataset didapatkan dari website Kaggle melalui link <https://www.kaggle.com/code/iabhishekmaurya/used-car-price-prediction>.

3.7.3 Pembersihan Data dan One Hot Encoding

Pembersihan data atau yang biasa disebut sebagai data cleansing merupakan tahap yang melibatkan penghapusan, perbaikan, atau memodifikasi data yang tidak akurat, tidak lengkap, tidak relevan, serta data yang terduplikasi dari suatu kumpulan data. Setelah mengunduh data dari situs website Kaggle, tahap pembersihan data ini dilakukan dengan menggunakan Google Colaboratory dan diimplementasikan dalam Bahasa pemrograman Python. Berikut adalah penjelasannya:

1 Proses Data Preprocessing

Proses Data Preprocessing adalah proses pengolahan data agar data yang tidak akurat dapat diperbaiki atau dibersihkan sehingga dapat digunakan. Selama *preprocessing*, data awal yang digunakan selalu mentah dan kemudian dilakukan proses untuk menyesuaikannya sesuai kebutuhan. Langkah ini mencakup semua data yang dikumpulkan, yang kemudian akan mengalami *preprocessing data* dalam bahasa pemrograman *python* menggunakan alat kolaboratif Google untuk mempersiapkan data untuk digunakan dalam pelatihan model. (Srinath, 2017) menjelaskan bahwa *python* memiliki beberapa fitur yang luar biasa. Pertama, Python menonjol karena perpustakaannya yang luas, yang menawarkan modul dalam distribusinya. Kedua, tata bahasa *Python* terkenal karena jelas dan mudah dipelajari. Selain itu, aturan presentasi kode sumbernya memudahkan proses verifikasi. *Python* juga mendukung pengoreksian dan penulisan ulang kode sumber, memberikan fleksibilitas dalam pengembangan. Selain itu, bahasa pemrograman ini berorientasi pada objek, yang mencerminkan pendekatan terstruktur untuk pengembangan perangkat lunak. Terakhir, *python* dapat dibangun menggunakan bahasa *Python* itu sendiri dan juga C/C++, menambah fleksibilitas dalam penggunaannya. Wes McKinney dalam (E. Hasibuan & Karim, 2022) menambahkan bahwa optimasi *python* juga dibantu dengan tersedianya paket inti seperti Pandas yang mendukung penyediaan struktur dan fungsi data yang kaya, dan dirancang khusus untuk mendukung analisis data yang kuat dan efisien.

2 Data Transformation

Langkah ini melibatkan perubahan data dari bentuk aslinya ke bentuk baru yang

lebih mudah untuk diproses dan diinterpretasikan. Tujuan dari transformasi data adalah untuk mengubah atau memodifikasi data ke dalam bentuk lain atau bentuk yang lain sehingga cocok untuk pemodelan *machine learning*.

3 Verifikasi Data

Tahap ini melibatkan verifikasi terhadap keakuratan data serta memastikan kembali bahwa data sudah bersih dan siap digunakan.

Dalam penelitian ini juga diterapkan metode One-Hot Encoding untuk memproses data, yang bertujuan untuk mengubah variabel kategorikal menjadi numerikal. Metode ini efektif untuk merepresentasikan variabel kategorikal sebagai vektor biner, di mana setiap kategori atau label diubah menjadi vektor biner dengan panjang yang sesuai dan jumlah kategori dalam variabel tersebut. Setiap elemen dalam vektor diisi dengan nilai nol, kecuali untuk indeks yang sesuai dengan kategori yang bersangkutan, yang diwakili oleh angka satu. Dengan teknik ini, One-Hot Encoding memungkinkan model regresi linier untuk memahami dan memproses informasi dari variabel kategorikal, sehingga variabel tersebut tetap dapat digunakan dalam pemodelan regresi linier dan dapat meningkatkan akurasi hasil prediksi harga mobil bekas. Penggunaan metode ini sangat penting dalam konteks penelitian ini, karena banyak fitur mobil bekas, seperti jenis bahan bakar atau lokasi, yang bersifat kategorikal dan perlu diolah agar dapat digunakan dalam analisis lebih lanjut. Dengan demikian, penerapan One-Hot Encoding tidak hanya mempermudah pemrosesan data, tetapi juga berkontribusi pada pengembangan model yang lebih robust dan akurat dalam memprediksi harga mobil bekas.

3.7.4 Penanganan Data Outlier

Outlier adalah nilai yang sangat berbeda dari nilai-nilai lain dalam dataset, Tahap ini melibatkan identifikasi dan penanganan outlier, baik dengan cara menghapus outlier yang sangat ekstrim atau menggunakan metode tertentu untuk menyesuaikan dampak outlier tersebut. Misalnya yang dilakukan dalam penelitian ini adalah dengan menggunakan metode IQR (*Interquartile Range*) untuk mengidentifikasi dan menangani data *outlier*.

3.7.5 Menentukan Variabel Independen dan Dependen

Variabel dependen dalam penelitian ini adalah harga jual mobil bekas, sedangkan variabel independen adalah fitur-fitur lain seperti yang terdapat dalam dataset yang digunakan yang meliputi tahun produksi, lokasi, kilometer mobil, dan lain sebagainya.

Pada tahap ini, juga dilakukan analisis korelasi untuk memahami seberapa kuat hubungan antara setiap variabel independent dengan variabel dependen dalam hal ini adalah harga bekas mobil. Hasil analisis ini disajikan dalam bentuk matriks korelasi dan heatmap untuk memperjelas hubungan antar variabel.

3.7.6 Verifikasi Data

Verifikasi data dilakukan untuk memastikan bahwa data yang telah diproses akurat dan siap digunakan dalam model analisis. Proses ini meliputi pengecekan distribusi data untuk mengidentifikasi anomali, penanganan nilai yang hilang melalui penghapusan atau imputasi untuk menghindari bias, serta identifikasi outlier yang dapat mempengaruhi hasil analisis. Selain itu, validasi proses encoding, seperti One-Hot Encoding, penting untuk memastikan variabel kategorikal dapat diinterpretasikan dengan benar. Dengan verifikasi data yang menyeluruh, peneliti dapat meningkatkan keandalan dan validitas model serta memastikan hasil analisis yang representatif.

3.7.7 Analisis Regresi Linier

Dalam penelitian ini, peneliti akan menganalisis data yang telah dikumpulkan dengan menggunakan teknik algoritma regresi linier berganda untuk memodelkan hubungan antara variabel dependen (harga mobil bekas) dan beberapa variabel independen seperti tahun produksi, lokasi, kilometer mobil, dan lain sebagainya. Proses ini merupakan penerapan algoritma yang bertujuan untuk menemukan, mengenali, dan menampilkan pola dalam data. Metode regresi linier dipilih karena diasumsikan adanya hubungan linier antara variabel dependen (harga mobil bekas) dan variabel independen (seperti tahun produksi, lokasi, dan kilometer). Jika data menunjukkan pola linear, regresi linier dapat memberikan hasil yang akurat dan relevan. Metode ini juga memungkinkan prediksi harga mobil bekas berdasarkan variabel yang ada, sehingga membantu dalam pengambilan keputusan yang lebih baik. Dalam sebuah model regresi, biasanya diperoleh nilai positif dan negatif pada estimasi yang seharusnya memiliki pengaruh signifikan berdasarkan beberapa faktor tertentu. Oleh karena itu, pada tahap ini diperlukan metode machine learning atau statistik agar model dapat diterapkan dengan efektif. Algoritma regresi linier ini akan menghitung koefisien regresi untuk setiap variabel independen, yang menunjukkan seberapa besar pengaruh masing-masing variabel independen terhadap harga mobil. Setelah menetapkan metode, langkah selanjutnya

yaitu dengan mengaplikasikannya pada dataset yang telah diperoleh. Jika nantinya data disesuaikan, langkah ini dapat kembali ke langkah *preprocessing* data.

3.7.8 Pengujian Model (Akurasi dan Nilai Error)

Pengujian model dilakukan dengan menghitung akurasi dan nilai error, yang merupakan tahap krusial dalam pengembangan sistem prediksi harga mobil bekas. Tujuan utama dari pengujian ini adalah untuk menilai seberapa efektif model yang telah dibangun dalam memprediksi harga mobil bekas. Salah satu metode yang umum digunakan untuk mengukur akurasi model adalah dengan membagi dataset menjadi dua bagian: data latih (training data) dan data uji (testing data). Pembagian ini bertujuan untuk melatih model pada satu subset data dan mengujinya pada subset lainnya, sehingga memberikan gambaran yang lebih akurat tentang kinerja model di dunia nyata. Pembagian data dapat dilakukan dengan berbagai macam rasio, yang umum digunakan adalah:

- a. 90/10: 90% data digunakan untuk pelatihan dan 10% untuk pengujian.
- b. 80/20: 80% data digunakan untuk pelatihan dan 20% untuk pengujian.
- c. 70/30: 70% data untuk pelatihan dan 30% untuk pengujian.
- d. 60/40: 60% data untuk pelatihan dan 40% untuk pengujian.

Rasio yang dipilih dapat mempengaruhi hasil evaluasi model. Rasio yang lebih besar untuk data latih biasanya menghasilkan model yang lebih baik, tetapi mengurangi jumlah data yang tersedia untuk pengujian. Setelah pengujian dilakukan dan akurasi ditentukan, langkah selanjutnya adalah menghitung nilai error berdasarkan hasil dari nilai prediksi. Penghitungan ini dilakukan dengan menggunakan beberapa metrik, yaitu:

- a. MSE

Mean squared error (MSE) adalah rata-rata kuadrat kesalahan antara nilai aktual dan nilai prediksi. Metode MSE umumnya digunakan untuk memverifikasi nilai estimasi kesalahan dari suatu model prediksi. Nilai MSE yang rendah atau mendekati nol mengindikasikan bahwa hasil prediksi sesuai dengan data aktual. MSE sering digunakan untuk mengevaluasi model regresi.

- b. RMSE

Root mean square error (RMSE) adalah ukuran tingkat kesalahan pada hasil prediksi. Semakin kecil nilai RMSE (mendekati nol), maka semakin akurat hasil prediksinya. RMSE merupakan metode evaluasi model

regresi linier dengan mengukur tingkat akurasi hasil estimasi model. RMSE dihitung dengan mengkuadratkan kesalahan produksi dengan jumlah data dan mengambil akar kuadratnya.

c. R-Squared (R²)

Metrik yang memberi tahu proporsi varian dalam variabel respons dari model regresi yang dapat dijelaskan oleh variabel prediktor. Nilai ini berkisar dari 0 hingga 1. Semakin besar nilai R², semakin baik model tersebut cocok dengan set data.

d. MAE

Mean absolute error (MAE) adalah rata-rata perbedaan antara nilai aktual dan nilai prediksi model. Hasil MAE yang kecil mengindikasikan bahwa model tersebut baik dalam melakukan prediksi, sedangkan hasil MAE yang besar mengindikasikan bahwa model tersebut memiliki masalah dalam melakukan prediksi. Jika MAE memberikan nilai 0, maka model merupakan prediktor yang sempurna untuk hasil (Faiqoh, 2023).

Untuk menghitung metrik kesalahan untuk model regresi linier seperti MAE, MSE, RMSE, dan R-squared dapat menggunakan kode *Python* dengan menggunakan library Scikit-Learn, seperti yang ditunjukkan pada Gambar 3.4.

```
from sklearn import metrics
from sklearn import metrics import mean_squared_error, mean_absolute_error

print("\t\tError Table")
print('Mean Absolute Error      : ', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error       : ', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Absolute Error  : ', np.sqrt(metrics.mean_squared_error(y_test,
y_pred)))
print('R Squared Error           : ', metrics.r2_score(y_test, y_pred))
```

Gambar 3.4 Kode Program untuk Menghitung Metrik Kesalahan

3.7.9 ETL

Sebelum disajikan, data harus terlebih dahulu dilakukan ekstrak, transform, dan load. ETL bertindak sebagai pengintegrasian antara data yang akan disimpan di repositori sehingga memiliki format visualisasi yang sama. ETL dapat meningkatkan kualitas analisis data hingga hasil business intelligence melalui pemrosesan data yang efisien. Dalam konteks penelitian ini, data yang sudah dibersihkan dan diproses dimuat ke dalam system visualisasi untuk diolah lebih lanjut.

Proses ETL ini dibagi menjadi tiga, yaitu:

- 1 *Extract*, adalah proses penentuan source yang akan digunakan sebagai sumber data bagi *data warehouse*. Dalam proses ini ditentukan data apa saja yang dibutuhkan, tabel apa yang dijadikan sumber.
- 2 *Transform*, setelah data ditentukan barulah data tersebut diubah agar sesuai dengan standar yang ada pada *data warehouse*.
- 3 Dan yang terakhir *Load*, yaitu proses untuk memasukkan data-data yang sudah di *transform* ke dalam *data warehouse* untuk disimpan sebagai *summary*.

3.7.10 Pemodelan Data Warehouse

Pemodelan data warehouse ini merupakan tahapan penting dalam mendesain struktur data yang menunjang analisis bisnis multidimensi. Langkah pertama yang dilakukan adalah dengan mengidentifikasi dimensi dan fakta yang relevan dengan bisnis yang diteliti. Kemudian, dimensi beserta fakta-fakta disusun menjadi sebuah skema, seperti *star schema* atau *snowflake schema* guna menghasilkan keterkaitan yang jelas di antara keduanya. Selanjutnya, pemodelan dimensi dilakukan untuk mengilustrasikan hirarki beserta hubungan antar dimensi, untuk memudahkan dalam melakukan analisis yang lebih mendalam. Terakhir, data diorganisasikan dalam bentuk tabel fakta dan dimensi yang siap digunakan untuk membangun data warehouse. Dengan demikian, pemodelan data warehouse berperan penting dalam mentransformasi data mentah ke dalam informasi yang bernilai dalam pengambilan keputusan bisnis yang lebih akurat dan efisien.

3.7.11 Perancangan Dashboard Visualisasi

Langkah selanjutnya yaitu visualisasi data. Visualisasi data pada dasarnya merupakan cara penyajian data dengan memanfaatkan berbagai elemen visual berupa grafik, bagan, dan peta untuk memberikan informasi. Dengan mengubah data dalam tabel menjadi grafik dan bagan, visualisasi data mempermudah pembaca dalam memahami informasi, wawasan, serta berbagai aspek yang tersembunyi di balik data mentah. Visualisasi data terletak pada kemampuannya untuk membantu pembaca melihat, berinteraksi, dan memahami data secara lebih baik. Sederhana atau kompleks, visualisasi yang tepat dapat membawa semua pembaca pada pemahaman yang serupa, terlepas dari tingkat keahlian mereka.

Visualisasi data bukan hanya membuat data menjadi lebih atraktif secara visual, namun juga membantu pengguna dalam mengidentifikasi tren, pola, hingga anomali yang tidak terlihat pada tabel data biasanya. Melalui visualisasi, pemilik bisnis dapat memahami situasi dengan cepat dan mengambil tindakan yang diperlukan.

Terdapat beragam jenis diagram yang dapat diterapkan dalam visualisasi data, seperti diagram garis, diagram batang, diagram lingkaran, treemap, dan lain sebagainya. Masing-masing jenis visualisasi memiliki keunggulan tersendiri dalam hal membantu pembaca memahami data. Sebagai contoh, diagram garis sangat tepat untuk menampilkan tren dari waktu ke waktu, sedangkan diagram batang cocok untuk membuat perbandingan jumlah di antara beberapa kategori. Diagram lingkaran berguna untuk menyajikan proporsi atau persentase, sedangkan treemap memberi gambaran yang lebih jelas mengenai hirarki.

Di samping itu, visualisasi data dapat dikustomisasi sesuai kebutuhan. Dengan kemampuan untuk memilih jenis diagram yang paling sesuai berdasarkan data yang ditampilkan, visualisasi menjadi alat yang sangat fleksibel sekaligus efektif. Teknologi modern juga mendukung visualisasi data yang bersifat interaktif, di mana pengguna dapat menggali data lebih dalam dengan mengklik ataupun mengarahkan kursor ke elemen-elemen tertentu. Pada dunia bisnis, visualisasi data merupakan hal yang esensial untuk presentasi laporan ke stakeholder, analisis pasar, serta pengembangan strategi. Visualisasi yang baik dapat menyampaikan data yang rumit melalui cara yang sederhana namun mudah dimengerti, sehingga para pengambil keputusan dapat memperoleh informasi secara cepat dan efisien.

3.7.12 Pengujian

Dalam tahapan pengujian dilakukan dua pengujian yang diantaranya adalah pengujian akurasi data, dan pengujian dashboard dengan berdasarkan tingkat kepuasan pengguna yang dilakukan dengan metode *System Usability Scale* (SUS).

Pengujian akurasi data adalah langkah penting dalam penelitian ini, dengan tujuan utama untuk memastikan bahwa informasi yang ditampilkan akurat dan konsisten dengan sumber data yang digunakan. Dengan informasi yang tepat, risiko kesalahan dalam penentuan harga dapat diminimalkan. Dashboard dirancang dengan fokus pada interaktivitas dan kemudahan penggunaan, sehingga pemilik bisnis dapat dengan mudah mengakses dan memahami informasi yang disajikan. Berbagai elemen visual, seperti grafik batang dan diagram garis, ditambahkan untuk mempermudah analisis data dan memberikan gambaran yang jelas tentang tren harga serta pola pasar. Selain itu, terdapat visualisasi perbandingan antara harga aktual dan

prediksi, yang memberikan wawasan lebih mendalam mengenai perbedaan antara keduanya. Fitur ini memungkinkan pemilik bisnis untuk dengan mudah membandingkan harga aktual dan prediksi, sehingga mereka dapat mengidentifikasi perbedaan dan membuat keputusan yang lebih tepat.

Selain menguji akurasi data, dilakukan juga pengujian terhadap dashboard berdasarkan tingkat kepuasan pengguna yang diukur menggunakan metode *System Usability Scale* (SUS). Metode ini bertujuan untuk menilai apakah desain dashboard informasi yang dikembangkan memenuhi permasalahan mengenai kurangnya visualisasi yang efektif dalam mempresentasikan prediksi harga mobil bekas dan dapat digunakan dengan efektif. Untuk itu, perlu dilakukan pengujian validasi desain. Metode yang digunakan adalah *System Usability Scale* (SUS). Dalam pengujian ini, kuesioner yang berisi pernyataan akan digunakan untuk mengukur kebergunaan desain dashboard informasi yang telah dibuat. Kuesioner ini akan diisi oleh para pemilik bisnis penjualan mobil bekas dan untuk daftar kuesioner dapat dilihat pada Tabel 3.5.

Tabel 3.5 Daftar Kuesioner (Lewis & Sauro, 2018)

No	Pernyataan
1	Saya merasa bahwa saya akan menggunakan <i>dashboard</i> ini secara rutin.
2	Saya merasa bahwa <i>dashboard</i> ini terlalu rumit untuk digunakan.
3	Saya merasa bahwa <i>dashboard</i> ini mudah untuk digunakan.
4	Saya merasa bahwa saya perlu belajar banyak sebelum bisa menggunakan <i>dashboard</i> ini.
5	Saya merasa bahwa <i>dashboard</i> ini sangat intuitif.
6	Saya merasa bahwa saya menemukan banyak ketidakkonsistenan dalam visualisasi <i>dashboard</i> ini.
7	Saya merasa bahwa hampir semua orang akan belajar menggunakan <i>dashboard</i> ini dengan cepat.
8	Saya merasa bahwa <i>dashboard</i> ini sangat membingungkan untuk digunakan dan dipahami.
9	Saya merasa bahwa kombinasi warna dan tata letak konten di <i>dashboard</i> ini efektif dan nyaman untuk dilihat.
10	Saya merasa bahwa saya perlu menggunakan <i>dashboard</i> ini lebih sering untuk memahami semua fiturnya.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Implementasi Regresi Linier

Dalam sub bab ini akan menjelaskan proses mengenai implementasi regresi linier dari mulai pengumpulan data hingga pengujian pemodelan. Tujuan utama melakukan implementasi pemodelan menggunakan regresi linier ini adalah untuk mencari atau menentukan prediksi harga mobil bekas yang berguna untuk pemilik bisnis yang seringkali mengalami kesulitan dalam melakukan penentuan harga pada stok mobil yang dimilikinya. Sehingga, pemilik bisnis akhirnya memperoleh keuntungan yang kurang maksimal dan bahkan bisa mengalami kerugian akibat penentuan harga yang kurang efektif, serta stok mobil bekas yang dimilikinya pun akan mengalami penumpukan dan terbengkalai.

Pertama kali untuk melakukan pemodelan ini adalah dengan melakukan pengumpulan data yang dilakukan dengan cara mencari, mengobservasi, menemukan data yang nantinya akan digunakan menggunakan browser google chrome. Kemudian data selanjutnya disimpan dalam file Microsoft Excel dengan format Comma Separated Value (CSV). Setelah dataset didapatkan, penting untuk memahami hubungan antara variabel-variabel yang ada. Analisis korelasi dilakukan untuk menentukan seberapa kuat hubungan antara variabel independen dan harga mobil bekas. Dengan demikian, analisis korelasi memberikan pemahaman awal tentang interaksi antar variabel.

Tahun	0.299475
Kilometer Mobil	-0.008249
Tipe Pemilik	-0.091098
Konsumsi BBM (kmpl)	-0.339932
Kapasitas Mesin (CC)	0.658047
Tenaga (bhp)	0.772756
Jumlah Kursi	0.055547
Harga Bekas (IDR)	1.000000
Name: Harga Bekas (IDR), dtype: float64	

Gambar 4.1 Hasil Analisis Korelasi

Seperti yang ditampilkan pada Gambar 4.1, didapatkan koefisien korelasi variabel Tahun adalah 0.29, hal ini menunjukkan adanya korelasi positif yang lemah, yang berarti hubungan antara variabel tersebut tidak terlalu kuat. Sementara itu, koefisien -0.008249 yang dimiliki variabel Kilometer Mobil menunjukkan korelasi negatif yang sangat lemah, hampir tidak ada hubungan yang sama sekali. Koefisien -0.091098 yang dimiliki variabel Tipe Pemilik juga menunjukkan korelasi negatif yang lemah, yang menandakan bahwa hubungan tersebut tidak signifikan. Di sisi lain, koefisien -0.3399 yang dimiliki variabel Konsumsi BBM (kmpl) menunjukkan adanya korelasi negatif yang moderat, yang berarti terdapat hubungan negatif yang cukup jelas. Koefisien 0,65 yang dimiliki variabel Kapasitas Mesin (CC) menunjukkan korelasi positif yang moderat, menandakan adanya hubungan positif yang cukup signifikan. Selanjutnya koefisien 0,7727 yang dimiliki variabel Tenaga (bhp) menunjukkan korelasi positif yang kuat, yang berarti hubungan antara variabel tersebut sangat kuat. Koefisien 0.0555 yang dimiliki variabel Jumlah Kursi menunjukkan korelasi positif yang sangat lemah, hampir tidak ada hubungan. Terakhir, koefisien 1,000 yang dimiliki variabel Harga Bekas (IDR) menunjukkan korelasi positif yang sempurna, yang berarti hubungan antara kedua variabel tersebut sangat kuat dan langsung.

Setelah melakukan analisis korelasi untuk memahami hubungan variabel yang ada, selanjutnya data mentah yang telah disimpan perlu dipersiapkan dengan melakukan perubahan agar lebih mudah dipahami. Proses ini melibatkan eliminasi data yang tidak sesuai dan tidak dibutuhkan. Dalam penelitian ini, data yang diperlukan mencakup Nama Mobil, Lokasi, Tahun, Kilometer Mobil, Jenis Bahan Bakar, Transmisi, Tipe Pemilik, Konsumsi BBM, Kapasitas Mesin, Tenaga, Jumlah Kursi, Harga Baru, dan Harga Bekas. Oleh karena itu, tahapan pre-processing data harus dilakukan, termasuk menghapus baris atau kolom yang memiliki nilai kosong atau hilang.

1 Menghapus nilai *missing* (NaN)

Pada dataset ini terdapat data yang hilang dan jumlahnya tidak terlalu besar sehingga data tersebut akan dihapus. Dari data asli terdapat 6.019 baris dan 14 kolom. Setelah dilakukan penghapusan data dengan nilai NaN, maka menjadi 5.872 baris dan 13 kolom. Tabel 3.2 menunjukkan jumlah data yang hilang pada data ini:

Tabel 4.1 *Feature Missing*

Nama Mobil	0
Lokasi	0
Tahun	0
Kilometer Mobil	0
Jenis Bahan Bakar	0
Transmisi	0
Tipe Pemilik	0
Konsumsi BBM	2
Kapasitas Mesin	36
Tenaga	36
Jumlah Kursi	42
Harga Baru	5195
Harga Bekas	0
Dtype : int64	

Transformasi data juga dilakukan dalam beberapa tahapan dalam penelitian ini. Langkah ini melibatkan perubahan data dari bentuk aslinya ke bentuk baru yang lebih mudah untuk diproses dan diinterpretasikan. Tujuan dari transformasi data adalah untuk mengubah atau memodifikasi data ke dalam bentuk lain atau bentuk yang lain sehingga cocok untuk pemodelan *machine learning*. Beberapa langkah tersebut dijabarkan seperti berikut ini:

- 1 Mengganti variabel atribut Tipe Pemilik menjadi numerik, penggantian disini dilakukan pada variabel Tipe Pemilik, dimana dataset ini pada awalnya memiliki tipe data object (kategorikal), kemudian diubah menjadi bentuk numerik (integer) atau angka. Alternatif ini bertujuan untuk menciptakan data set yang berkualitas untuk melakukan model prediksi, selain pemodelan yang hanya dapat dilakukan dengan tipe data numerik.

```
# Penyesuaian Tipe Data (Tipe Pemilik)
pemetaan = {"First": 1, "Second": 2, "Third": 3, "Fourth & Above": 4}
# Menggunakan pemetaan manual untuk mengubah kolom "Tipe Pemilik" menjadi data
numerik
mobkas_df["Tipe Pemilik"] = mobkas_df["Tipe Pemilik"].map(pemetaan)
```

Gambar 4.2 Kode Program untuk Penyesuaian Tipe Data

- 2 Mengubah satuan variabel Konsumsi BBM, pada variabel ini satuan dari variabel akan disesuaikan sepenuhnya atau diubah menjadi "kmpl". Tujuannya adalah untuk menghasilkan data yang lebih berkualitas serta hasil pemodelan yang tentu saja akan lebih valid.

```
# Mengubah satuan variabel Konsumsi BBM data = {'Konsumsi BBM'}
df = pd.DataFrame(data)
def convert_to_kmpl(entry):
    if 'km/kg' in entry:
        value = float(entry.replace(' km/kg', ''))
        return str(value * 0.7457) + " kmpl" # Konversi dari km/kg ke kmpl
    return entry
mobkas_df['Konsumsi BBM'] = mobkas_df['Konsumsi BBM'].apply(convert_to_kmpl)
```

Gambar 4.3 Kode Program untuk Mengubah Satuan

- 3 Pada variabel Kapasitas Mesin, Tenaga, serta Konsumsi BBM, dilakukan penghapusan nilai satuan. Apabila tidak dilakukan penghapusan maka ketiga variabel tersebut tidak dapat digunakan karena memiliki tipe data object (kategorikal). Cara untuk mengatasi masalah ini adalah dengan melakukan perubahan dan penyesuaian kembali pada kedua variabel tersebut.

```
# Penghapusan nilai satuan variabel
mobkas_df['Kapasitas Mesin'] = mobkas_df['KapasitasMesin'].str.replace('
CC', '')
mobkas_df['Tenaga'] = mobkas_df['Tenaga'].str.replace(' bhp', '')
mobkas_df['Konsumsi BBM'] = mobkas_df['Konsumsi
BBM'].str.replace('kmpl', '').astype(float)
```

Gambar 4.4 Kode Program untuk Menghapus Satuan Variabel

- 4 Melakukan penggantian nama variabel atribut, dimaksudkan agar proses pemodelan lebih mudah dilakukan karena isi datanya sudah tidak lagi mengandung satuan. Maka untuk memudahkannya diberikan opsi penggantian nama pada beberapa variabel.

```
# Penggantian nama beberapa variabel
mobkas_df.rename(columns={'Konsumsi BBM': 'Konsumsi BBM (kmpl)'}, inplace=True)
mobkas_df.rename(columns={'Kapasitas Mesin': 'Kapasitas Mesin (CC)'},
inplace=True)
mobkas_df.rename(columns={'Tenaga': 'Tenaga (bhp)'}, inplace=True)
mobkas_df.rename(columns={'Harga Bekas': 'Harga Bekas (Lakh)'}, inplace=True)
```

Gambar 4.5 Kode Program untuk Penggantian Nama Beberapa Variabel

- 5 Membuat fitur baru dengan cara merekayasa fitur, rekayasa fitur ini adalah proses pembuatan fitur baru dari data mentah yang ada pada atribut Nama Mobil. Fitur baru yang dihasilkan adalah dengan nama Merk. Fitur baru ini dapat digunakan untuk menambah performa model machine learning atau membuat model lebih mudah dipahami. Membuat fitur baru untuk rekayasa fitur dapat dianggap sebagai bagian dari proses transformasi data karena melibatkan pengubahan data mentah ke dalam bentuk yang lebih sesuai untuk analisis atau pemodelan.

Tabel 4.2 Sampel Fitur Baru yang Dihasilkan

Nama Mobil	Merk
Maruti Wagon R LXI CNG	Maruti
Hyundai Creta 1.6 CRDi SX Option	Hyundai
Honda Jazz V	Honda
Maruti Ertiga VDI	Maruti
Audi A4 New 2.0 TDI Multitronic	Audi

- 6 Penyesuaian tipe data keseluruhan, penyesuaian ini dilakukan terhadap keseluruhan variabel dengan tipe data yang masih bersifat kategorikal (object). Tujuannya sendiri adalah agar dapat memaksimalkan data itu sendiri untuk dilakukan pemodelan prediksi dan dapat menghasilkan data yang berkualitas. Setelah penyesuaian tipe data secara keseluruhan dilakukan maka terdapat total 8 atribut dengan tipe data integer (numerik) dan sisanya 5 atribut tipe data berjenis object (kategorikal).

Selain melakukan preprocessing data, tahap selanjutnya adalah penerapan metode yang bernama One Hot Encoding dengan tujuan untuk mengubah variabel dengan tipe data kategorikal menjadi numerikal. Hal ini penting agar variabel tersebut tetap dapat digunakan dalam pemodelan regresi linier. Variabel kategorikal seperti Nama Mobil, Lokasi, Jenis Bahan Bakar, Transmisi, dan Merk akan diubah menjadi dalam bentuk angka (dalam One Hot Encoding disebut vektor biner), meskipun tidak ada perubahan pada variabel Nama Mobil dan Merk. Proses pelabelan ini dilakukan dengan menggunakan skala nominal, karena dalam

pemodelan machine learning, mesin hanya dapat membaca data dalam bentuk numerikal. Total dalam dataset yang digunakan pada penelitian ini mencakup 12 variabel sebagai variabel independen.

```
# Penerapan One-Hot Encoding
X = a[['Lokasi', 'Jenis Bahan Bakar', 'Transmisi']]
Y = a.Harga Bekas (IDR)
X = pd.get dummies(drop_first=False) (data=X)
```

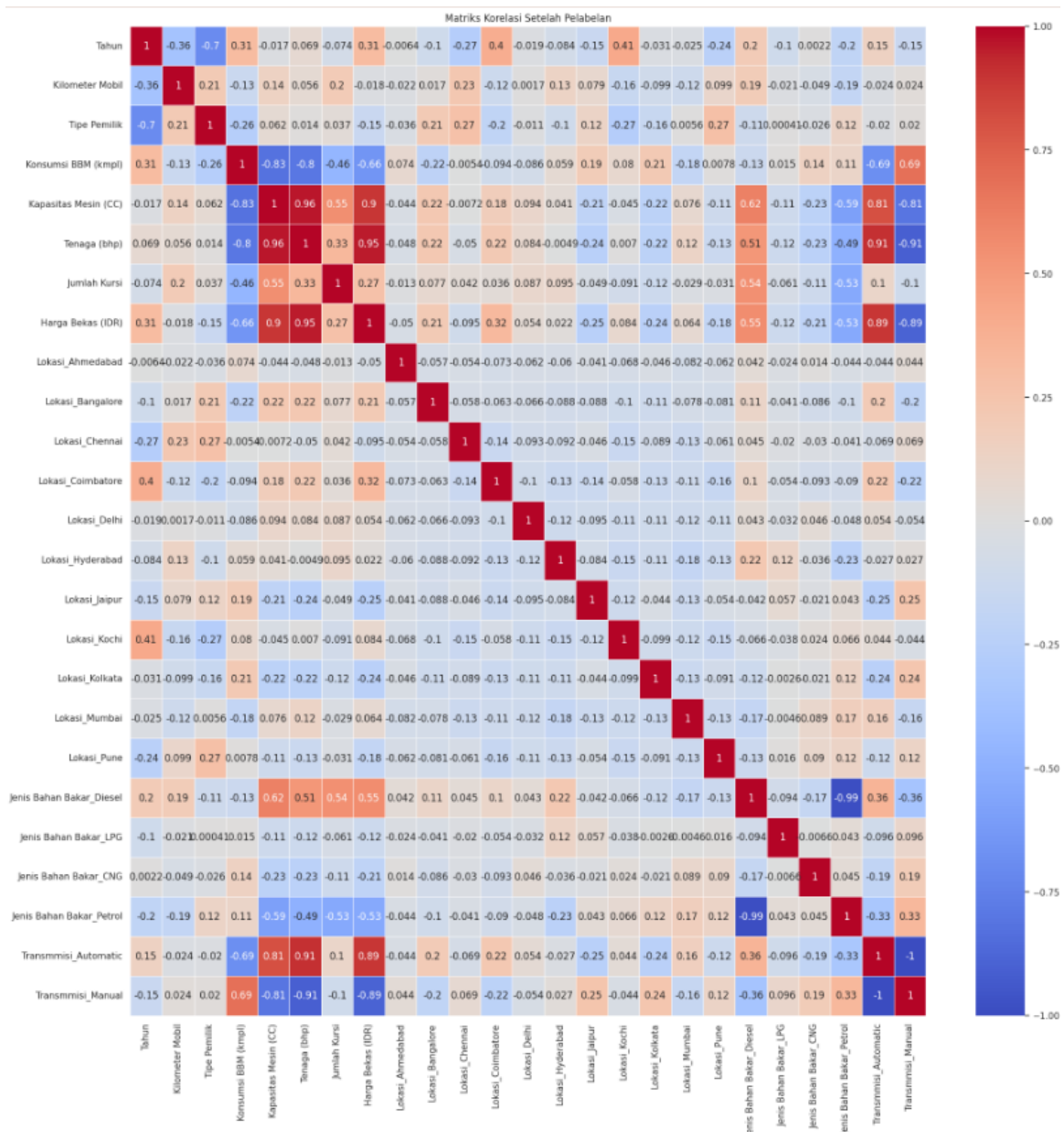
Gambar 4.6 Kode Program Penerapan *One-Hot Encoding*

...	Lokasi_Hyderabad	Lokasi_Jaipur	Lokasi_Kochi	Lokasi_Kolkata	Lokasi_Mumbai	Lokasi_Pune	Jenis Bahan Bakar_Diesel	Jenis Bahan Bakar_LPG	Jenis Bahan Bakar_Petrol	Transmisi_Manual
...	0	0	0	0	1	0	0	0	0	1
...	0	0	0	0	0	0	1	1	0	1
...	0	0	0	0	0	0	0	0	1	1
...	0	0	0	0	0	0	0	1	0	1
...	0	0	0	0	0	0	0	1	0	0

Gambar 4.7 *Label Encoding*

Hasil dari proses pelabelan data menggunakan *One Hot Encoding* dapat dilihat pada Gambar 4.6, yang menunjukkan adanya variabel baru yang dibentuk. Proses ini menciptakan kolom baru dari variabel kategorik yang ada, di mana setiap kolom baru berisi nilai 0 atau 1, dengan arti 0 berarti tidak ada dan 1 berarti ada. Berdasarkan gambar tersebut, dapat diketahui bahwa variabel lokasi terdiri atas 11 kategori, jenis bahan bakar terdiri atas 4 kategori, serta transmisi terdiri atas 2 kategori. Setelah dilakukan pelabelan data, terdapat 14 variabel independen baru yang terbentuk, yaitu Lokasi_Ahmedabad, Lokasi_Bangalore, Lokasi_Chennai, Lokasi_Coimbatore, Lokasi_Delhi, Lokasi_Hyderabad, Lokasi_Jaipur, Lokasi_Kochi, Lokasi_Kolkata, Lokasi_Mumbai, Lokasi_Pune, Jenis Bahan Bakar_Diesel, Jenis Bahan Bakar_LPG, Jenis Bahan Bakar_Petrol, Jenis Bahan Bakar_CNG, Transmisi_Automatic dan Transmisi_Manual.

Setelah melakukan pelabelan data dengan metode One Hot Encoding, perlu untuk dilakukan pengecekan korelasi antar variabel sebelum dilakukan tahapan mempertimbangkan multikolinearitas. Pengecekan ini bertujuan untuk melihat hubungan awal pada variabel yang sudah dilakukan pelabelan data kategorikal menjadi numerikal yang dihasilkan pada Gambar 4.8.

Gambar 4.8 Pengecekan Korelasi dengan *Heatmap*

Dalam menerapkan One Hot Encoding, peneliti mempertimbangkan aspek multikolinearitas, yang penting untuk dilakukan karena pelabelan data kategorikal menjadi numerikal dapat menyebabkan korelasi tinggi antar variabel independen. Untuk menghindari masalah ini, peneliti menggunakan parameter “*drop_first=True*” yang bertujuan untuk menghapus satu kolom kategori dalam proses One Hot Encoding. Setelah proses encoding, peneliti memeriksa korelasi antar variabel. Dengan demikian, penulis memastikan bahwa pelabelan data tidak hanya menghasilkan variabel-variabel baru yang relevan, tetapi juga menjaga keakuratan analisis dengan mempertimbangkan multikolinearitas.

Selain melakukan pelabelan data dengan metode *One Hot Encoding*, mengidentifikasi data *outlier* juga perlu dilakukan. Data *outlier* dapat diartikan sebagai data yang menyimpang terlalu jauh dari rentang normal dalam kumpulan data atau biasa disebut dengan nilai-nilai ekstrem. Tujuannya adalah untuk membantu menganalisa data secara lebih jelas serta meningkatkan persentase keakuratan data. Untuk mendeteksi *outlier*, teknik berbasis IQR yang diusulkan pada awalnya mengidentifikasi kuartil dan kemudian menentukan rentang data yang diinginkan. Artinya nilai batas bawah dan batas atas sudah teridentifikasi. Titik data yang berada di bawah batas bawah dan di atas batas atas dianggap sebagai outlier (Magar, Ruikar, Bhoite, & Mente, 2024). Sebelum melakukan teknik IQR ini penting untuk melakukan penghapusan data yang ada dalam *dataframe* yang bersifat duplikat, Tujuannya agar saat menghapus data yang bersifat *outlier* tadi menjadi maksimal.

Berikut Gambar 4.9 merupakan kode program untuk mendeteksi *outlier* dengan Teknik IQR, dan Gambar 4.10 merupakan hasil dari keluaran programnya

```
numeric_cols = final_mobkas_cleaned.select_dtypes(include='number')

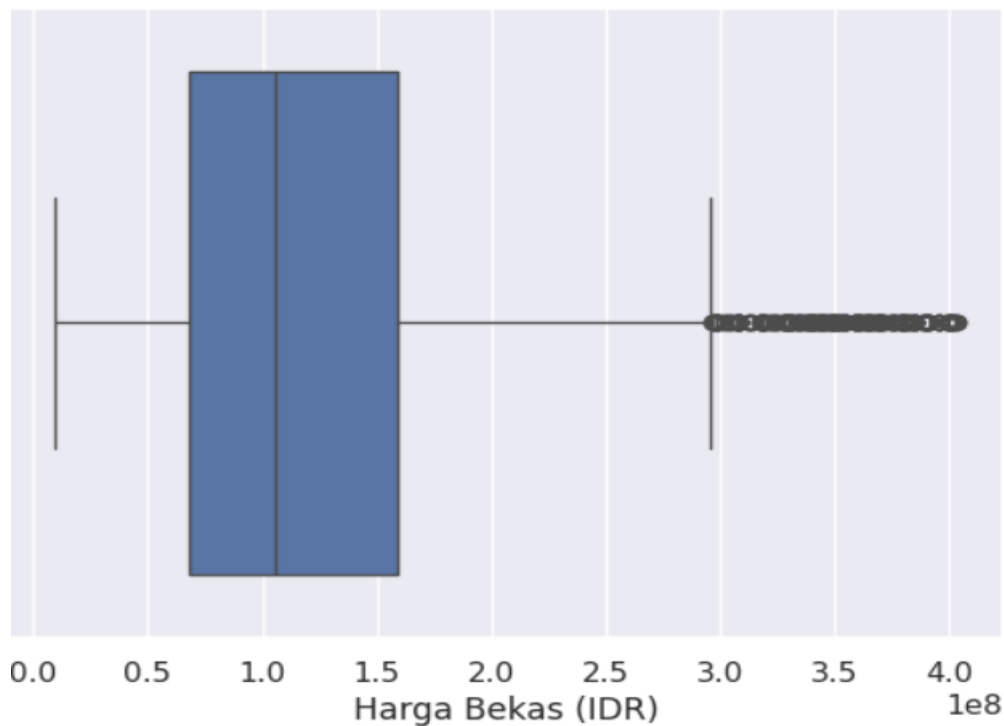
# Menghitung Quartile dan Interquartile Range (IQR)
Q1 = numeric_cols['Harga Bekas (IDR)'].quantile(0.25)
Q3 = numeric_cols['Harga Bekas (IDR)'].quantile(0.75)
IQR = Q3 - Q1

# Menentukan batas bawah dan atas untuk outliers
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

# menghitung outlier
no_outliers = final_mobkas_cleaned[~((numeric_cols < lower_bound) | (numeric_cols >
upper_bound)).any(axis=1)]

# Menampilkan kolom dengan outliers
print("\nData Setelah menghapus outlier:")
print(no_outliers)
```

Gambar 4.9 Kode Program Identifikasi Data Outliers



Gambar 4.10 Hasil Identifikasi Data Outliers

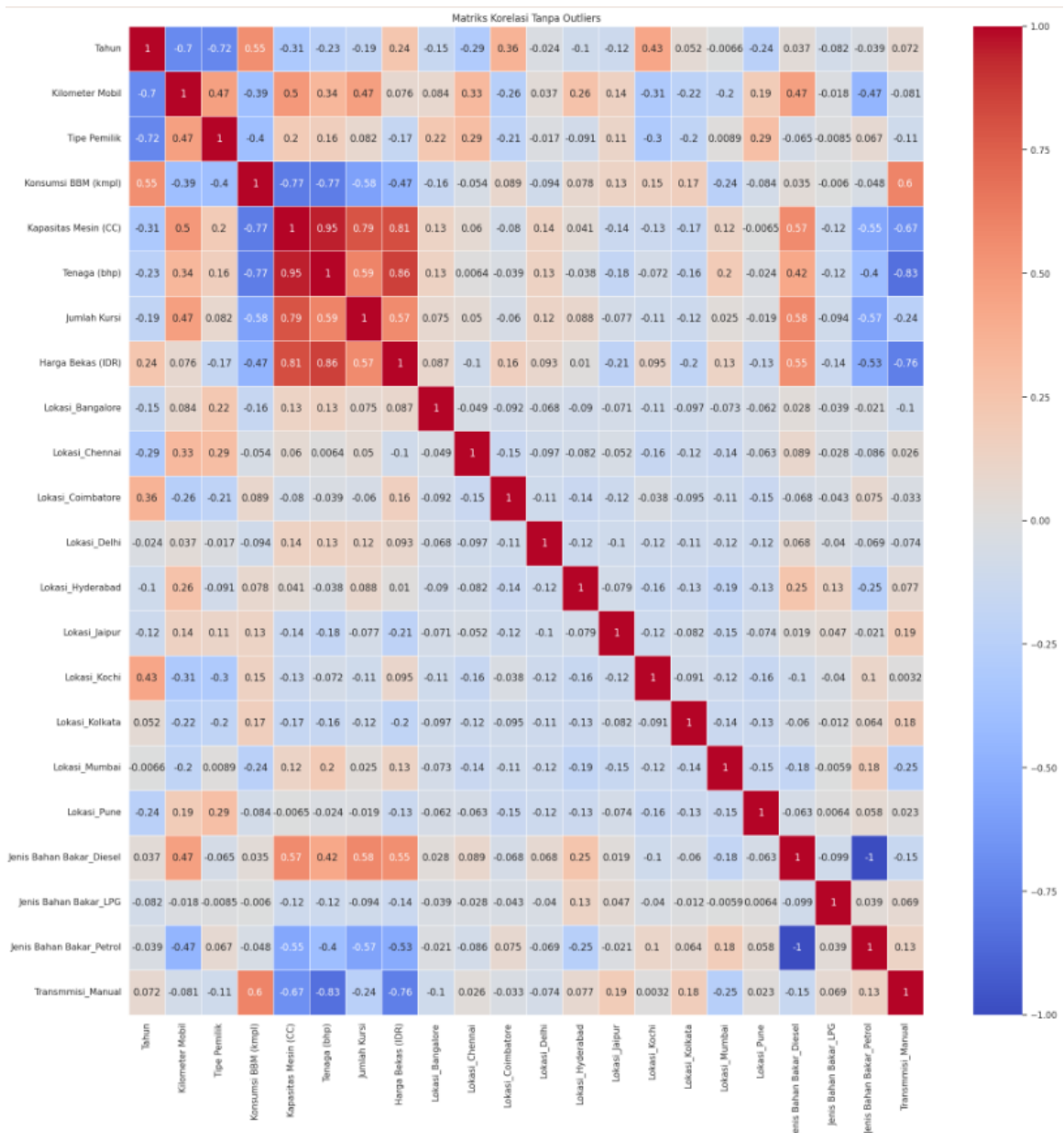
Pada Gambar 4.10 keluaran yang dihasilkan sudah menghasilkan data mana saja yang teridentifikasi sebagai *outliers*. Nilai *outliers* ini biasanya adalah nilai yang terletak lebih dari 1,5 kali rentang interkuartil (IQR) dari kuartil, dan visualisasi ini menunjukkan beberapa titik yang terletak di sisi kanan untuk variabel “Harga Bekas (IDR)” merupakan nilai yang bersifat *outliers*.

Langkah berikutnya dalam proses analisis data adalah melakukan analisis matriks korelasi yang telah divisualisasikan dalam bentuk *heatmap*. Matriks korelasi ini berfungsi untuk menilai sejauh mana variabel-variabel lain berhubungan dengan variabel dependen. Dengan menggunakan *heatmap*, hubungan antar variabel dapat dilihat dengan jelas, di mana warna dan intensitas menunjukkan kekuatan serta arah dari korelasi tersebut. Korelasi yang tinggi antara variabel dapat menunjukkan adanya keterkaitan yang kuat, meskipun penting untuk diingat bahwa korelasi tidak selalu berarti adanya hubungan sebab akibat.

Analisis ini memberikan wawasan yang berharga dalam memahami struktur data dan interaksi antar variabel. Misalnya, jika variabel independen menunjukkan korelasi positif yang signifikan dengan variabel dependen, ini dapat mengindikasikan bahwa peningkatan pada variabel independen tersebut berpotensi meningkatkan nilai variabel dependen.

Sebaliknya, korelasi negatif dapat menunjukkan bahwa peningkatan pada satu variabel dapat berhubungan dengan penurunan pada variabel lainnya.

Temuan dari analisis korelasi ini menjadi landasan yang solid untuk pengambilan keputusan dalam analisis regresi. Dengan mengetahui variabel mana yang memiliki hubungan signifikan dengan variabel dependen, peneliti dapat memilih variabel yang tepat untuk dimasukkan ke dalam model regresi. Hal ini memastikan bahwa model yang dibangun tidak hanya relevan tetapi juga dapat memberikan prediksi yang akurat dan dapat diandalkan. Dengan demikian, analisis matriks korelasi dan visualisasi heatmap menjadi langkah krusial dalam proses pemodelan. Mereka tidak hanya membantu dalam mengidentifikasi variabel kunci yang akan mempengaruhi hasil akhir dari analisis regresi, tetapi juga memberikan pemahaman yang lebih dalam tentang dinamika data. Proses ini memastikan bahwa model yang dibangun tidak hanya akurat dalam prediksi, tetapi juga memberikan wawasan yang dapat digunakan untuk pengambilan keputusan yang lebih baik dalam konteks penelitian atau aplikasi praktis.



Gambar 4.11 Analisis Matriks Korelasi

Setelah semua tahapan dilakukan, tahap terakhir sebelum melakukan analisis Regresi Linier adalah data verification. Tahap ini melibatkan verifikasi terhadap keakuratan data serta memastikan kembali bahwa data sudah bersih dan siap digunakan. Gambar 4.12 menunjukkan bahwa data sudah tidak ada lagi nilai yang terduplikasi, hilang bahkan

missing, selain itu juga data sudah menjadi kedalam bentuk numerik sehingga sudah dapat dilakukan untuk pemodelan dengan *machine learning*.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5872 entries, 0 to 5871
Data columns (total 24 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Nama Mobil                            5872 non-null   object
1   Tahun                                 5872 non-null   int64
2   Kilometer Mobil                       5872 non-null   int64
3   Tipe Pemilik                          5872 non-null   int64
4   Konsumsi BBM (kmpl)                   5872 non-null   int64
5   Kapasitas Mesin (CC)                  5872 non-null   int64
6   Tenaga (bhp)                          5872 non-null   int64
7   Jumlah Kursi                          5872 non-null   int64
8   Harga Bekas (IDR)                     5872 non-null   int64
9   Merk                                  5872 non-null   object
10  Lokasi_Bangalore                       5872 non-null   int64
11  Lokasi_Chennai                         5872 non-null   int64
12  Lokasi_Coimbatore                      5872 non-null   int64
13  Lokasi_Delhi                           5872 non-null   int64
14  Lokasi_Hyderabad                       5872 non-null   int64
15  Lokasi_Jaipur                          5872 non-null   int64
16  Lokasi_Kochi                           5872 non-null   int64
17  Lokasi_Kolkata                         5872 non-null   int64
18  Lokasi_Mumbai                          5872 non-null   int64
19  Lokasi_Pune                            5872 non-null   int64
20  Jenis Bahan Bakar_Diesel               5872 non-null   int64
21  Jenis Bahan Bakar_LPG                  5872 non-null   int64
22  Jenis Bahan Bakar_Petrol               5872 non-null   int64
23  Transmisi_Manual                       5872 non-null   int64
dtypes: int64(22), object(2)
memory usage: 1.1+ MB
```

Gambar 4.12 *Data Verification*

Data mobil bekas siap untuk diolah dan dianalisa secara lebih detail untuk mendapatkan wawasan dan *insight* yang dapat digunakan untuk memastikan kualitas data itu sendiri, mencegah terjadinya *overfitting*, memastikan konsistensi, menangani nilai yang hilang, dan masih banyak lagi kegunaan lainnya setelah tahap pembersihan data selesai.

Setelah data dipastikan siap, langkah selanjutnya adalah menerapkan algoritma regresi linier. Regresi linier yang digunakan dalam penelitian ini adalah regresi linier berganda (*multiple linear regression*). Pemilihan teknik Regresi Linier Berganda karena melibatkan dua atau lebih variabel. Dengan menerapkan regresi linier berganda, peneliti dapat menghasilkan model yang dapat digunakan untuk menentukan prediksi harga mobil bekas dari fungsi regresi linier berganda yang dihasilkan.

```
# kemudian dapat dihasilkan menjadi sebuah fungsi model regresi linier berganda
# fungsi regresi linier berganda = (-4.30e+10) + 2.14e+07*tahun + 3.89e+01*kilometer mobil + (-6.52e+06)*tipe pemilik
# + (-5.47e+06)*konsumsi_bbm (kmpl) + (-2.17e+03)*kapasitas mesin (cc) + 2.73e+06*tenaga (bhp)
# + (-1.95e+07)*jumlah_kursi + 2.52e+07*lokasi_bangalore + 4.98e+06*lokasi_chennai
# + 3.21e+07*lokasi_coimbatore + (-1.36e+07)*lokasi_delhi + 2.44e+07*lokasi_hyderabad
# + 8.06e+06*lokasi_jaipur + (-1.78e+07)*lokasi_kochi + (-3.15e+07)*lokasi_kolkata
# + (-3.56e+07)*lokasi_mumbai + (-7.44e+06)*lokasi_pune + 2.29e+07*jenis_bahan_bakar_diesel
# + 6.40e+06*jenis_bahan_bakar_lpg + (-2.85e+07)*jenis_bahan_bakar_petrol
# + (-4.87e+07)*transmisi_manual
```

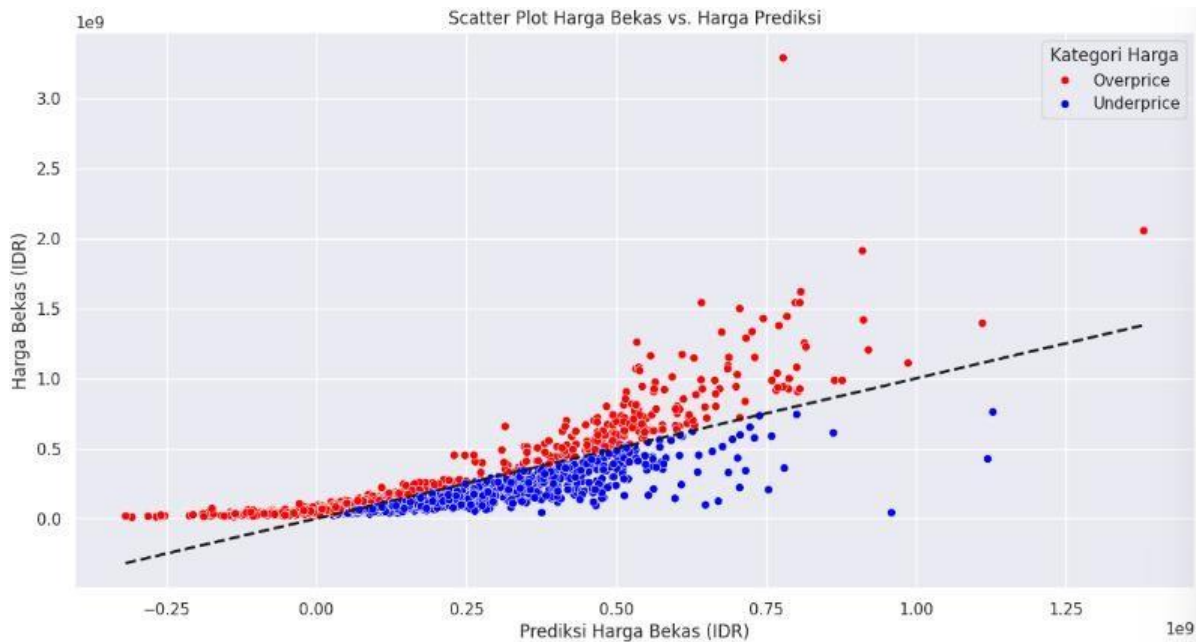
Gambar 4.13 Fungsi Regresi Linier Berganda

Fungsi regresi linier berganda yang dihasilkan dalam penelitian ini ditampilkan pada Gambar 4.13, dengan variabel yang digunakan meliputi Tahun, Kilometer Mobil, Tipe Pemilik, Konsumsi BBM (kmpl), Kapasitas Mesin (CC), Tenaga (bhp), Jumlah Kursi, Lokasi_Bangalore, Lokasi_Chennai, Lokasi_Coimbatore, Lokasi_Delhi, Lokasi_Hyderabad, Lokasi_Jaipur, Lokasi_Kochi, Lokasi_Kolkata, Lokasi_Mumbai, Lokasi_Pune, Jenis Bahan Bakar_Diesel, Jenis Bahan Bakar_LPG, Jenis Bahan Bakar_Petrol, Transmisi_Manual sebagai variabel yang menentukan.

Tabel 4.3 Nilai Prediksi dan Nilai Aktual

No	Nama Mobil	Harga Bekas (IDR)	Prediksi Harga Bekas (IDR)
1	Maruti Wagon R LXI CNG	35936250	-70245580
2	Hyundai Creta 1.6 CRDi SX Option	256687500	271428769
3	Honda Jazz V	92407500	49487795
4	Maruti Ertiga VDI	123210000	73347596
5	Audi A4 New 2.0 TDI Multitronic	364290899	370770219
...	...	...	...
5871	Chevrolet Beat Diesel	51337500	-3707426

Tabel 4.3 adalah nilai prediksi harga bekas berdasarkan beberapa variabel *independent* yang mempengaruhi. Prediksi harga mobil bekas yang dihasilkan oleh model tidak berbeda jauh dari nilai aslinya. Seperti contoh data ke 5, prediksi harga mobil yang dihasilkan adalah 370770219 IDR, sedangkan harga asli mobil tersebut adalah 364290899 IDR, dan begitu juga yang lainnya.



Gambar 4.14 Grafik Harga Prediksi dan Harga Aktual

Grafik scatter plot ini memberikan wawasan penting bagi pemilik bisnis tentang perbandingan antara harga bekas mobil yang sebenarnya (Harga Bekas IDR) dan harga yang diprediksi oleh model regresi linier (Prediksi Harga Bekas IDR). Garis diagonal putus-putus berfungsi sebagai referensi, menunjukkan harga yang diprediksi oleh model sama dengan harga sebenarnya, yang kita anggap sebagai harga wajar. Titik-titik biru yang berada di bawah garis diagonal mengindikasikan mobil-mobil yang dihargai lebih rendah dari prediksi model, menunjukkan bahwa mobil-mobil ini underprice. Ini dapat menjadi peluang strategis untuk meningkatkan margin keuntungan atau menarik lebih banyak pelanggan dengan menawarkan nilai lebih tinggi dibandingkan harga pasar yang diprediksi.

Sebaliknya, titik-titik merah yang berada di atas garis diagonal menunjukkan mobil-mobil yang dihargai lebih tinggi dari prediksi model, mengindikasikan bahwa mobil-mobil ini overprice. Mengetahui hal ini memungkinkan kita untuk meninjau kembali strategi penetapan harga kita agar lebih kompetitif, menghindari harga yang terlalu tinggi yang bisa mengurangi minat pembeli dan menghambat penjualan.

Distribusi titik-titik ini memberikan pemahaman yang mendalam tentang variasi harga mobil bekas di pasar dan membantu kita mengidentifikasi tren apakah mobil-mobil tersebut cenderung dijual lebih murah atau lebih mahal daripada prediksi. Informasi ini sangat berharga dalam merumuskan strategi penetapan harga yang lebih efektif dan meningkatkan daya saing bisnis.

Tabel 4.4 Hasil Pengukuran Metrik Kesalahan

MAE	MSE	RMSE	R²
78357505.4402942	1.7565720632247318e+16	132535733.41649157	0.6683129324399479

Setelah mendapatkan model dilakukan evaluasi model. Tahap ini digunakan untuk mengidentifikasi keakuratan suatu model. Ukuran yang digunakan untuk mengevaluasi hasil prediksi dari model adalah ukuran *R squared*, *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Root Mean Square Error (RMSE)*. Berdasarkan tabel 4.4 diketahui nilai RMSE sebesar 132535733.41649157 (pembulatan 132535733). Nilai RMSE digunakan untuk menggambarkan tingkat error data model yang digunakan. Semakin kecil nilai RMSE maka semakin tinggi nilai akurasi sistem. Kemudian nilai MAE diketahui sebesar 78357505.4402942 (pembulatan 78357505). Nilai MAE mengukur rata-

rata perbedaan absolut antara nilai aktual dan nilai prediksi. Semakin rendah nilai MAE, semakin baik model dalam memprediksi nilai. Selain itu juga diketahui nilai MSE sebesar $1.7565720632247318e+16$ (pembulatan $1.76e+16$). Nilai MSE ini mengukur rata-rata kuadrat kesalahan antara nilai aktual dan nilai prediksi. Nilai MSE yang rendah atau mendekati nol mengindikasikan bahwa hasil prediksi sesuai dengan data aktual. Selanjutnya yang terakhir adalah diketahui nilai R-square yang dihasilkan model sebesar 0.6683129324399479 (pembulatan 0.67 atau 67%). Semakin nilai R-square mendekati angka 1 berarti model yang dihasilkan memberikan keakuratan hasil yang baik.

Tabel 4.5 Pengujian Prediksi

No	Rasio Data	Akurasi
1	60:40	0.7318586075763591
2	70:30	0.7064803876716019
3	80:20	0.7089949704181817
4	90:10	0.7074626652635005

Berdasarkan hasil pengujian prediksi dengan beragam rasio data latih dan data uji seperti pada Tabel 4.5, rasio data uji dan data latih 60%:40% mendapatkan hasil akurasi prediksi tertinggi yaitu 73.18%.

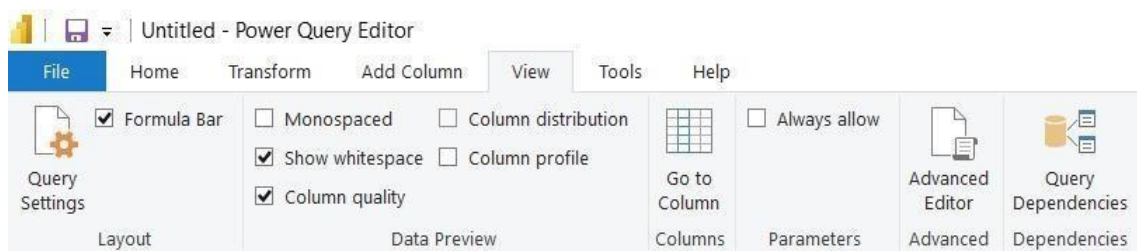
```
# Save Model
df = pd.DataFrame(final_mobkas)
df.to_csv('Final dataset with Predicted.csv', index=False)
from google.colab import files
files.download('Final dataset with Predicted.csv')
```

Gambar 4.15 Save Model

Setelah tahap pembuatan model prediksi dan analisis selesai, terdapat dua jenis dataset yang hampir memiliki kesamaan informasi atau *granularity*. Perbedaannya terletak pada satu dataset yang memiliki variabel prediksi harga mobil bekas, di mana beberapa variabel telah dilakukan pelabelan data. Sementara itu, dataset lainnya tidak memiliki variabel prediksi harga, dan variabel-variabel di dalamnya masih dalam kondisi awal (belum dilakukan pelabelan data). Untuk mempermudah tahapan selanjutnya, kedua dataset ini akan digabungkan menggunakan parameter merge.

4.2 ETL dan Perancangan Data Warehouse

Sebelum dilakukan visualisasi data, tahapan ETL atau extract, transform, load perlu dilakukan. Tahapan ini penting untuk mempersiapkan data agar visualisasi dapat optimal. Pada tahapan ini dilakukan pengecekan kembali data yang akan digunakan untuk memastikan kualitasnya sebelum proses ini dilakukan. Pada tahap pengecekan dataset ini bertujuan untuk memastikan bahwa dataset tersebut telah siap untuk divisualisasikan. Ini merupakan tahap yang penting untuk memverifikasi bahwa dataset telah dibersihkan dengan baik sebelumnya menggunakan baris kode PySpark. Caranya adalah dengan memilih *Transform Data*, lalu navigasi ke bagian *View*. Di bagian *View*, pilih opsi *Show whitespace* untuk melihat apakah masih ada data yang kosong atau null, serta opsi *Column quality* untuk memastikan tingkat validitas data itu sendiri sudah sempurna atau belum.



Gambar 4.16 Pengecekan Dataset

Selanjutnya adalah melihat hasil dari pengecekan yang sudah dilakukan, didapatkan bahwa sudah tidak ada lagi data yang kosong atau *null* serta pada bagian validitas semua variabelnya sudah mencapai 100% atau sempurna seperti gambar 4.17 untuk dataset yang sudah dilakukan penggabungan sebelumnya. Dari hasil ini dapat disimpulkan bahwa dataset sudah siap untuk dilakukan visualisasi data.

Queries [8] Table.SelectRows("#Reordered Columns1", each true)

	Tahun	Kilometer Mobil	Tipe Pemilik	Konsumsi BBM (kmpl)	Nama Mobil	Kapasitas Mesin (CC)
1	2010	72000	1	1	19 Maruti Wagon R LXI CNG	99
2	2018	25692	1	1	21 Maruti Ciaz Zeta	146
3	2018	10575	1	1	21 Maruti Ciaz Zeta	146
4	2015	64424	1	1	25 Maruti Swift VDI BSIV	124
5	2015	76496	1	1	25 Maruti Swift VDI BSIV	124
6	2015	63000	1	1	25 Maruti Swift VDI BSIV	124
7	2015	60095	1	1	25 Maruti Swift VDI BSIV	124
8	2015	23000	1	1	25 Maruti Swift VDI BSIV	124
9	2017	25000	1	1	28 Maruti Swift DDIS VDI	124
10	2015	55392	2	2	23 Maruti S Cross DDIS 200 Alpha	124
11	2013	54000	2	2	20 Maruti Alto K10 2010-2014 VXI	99
12	2012	87000	1	1	20 Maruti Ertiga VDI	124
13	2012	82085	1	1	20 Maruti Ertiga VDI	124
14	2012	168000	1	1	20 Maruti Ertiga VDI	124
15	2007	262000	4	4	12 Toyota Innova 2.5 V Diesel 7-seater	249
16	2016	36000	1	1	11 Toyota Innova Crysta 2.8 GX AT 8S	275
17	2016	40000	1	1	11 Toyota Innova Crysta 2.8 GX AT 8S	275
18	2015	41000	1	1	19 Hyundai Creta 1.6 CRDI SX Option	158
19	2015	41000	1	1	19 Hyundai Creta 1.6 CRDI SX Option	158
20	2015	61786	1	1	19 Hyundai Creta 1.6 CRDI SX Option	158
21	2012	75000	1	1	15 Hyundai EON LPG Era Plus Option	81
22	2010	45807	1	1	18 Hyundai i20 1.2 Magna	119
23	2010	67714	1	1	18 Hyundai i20 1.2 Magna	119
24	2010	18000	1	1	18 Hyundai i20 1.2 Magna	119
25	2010	18000	1	1	18 Hyundai i20 1.2 Magna	119

Gambar 4.17 Validitas Dataset yang sudah digabungkan

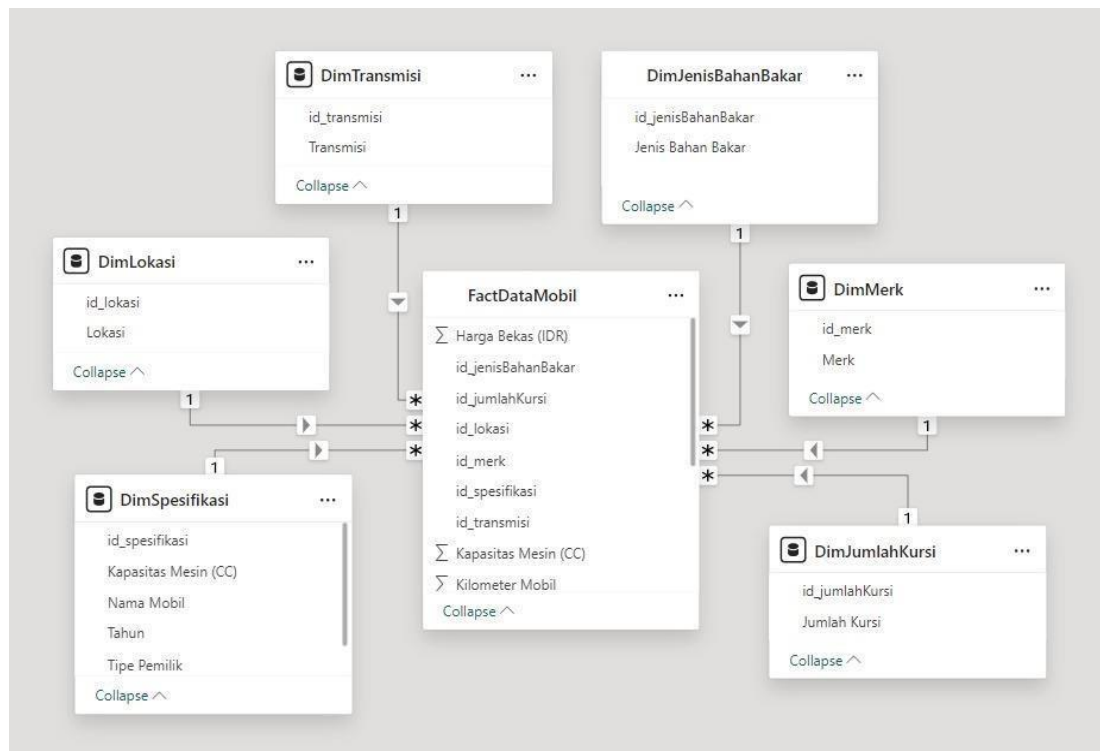
Setelah selesai dilakukannya proses *extract* serta *transform*, kini saatnya melanjutkan ke proses yang terakhir yaitu *load*. *Load* merupakan proses untuk memuat data yang telah dilakukan transformasi data pada proses sebelumnya ke dalam *data warehouse* Power BI. Data yang sudah selesai melalui proses transformasi sebelumnya berupa pengecekan dataset, kini data tersebut kemudian dimuat ke dalam *data warehouse* dengan menerapkan proses *apply* pada *power query* lalu data tersebut akan secara otomatis tersimpan di dalam *data warehouse* Power BI untuk kemudian proses pembuatan dashboard visualisasi dapat dilakukan.

Setelah melakukan proses ETL, perlu untuk membangun relasi tabel data sebelum merancang *data warehouse* pada Power BI yang berguna untuk mengintegrasikan atribut tabel satu dengan atribut tabel lain yang memiliki nilai serupa. Relasi yang terbentuk ditunjukkan pada Gambar 4.18.

Active	From: Table (Column)	To: Table (Column)
☒	FactDataMobil (id_jenisBahanBakar)	DimJenisBahanBakar (id_jenisBahanBakar)
☒	FactDataMobil (id_jumlahKursi)	DimJumlahKursi (id_jumlahKursi)
☒	FactDataMobil (id_lokasi)	DimLokasi (id_lokasi)
☒	FactDataMobil (id_merk)	DimMerk (id_merk)
☒	FactDataMobil (id_spesifikasi)	DimSpesifikasi (id_spesifikasi)
☒	FactDataMobil (id_transmisi)	DimTransmisi (id_transmisi)

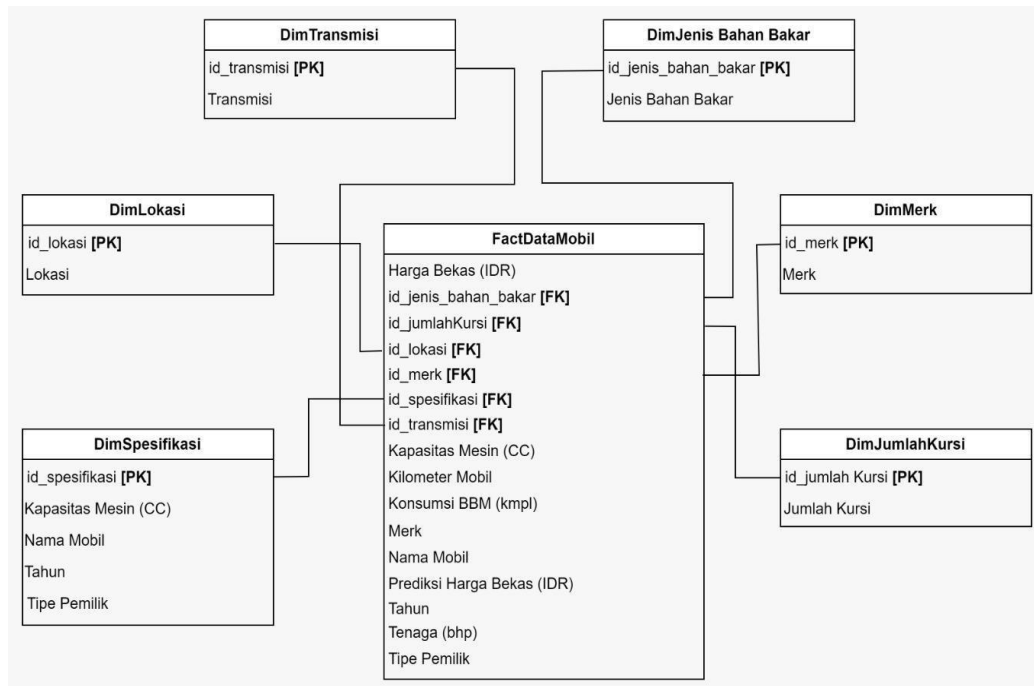
Gambar 4.18 Pengaturan Relasi Tabel

Untuk gambaran dari relasi tabel yang terbentuk setelah dilakukan pengaturan, dapat dilihat pada Gambar 4.19.



Gambar 4.19 Relasi Tabel Data yang Terbentuk

Setelah relasi tabel data terbentuk, kemudian langkah yang selanjutnya adalah dengan melakukan pemodelan *dimensional modelling*. Pada penelitian ini digunakan star schema. Opsi *star schema* digunakan dikarenakan tabel fakta terletak di tengah dan tabel-tabel lainnya, seperti tabel dimensi, terletak di sekelilingnya. Kelebihan dari *star schema* adalah strukturnya yang sederhana, dan perawatannya yang rendah. Namun, kekurangannya adalah kinerjanya yang kurang baik jika dibandingkan skema lainnya, dan jumlah tabel dimensi bisa sangat banyak. Hasil dari *star schema* pada Gambar 4.20 menggunakan "FactDataMobil" sebagai tabel fakta, sedangkan untuk tabel dimensi menggunakan spesifikasi, lokasi, merk, transmisi, jumlah kursi, dan jenis bahan bakar.



Gambar 4.20 *Dimensional Modelling*

Pada *star schema* diatas yang berperan sebagai *fact table* adalah tabel Data Mobil itu sendiri, dimana dalam tabel ini merupakan hasil dari penggabungan sebelumnya antara dataset yang sudah disertai hasil prediksi dengan dataset yang belum disertai hasil prediksi.

Tabel Data Mobil dijadikan sebagai fact table karena tabel data mobil merupakan tabel utama yang berisikan kumpulan *primary key* dari tabel-tabel lainnya. *Dimension table* untuk skema diatas adalah tabel spesifikasi, tabel lokasi, tabel transmisi, tabel jenis bahan bakar, tabel merk, serta tabel jumlah kursi karena *primary key* dari tabel-tabel tersebut berhubungan dengan salah satu *composite key* yang ada pada *fact table*.

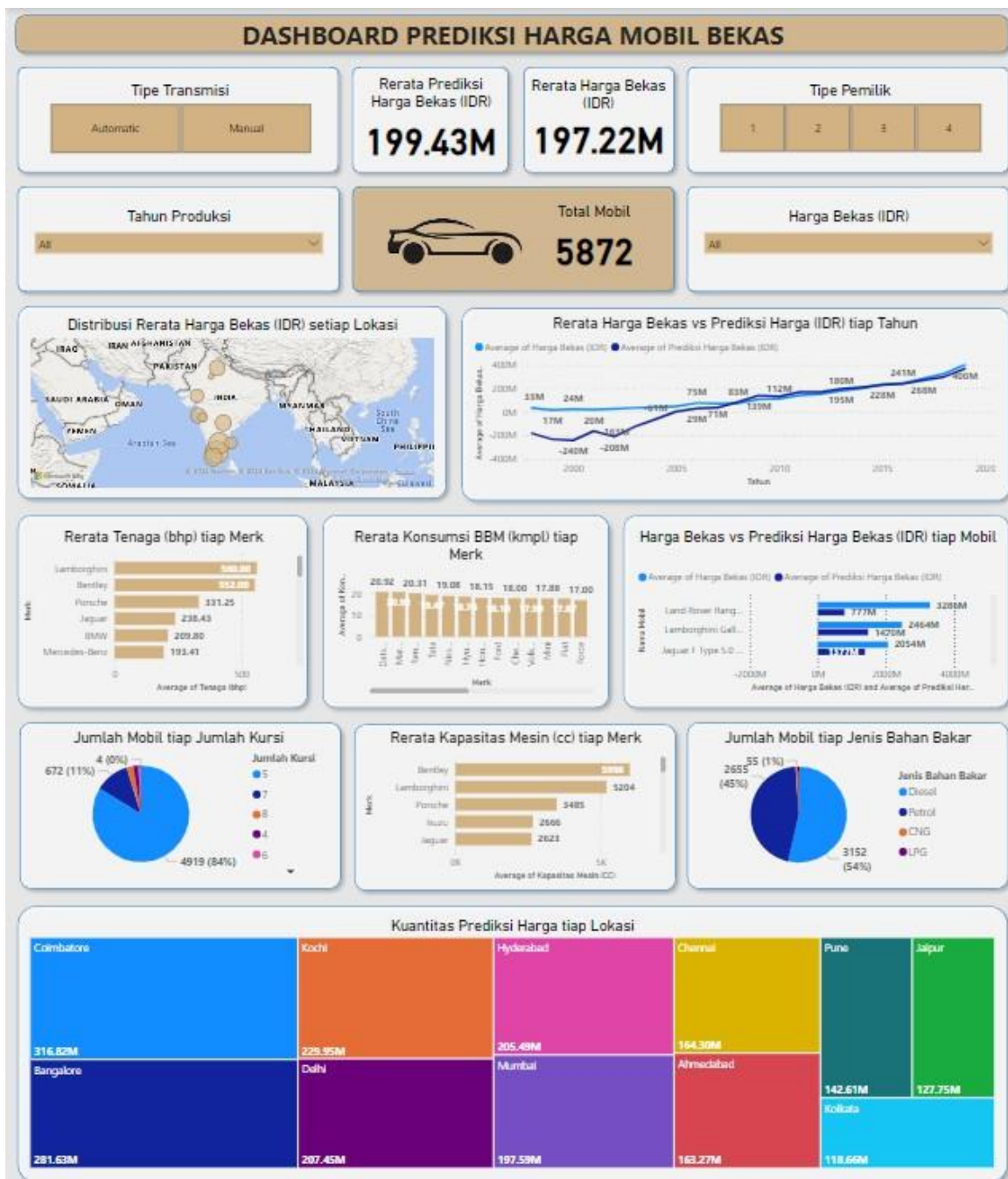
Untuk atribut-atribut diatas memiliki tipe data sebagai berikut:

- Atribut “Merk” memiliki tipe data *string*,
- Atribut “Transmisi” memiliki tipe data *string*,
- Atribut “Jenis Bahan Bakar” memiliki tipe data *string*,
- Atribut “Jumlah Kursi” memiliki tipe data *integer*,
- Atribut “Nama Mobil” memiliki tipe data *string*,
- Atribut “Kapasitas Mesin (CC)” memiliki tipe data *integer*,
- Atribut “Tenaga (bhp)” memiliki tipe data *integer*,
- Atribut “Lokasi” memiliki tipe data *string*,

- i. Atribut “Harga Bekas (Lakh)” memiliki tipe data *integer*,
- j. Atribut “Kilometer Mobil” memiliki tipe data *integer*,
- k. Atribut “Konsumsi BBM (kmpl)” memiliki tipe data *integer*,
- l. Atribut “Tahun” memiliki tipe data *integer*,
- m. Atribut “Tipe Pemilik” memiliki tipe data *integer*,
- n. Atribut “Prediksi Harga Bekas (Lakh)” memiliki tipe data *integer*.

4.3 Visualisasi Dashboard

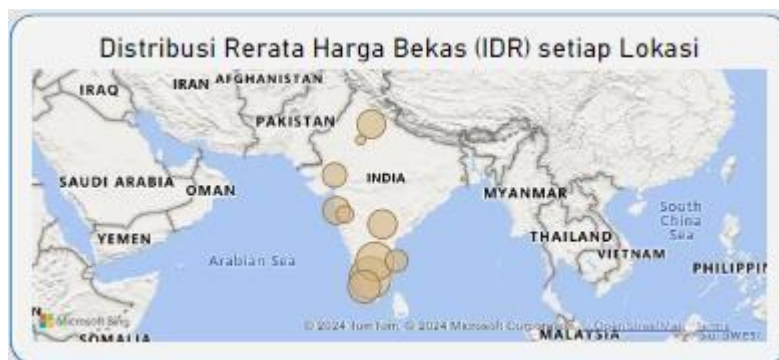
Setelah data diolah dan dimodelkan menggunakan regresi linier, langkah terakhir adalah merancang dashboard yang menyajikan informasi dalam bentuk visual. Dashboard ini bertujuan untuk memberikan wawasan menyeluruh dan membantu pemilik bisnis dalam pengambilan keputusan, seperti mengoptimalkan stok, meminimalkan kerugian, dan memaksimalkan keuntungan. Oleh karena itu, dashboard yang efektif tidak hanya menyajikan data, tetapi juga menyediakan alat bagi pengguna, dalam hal ini pemilik bisnis, untuk membuat keputusan yang lebih baik. Bentuk visualisasi dashboard yang terbentuk disajikan dalam Gambar 4.21.



Gambar 4.21 Rancangan Dashboard Power BI

Dashboard yang terbentuk ini menyajikan berbagai visualisasi data seperti distribusi rata-rata harga bekas setiap lokasi, perbandingan rata-rata harga bekas dengan rata-rata prediksi harga tiap tahunnya, rata-rata tenaga, konsumsi bbm, serta kapasitas mesin untuk setiap merk, perbandingan harga bekas dengan harga prediksi untuk tiap mobil, distribusi jumlah mobil tiap jumlah kursi dan jenis bahan bakar, serta kuantitas prediksi harga di tiap lokasinya.

Untuk meningkatkan interaktivitas *dashboard*, perlu ditambahkan elemen seperti slicer yang berfungsi untuk melakukan filter, sehingga pengguna dapat menyesuaikan tampilan data sesuai kebutuhan dan fokus pada segmen tertentu. Selain itu, menyertakan ringkasan statistik, seperti rata-rata dan *median*, dapat memberikan konteks tambahan tentang data yang ditampilkan, mempermudah pemilik bisnis dalam memperoleh informasi. Gambar 4.22 menampilkan grafik pertama dalam visualisasi, yang menunjukkan distribusi rata-rata harga bekas setiap lokasi dalam bentuk peta visual. Visualisasi ini memberikan informasi tambahan ketika pemilik bisnis melakukan navigasi atau menyeleksi area tertentu, memungkinkan mereka untuk meninjau data rinci seperti rata-rata harga mobil bekas, jumlah mobil, dan prediksi harga bekas untuk lokasi tertentu. Dengan fitur interaktif ini, pemilik bisnis dapat melakukan analisis lebih mendalam, membantu dalam pengambilan keputusan yang lebih tepat.



Gambar 4.22 Distribusi Rerata Harga Bekas (IDR) setiap Lokasi

Grafik selanjutnya menggambarkan perbandingan rata-rata antara harga aktual dan harga prediksi setiap tahunnya, ditampilkan pada Gambar 4.23. Visualisasi ini menggunakan line chart, dengan sumbu horizontal mewakili tahun dan sumbu vertikal mewakili harga rata-rata mobil. Terdapat dua garis yang menyajikan tren harga aktual dan harga prediksi: garis berwarna biru muda untuk rata-rata harga aktual dan garis berwarna biru tua untuk rata-rata harga prediksi yang diperoleh dari model regresi. Grafik ini menunjukkan bahwa dari tahun 1999 hingga 2004, perbandingan antara harga aktual dan harga prediksi cukup signifikan, dengan perbedaan mencolok. Namun, dari tahun 2005 hingga 2019, perbandingan antara harga aktual dan harga prediksi cenderung stabil. Hasil

analisis visual ini memberikan informasi bermanfaat bagi para pemilik bisnis mobil bekas, yang dapat digunakan untuk perencanaan bisnis dan strategi pemasaran.



Gambar 4.23 Analisis Tren Harga

Setelah visualisasi distribusi rata-rata harga bekas setiap lokasi dan perbandingan rata-rata antara harga aktual dengan harga prediksi, visualisasi selanjutnya menunjukkan informasi mendalam mengenai rata-rata tenaga (bhp), konsumsi bahan bakar (kmpl), dan kapasitas mesin (cc) tiap merek. Visualisasi ini menunjukkan bahwa semakin besar kapasitas mesin, umumnya mobil akan memiliki potensi tenaga yang lebih besar. Tenaga yang dihasilkan oleh mesin mobil berpengaruh pada performa dan kinerja kendaraan, termasuk akselerasi dan kecepatan maksimum, serta konsumsi bahan bakar.

Ketiga informasi ini digunakan sebagai dasar untuk strategi penetapan harga, pemasaran, atau alokasi sumber daya berdasarkan tenaga rata-rata (bhp), konsumsi bahan bakar (kmpl), dan kapasitas mesin (cc) dari setiap merek. Dengan cara ini, pemilik bisnis dapat mengambil kebijakan yang lebih tepat dan akurat berdasarkan pemahaman yang diperoleh dari hasil visualisasi data mengenai efisiensi, performa kendaraan, dan spesifikasi mesin dari berbagai merek.

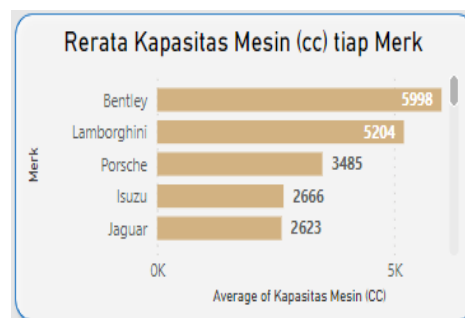
Grafik yang menunjukkan rata-rata tenaga pada merek disajikan dalam Gambar 4.24. Lamborghini dan Bentley memiliki perbedaan signifikan dibandingkan merek lain, karena keduanya fokus pada performa tinggi dalam desain mesin, menggunakan konfigurasi mesin yang lebih besar serta teknologi turbocharging dan supercharging. Meskipun demikian, kedua merek ini memiliki mesin yang dioptimalkan untuk efisiensi pembakaran dan menghasilkan tenaga maksimum, meskipun konsumsi bahan bakar mereka mungkin tidak lebih efisien dibandingkan merek lain. Target pasar mereka adalah mobil mewah dan performa tinggi, memungkinkan keduanya untuk fokus pada performa yang lebih tinggi.



Gambar 4.24 Rata-rata Tenaga (bhp) tiap Merk

Analisis lebih lanjut menunjukkan bahwa kedua merek tersebut kembali menduduki posisi teratas, seperti yang tercantum pada Gambar 4.25. Hal ini disebabkan oleh korelasi antara tenaga yang dihasilkan dan kapasitas mesin yang besar. Kedua faktor ini memang erat kaitannya, karena semakin besar tenaga yang dihasilkan, maka kapasitas mesin juga cenderung lebih besar.

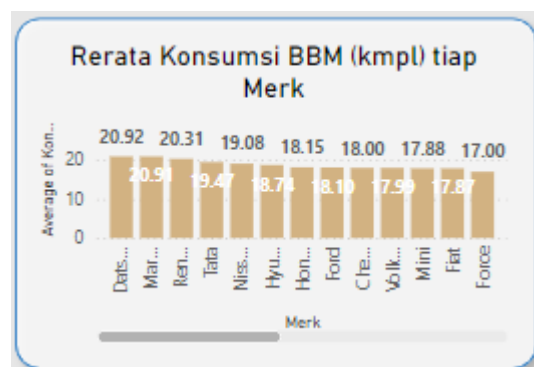
Pemahaman mendalam tentang hubungan antara tenaga dan kapasitas mesin memberikan dasar yang kuat bagi pengambilan keputusan bisnis yang cerdas. Melalui konteks ini, bisnis dapat merancang strategi pemasaran yang lebih terarah, mengoptimalkan, dan meningkatkan penjualan. Semua langkah ini didasarkan pada data dan tren yang jelas, sehingga memungkinkan pemilik bisnis untuk lebih responsif terhadap kebutuhan pasar.



Gambar 4.25 Rata-rata Kapasitas Mesin (cc) tiap Merk

Visualisasi selanjutnya memvisualisasikan rata-rata konsumsi bahan bakar tiap merek. Gambar 4.26 menunjukkan bahwa merek Datsun menunjukkan efisiensi bahan bakar yang lebih baik dibandingkan merek lain, meskipun perbedaannya tipis. Ini terjadi karena Datsun memiliki tenaga dan kapasitas mesin yang tidak terlalu besar, yang menghasilkan konsumsi bahan bakar yang lebih efisien.

Pemahaman mendalam tentang hubungan antara tenaga, kapasitas mesin, dan konsumsi bahan bakar memberikan landasan yang kuat untuk pengambilan keputusan bisnis yang cerdas. Dalam konteks ini, bisnis dapat merancang strategi pemasaran yang lebih terarah dan tepat sasaran, serta melakukan optimalisasi dan peningkatan penjualan. Dengan menyoroti keunggulan efisiensi bahan bakar dari merek Maruti, strategi pemasaran dapat difokuskan pada atribut ini untuk menarik pelanggan yang peduli dengan efisiensi dan hemat biaya.



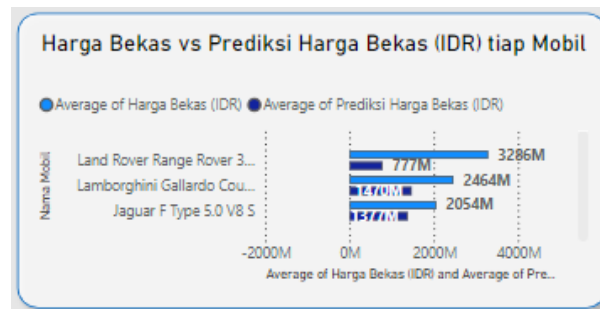
Gambar 4.26 Rata-rata Konsumsi BBM (kmpl) tiap Merk

Setelah melakukan analisis mengenai hubungan antara tenaga, kapasitas mesin, dan konsumsi bahan bakar, langkah selanjutnya adalah melakukan komparasi antara harga bekas dengan prediksi harga bekas untuk setiap mobil. Dengan komparasi ini, pemilik bisnis dapat mengevaluasi keakuratan prediksi harga bekas dan mengidentifikasi mobil-mobil yang memiliki perbedaan harga yang signifikan antara harga aktual dan harga prediksi, seperti yang terlihat pada Gambar 4.27.

Informasi ini sangat penting bagi pemilik bisnis untuk mengambil keputusan secara strategis yang berhubungan dengan penilaian aset dan pengambilan keputusan. Tidak hanya itu, visualisasi ini juga membantu pemilik bisnis untuk mengetahui performa prediksi harga bekas dengan harga aktual berdasarkan masing-masing mobil yang berbeda dalam satu tampilan, sehingga memudahkan dalam penentuan strategi pemasaran ataupun alokasi sumber daya yang efisien dan melihat secara lebih mendetail mengenai selisih harga aktual dan harga prediksi untuk masing-masing mobil.

Grafik yang digunakan dalam visualisasi ini adalah diagram batang yang ditampilkan secara horizontal, memperlihatkan dua batang untuk setiap nama mobil dengan satu bilah mewakili harga prediksi dan bilah lainnya mewakili harga aktual, sedangkan sumbu

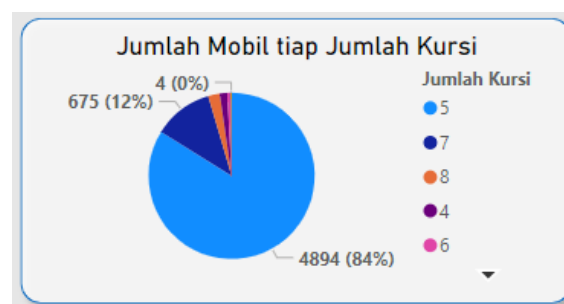
vertikal menyatakan harga mobil tersebut. Grafik ini menyediakan wawasan mendalam mengenai sejauh mana model prediksi berhasil atau tidak untuk tiap mobil, sehingga dapat dengan mudah mendeteksi peningkatan maupun penyempurnaan model. Secara keseluruhan, visualisasi ini menyajikan gambaran yang jelas mengenai bagaimana prediksi harga mobil dengan harga aktual setiap mobil.



Gambar 4.27 Komparasi Harga Bekas vs Prediksi Harga tiap Mobil

Visualisasi selanjutnya dalam dashboard ini juga terdapat sejumlah visualisasi yang menyajikan gambaran umum mengenai data. Gambar 4.28 menghadirkan visualisasi data dalam bentuk grafik pie yang membahas secara rinci dalam data tersebut. Grafik pie pada Gambar 4.28 memberikan gambaran yang jelas tentang pembagian persentase jumlah mobil berdasarkan jumlah kursinya.

Dengan sejumlah warna yang berbeda, grafik pie ini mencerminkan proporsi mobil berdasarkan jumlah kursi yang dimilikinya. Penggunaan warna dan proporsi yang tepat memudahkan pemilik bisnis untuk memahami total jumlah mobil yang ada berdasarkan jumlah kursi. Selain itu, membantu dalam pengambilan keputusan terkait inventarisasi dan penjualan mobil. Ini memberikan wawasan yang lebih dalam tentang preferensi konsumen terkait jenis mobil yang mereka pilih, serta memandu keputusan bisnis terkait stok dan pemasaran mobil yang efektif.

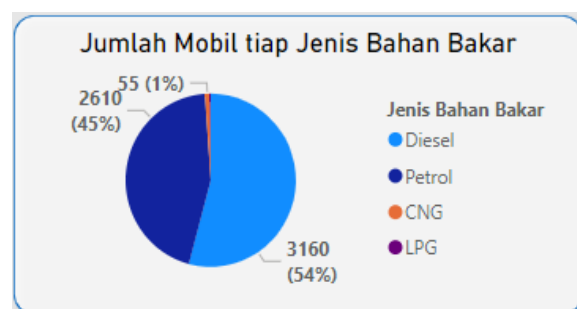


Gambar 4.28 Grafik Proporsi Jumlah Mobil tiap Jumlah Kursi

Dalam visualisasi yang disajikan di atas, mobil dengan jumlah kursi sebanyak lima kursi menduduki peringkat teratas. Hal ini disebabkan karena dengan lima kursi sering dianggap ideal untuk keluarga kecil karena dapat menampung anggota keluarga inti dengan nyaman. Selain itu, mobil lima kursi menawarkan keseimbangan yang baik antara kapasitas penumpang dan ruang bagasi, menjadikannya cocok untuk berbagai keperluan, baik harian maupun perjalanan jarak jauh. Fleksibilitasnya membuatnya menjadi pilihan yang populer di pasaran.

Untuk visualisasi yang terakhir, terdapat gambaran yang jelas mengenai klasifikasi mobil berdasarkan jenis bahan bakar yang masih dalam bentuk grafik pie yang membahas secara rinci dalam data tersebut. Grafik pie pada Gambar 4.29 memberikan gambaran yang jelas tentang pembagian persentase jumlah mobil berdasarkan jenis bahan bakarnya.

Dengan sejumlah warna yang berbeda, grafik pie ini mencerminkan proporsi mobil berdasarkan jumlah kursi yang dimilikinya. Penggunaan warna dan proporsi yang tepat memudahkan pemilik bisnis untuk memahami total jumlah mobil yang ada berdasarkan jenis bahan bakar. Selain itu, membantu dalam pengambilan keputusan terkait inventarisasi dan penjualan mobil. Dengan memberikan wawasan yang lebih mendalam tentang preferensi konsumen terkait jenis mobil yang mereka pilih, grafik ini juga dapat membimbing keputusan bisnis terkait stok dan strategi pemasaran mobil yang efektif.



Gambar 4.29 Komparasi Harga Bekas vs Prediksi Harga tiap Mobil

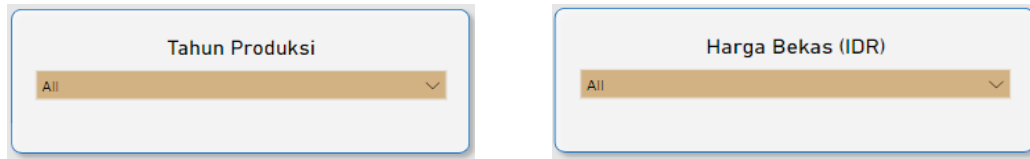
Terlihat pada Gambar 4.29 bahwa mobil dengan bahan bakar diesel lebih dominan dibandingkan dengan mobil berbahan bakar bensin, CNG, dan LNG. Ini karena di India, diesel biasanya lebih murah sehingga lebih ekonomis. Efisiensi bahan bakar yang tinggi, torsi yang besar, dan harga yang lebih rendah membuat mesin diesel menjadi pilihan yang menarik, terutama untuk kendaraan komersial dan mereka yang mencari efisiensi bahan bakar yang baik.

Seperti yang telah dijabarkan diatas bahwa dalam dashboard visualisasi perlu ditambahkan elemen seperti slicer yang berfungsi untuk melakukan filter, sehingga pengguna dapat menyesuaikan tampilan data sesuai kebutuhan dan fokus pada segmen tertentu. Selain itu, menyertakan ringkasan statistik, seperti rata-rata dan median, dapat memberikan konteks tambahan tentang data yang ditampilkan, mempermudah pemilik bisnis dalam memperoleh informasi. Elemen slicer pertama yang digunakan dalam dashboard ini, berfokus terhadap pemilihan Tipe Transmisi dan Tipe Pemilik menggunakan slicer visual dengan style yang digunakan adalah tile dikarenakan jumlah pilihan data di dalamnya tidak terlalu banyak. Tipe transmisi memiliki dua pilihan diantaranya automatic dan manual, sementara untuk tipe pemilik terdiri dari empat kategori yaitu pertama, kedua, ketiga, dan keempat. Visualisasi ini memiliki kemudahan untuk memilih dan melihat data yang lebih detail mengenai pilihan yang dipilih, seperti yang ditampilkan dalam Gambar 4.30. Dengan adanya menu pilihan ini, pemilik bisnis dapat membuat kombinasi guna melakukan segmentasi pasar yang lebih efektif serta melihat mobil mana saja yang masuk ke dalam kategori transmisi dan tipe pemilik yang telah ditentukan.



Gambar 4.30 Slicer Tipe Transmisi dan Tipe Pemilik

Elemen *slicer* berikutnya adalah pemilihan Tahun Produksi dan Harga Bekas (IDR) menggunakan *slicer* visual dengan *style* yang digunakan adalah *dropdown* karena banyaknya pilihan data. Untuk tahun produksi sendiri mencakup pilihan dari tahun 1998 sampai dengan 2019, sementara untuk harga bekas memiliki pilihan dari yang terendah 9.035.400 IDR sampai yang tertinggi 3.285.600.000 IDR. Visualisasi ini memudahkan pemilik bisnis untuk memfilter data mobil berdasarkan tahun produksi dan rentang harga bekas yang dipilih sebagaimana ditampilkan Gambar 4.31. Dengan fitur ini, pemilik bisnis dapat dengan mudah mengidentifikasi tren harga dan mobil dari berbagai tahun produksi.



Gambar 4.31 Slicer Tahun Produksi dan Harga Bekas (IDR)

Selain elemen *slicer* diatas terdapat juga kebutuhan akan visualisasi data yang mampu menyimpulkan analisis bisnis dengan lebih menyeluruh, seperti rata-rata harga bekas, harga prediksi, serta jumlah mobil yang disajikan dengan menggunakan *card* visual. Visualisasi ini memiliki beberapa manfaat, termasuk menampilkan informasi penting secara ringkas dan mudah dipahami. Informasi tersebut akan secara langsung ditampilkan tanpa perlu dilakukan interpretasi lebih lanjut, hal tersebut dapat membantu pemilik bisnis untuk memperoleh gambaran umum dari data yang disajikan dengan cepat. Tidak hanya itu, melalui visualisasi ini, pemilik bisnis dapat dengan mudah melihat perbandingan antara rata-rata harga bekas, prediksi harga, dan jumlah mobil berdasarkan opsi yang telah dipilih sebelumnya.

Sehingga, dengan adanya visualisasi data yang menyeluruh dan relevan, bisnis dapat mengumpulkan informasi penting yang dapat menentukan arah kebijakan maupun rencana bisnis yang lebih baik. Analisis ini bukan hanya menyediakan informasi terperinci mengenai rata-rata harga aktual, harga prediksi, dan jumlah mobil, melainkan dapat membantu pemilik bisnis untuk melihat tren dan memudahkan dalam proses pengambilan keputusan.



Gambar 4.32 Card

Terlihat pada Gambar 4.32 bahwa untuk rata-rata prediksi harga bekas yang diperoleh dari hasil pemodelan sebesar 199,43 juta rupiah, sedangkan untuk rata-rata harga bekasnya tercatat sebesar 197,22 juta rupiah. Rerata Harga Bekas (IDR) menunjukkan harga rata-rata sebenarnya dari mobil bekas, sementara bagian "Total Mobil" menampilkan jumlah total mobil yang terdaftar, yaitu 5.872 unit.

4.4 Pengujian Dasbor

Selain dengan memvisualisasikan hasil dari pemodelan regresi linier, perlu juga untuk dilakukan beberapa pengujian seperti pengujian akurasi data serta pengujian untuk mengukur tingkat kepuasan pengguna terhadap dashboard yang telah dibuat.

4.4.1 Pengujian Akurasi Data

Pengujian akurasi data dilakukan dengan cara menguji *dashboard* yang membandingkan harga *rill* mobil bekas yang terdaftar di berbagai *marketplace* di India dengan harga prediksi yang dihasilkan dari pemodelan sebelumnya. Untuk mendapatkan harga *rill*, data dikumpulkan dari *platform* seperti OLX dan Cars24, yang merupakan situs populer untuk membeli dan menjual mobil bekas di India. Proses pengumpulan data ini melibatkan pencatatan informasi penting mengenai variabel mobil, seperti tahun produksi, merek, model, dan lokasi penjualan, yang dilakukan secara acak. Variabel-variabel ini dipilih karena data yang tersedia di *marketplace* tidak mungkin tersedia secara detail, sehingga peneliti melakukannya secara acak berdasarkan informasi umum.

Tabel 4.6 Pencatatan Informasi secara Acak

No	Tahun	Merek	Model	Lokasi	Sumber	Harga Bekas	Harga Bekas (dataset)
1	2012	Maruti	Ertiga VDI	Chennai	OLX	₹ 5,87,000	₹ 6,00,000
2	2015	Hyundai	Creta 1.6 SX Option	Pune	Cars24	₹ 7.63,000	₹ 12,50,000

Setelah mengumpulkan data harga *rill*, langkah selanjutnya adalah menggunakan hasil pemodelan regresi linier untuk membandingkan prediksi harga mobil bekas dengan harga *rill* yang telah diperoleh. Perbandingan ini dilakukan dengan mengevaluasi mobil-mobil berdasarkan data harga *rill* yang dikumpulkan. Jika harga *rill* lebih rendah dari harga

prediksi, mobil tersebut dianggap *underprice*, menunjukkan bahwa harga jualnya mungkin lebih murah dibandingkan nilai yang diprediksi oleh model. Sebaliknya, jika harga *rill* lebih tinggi dari harga prediksi, mobil tersebut dianggap *overprice*, yang berarti harga jualnya lebih tinggi dari yang diperkirakan. Melalui analisis ini, dapat mengukur seberapa akurat model regresi linier dalam memprediksi harga mobil bekas sesuai dengan harga pasarnya.

Tabel 4.7 Perbandingan Prediksi Harga dengan Harga *Rill*

No	Tahun	Merk	Model	Lokasi	Sumber	Harga Bekas (<i>marketplace</i>)	Harga Bekas (dataset)	Prediksi Harga (IDR)	Prediksi Harga (Lakh)
1	2012	Maruti	Ertiga VDI	Chennai	OLX	₹ 5,87,000	₹ 6,00,000	73.347 .596	₹ 3,94,000
2	2015	Hyundai	Creta 1.6 SX Option	Pune	Cars24	₹ 7.63,000	₹ 12,50,000	271.42 8.769	₹ 14,58,000

Tabel 4.7 menunjukkan perbandingan antara harga mobil bekas secara *rill* yang didapatkan dari dua *marketplace* populer untuk membeli dan menjual mobil bekas di India dengan prediksi harga yang telah dihasilkan dari pemodelan regresi linier sebelumnya. Pada Tabel ini dapat diketahui bahwa mobil pertama dengan tahun produksi 2012, merk Maruti dan model Ertiga VDI yang dijual pada lokasi Chennai yang diambil dari data OLX diketahui bahwa mengalami *overprice*, dikarenakan untuk harga *rill* lebih tinggi daripada harga prediksi dari pemodelan. Selanjutnya untuk data mobil kedua dengan tahun produksi 2015, merk Hyundai dan model Creta 1.6 SX Option yang dijual pada lokasi Pune yang diambil dari data Cars24 diketahui bahwa mengalami *underprice*, yang dikarenakan harga *rill* lebih rendah jika dibanding dengan harga prediksi yang dihasilkan dari pemodelan. Dapat disimpulkan bahwa, perbandingan data dari dataset dan data *rill* yang dikumpulkan dari OLX dan Cars24, dapat disimpulkan bahwa model prediksi yang dibangun mampu memberikan prediksi harga yang mendekati harga sebenarnya di pasar. Terlihat bahwa data yang berasal dari dataset menunjukkan hasil yang berbeda dengan data riil. Hal ini mengindikasikan bahwa data riil yang dikumpulkan dari pasar lebih dinamis dan fluktuatif, sedangkan data dataset cenderung memiliki rentang harga yang lebih stabil. Meskipun demikian, model prediksi masih mampu menangkap pola dan tren harga umum, sehingga dianggap cukup akurat untuk memprediksi harga mobil bekas.

4.4.2 Pengujian Dasbor

Teknik yang diterapkan dalam tahap pengujian ini adalah *System Usability Scale* (SUS), yang bertujuan untuk menilai tingkat kepuasan pengguna terhadap dashboard yang telah dikembangkan. Pengujian dilakukan dengan meminta tiga responden untuk menggunakan dashboard tersebut dan mengisi kuesioner yang telah disiapkan. Dengan responden 1 adalah pemilik *showroom* mobil bekas yang sudah berjalan selama 5 tahun, responden 2 adalah pemilik *showroom* mobil bekas yang berpengalaman lebih dari 5 tahun, serta untuk responden 3 adalah pemilik *showroom* mobil bekas yang baru saja berdiri kurang lebih satu tahun terakhir ini. Kuesioner ini terdiri dari 10 pernyataan, yang diberi label dari Q1 hingga Q10. Hasil dari pengujian ini dapat dilihat pada Tabel 4.8.

Tabel 4.8 Hasil Pengujian SUS

Responden	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Responden 1	4	1	5	2	4	1	4	2	5	2
Responden 2	3	2	4	1	5	1	5	1	4	3
Responden 3	4	2	5	1	5	2	5	2	5	3

Perhitungan SUS dilakukan dengan cara untuk pertanyaan bernomor ganjil seperti nomor 1,3,5,7, dan 9 kontribusi skor adalah skala posisi dikurangi 1. Untuk pertanyaan bernomor genap seperti nomor 2,4,6,8 dan 10, kontribusi adalah 5 minus posisi skala. Kalikan jumlah nilai sebesar 2,5 untuk mendapatkan nilai keseluruhan sehingga didapatkan hasil seperti Tabel 4.9.

Tabel 4.9 Hasil Akhir Perhitungan Pengujian SUS

No	Q1	Q2	Q 3	Q4	Q5	Q6	Q 7	Q8	Q 9	Q1 0	Juml ah	Nilai
Responden 1	3	4	4	3	3	4	3	3	4	3	34	85
Responden 2	2	3	3	4	4	4	4	4	3	2	33	82,5
Responden 3	3	3	4	4	4	3	4	3	4	2	34	85
												84,16

Berdasarkan perhitungan menggunakan pengujian SUS didapatkan hasil akhir sebesar 84,16, sehingga *dashboard* yang dibuat dapat dianggap baik sekali (*excellent*).

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan studi kasus ini, dapat ditarik beberapa kesimpulan sebagai berikut:

- 1 Penelitian ini berhasil mengatasi kesulitan dalam menentukan harga mobil bekas dengan menerapkan model regresi linier, yang mencapai akurasi 73.18% dengan rasio data 60:40. Hasil ini membantu mengurangi potensi overpriced atau underpriced dalam transaksi.
- 2 Penelitian ini juga meningkatkan efektivitas presentasi hasil prediksi harga melalui visualisasi yang baik. Pengujian System Usability Scale (SUS) menunjukkan nilai *excellent*, menegaskan bahwa hasil prediksi disajikan dengan cara yang informatif dan mudah dipahami.

5.2 Saran

Berdasarkan penelitian yang dilakukan, terdapat beberapa saran yang diberikan oleh peneliti sebagai berikut:

- 1 Penelitian ini menggunakan metode regresi linier. Penelitian selanjutnya dapat mengeksplorasi dan menerapkan metode machine learning lain yang lebih kompleks serta efektif dalam memprediksi harga mobil bekas, yang harapannya dapat menghasilkan model secara maksimal dan dapat lebih memperbaiki kembali untuk hasil pengujian prediksinya.
- 2 Penelitian selanjutnya diharapkan dapat meningkatkan hasil prediksi secara maksimal dengan menambahkan variabel-variabel penelitian yang lebih sempurna dan memiliki wawasan yang lebih luas terhadap aspek lainnya dalam penelitian.

DAFTAR PUSTAKA

- Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics: Conference Series*, 1142, 012012. <https://doi.org/10.1088/1742-6596/1142/1/012012>
- Azmi, U., Hadi, Z. N., & Soraya, S. (2020). ARDL METHOD: Forecasting Data Curah Hujan Harian NTB. *Jurnal Varian*, 3(2), 73–82. <https://doi.org/10.30812/varian.v3i2.627>
- Baihaqi, W. M., & Dianingrum, M. (2019). Regresi Linier Sederhana untuk Memprediksi Kunjungan Pasien di Rumah Sakit Berdasarkan Jenis Layanan dan Umur Pasien. *Jurnal SIMETRIS*, 10(2).
- Becker, L., & Gould, E. (2019). Microsoft Power BI: Extending Excel to Manipulate, Analyze, and Visualize Diverse Data. *Serials Review*, 45, 1–5. <https://doi.org/10.1080/00987913.2019.1644891>
- Bengnga, A., & Ishak, R. (2018). Prediksi Jumlah Mahasiswa Registrasi Per Semester Menggunakan Linier Regresi Pada Universitas Ichsan Gorontalo. *ILKOM Jurnal Ilmiah*, 10(2), 136–143. <https://doi.org/10.33096/ilkom.v10i2.274.136-143>
- BlueBook, I. (n.d.). *Used Car Industry Report FY2022-23 6th Edition*. DasWeltAuto and Car&Bike.
- Bororing, J. E. (2022). Implementasi Dashboard Microsoft Power BI Untuk Visualisasi Data Covid 19 Indonesia. *Jurnal Informasi Interaktif*, 7.
- Budilaksana, D., Sukarsa, I. M., & Wiranatha, A. A. K. A. C. (2021). Implementing kNearest Neighbor Methods to Predict Car Prices. *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)*, 58. <https://doi.org/10.24843/JIM.2021.v09.i01.p06>
- Carneiro, T., Medeiros Da Nobrega, R. V., Nepomuceno, T., Bian, G.-B., De Albuquerque, V. H. C., & Filho, P. P. R. (2018). Performance Analysis of Google Colaboratory as a Tool

- for Accelerating Deep Learning Applications. *IEEE Access*, 6, 61677–61685.
<https://doi.org/10.1109/ACCESS.2018.2874767>
- Chen, X., Gu, S., Deng, X., & Huang, L. (2022). Used Car Prices in India: What about Future? *Business and Management Research*, 211.
- Connolly, T. M., & Begg, C. E. (2005). Database systems: A practical approach to design, implementation, and management (4th ed). Harlow, Essex, England ; New York: *Addison-Wesley*.
- Connolly, T. M., & Begg, C. E. (2010). Database systems: A practical approach to design, implementation, and management (5. ed., internat. ed). Boston: *Pearson Education*.
- Darman, R. (2018). Analisis Visualisasi dan Pemetaan Data Tanaman Padi di Indonesia Menggunakan Microsoft Power BI. *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, 4(2), 156. <https://doi.org/10.24014/rmsi.v4i2.5271>
- Ehsani, Z., & Ehsani, M. H. (2015). Effect Of Quality and Price on Customer Satisfaction and Commitment in Iran Auto Industry. *International Journal of Service Science, Management and Engineering*, 52–56.
- Fadhilah, M., & Ali, I. (2023). Prediksi Jumlah Produksi Sablon Tahun Menggunakan Algoritma Regresi Linear di Nolbas SVN. *INTERNAL (Information System Journal)*, 22-32.
- Faiqoh, H. (2023). Penerapan Ensemble Feature Selection untuk Mengurangi Dimensionalitas dalam Prediksi Data Time Series (Skripsi, Universitas Islam Negeri Syarif Hidayatullah Jakarta). <https://repository.uinjkt.ac.id/dspace/bitstream/123456789/67099/1/HANA%20FAIQOH-FST.pdf>
- Farizal, Rachman, A., & Al Rasyid, H. (2014). Model Peramalan Konsumsi Bahan Bakar Jenis Premium di Indonesia dengan Regresi Linier Berganda. *Jurnal Ilmiah Teknik Industri*, 13(2).

- Febriansyah, R. A. (2022). Dashboard Informatif Persebaran Covid-19 di Indonesia Pada Data dan Artificial Intelligence Menggunakan Microsoft Azure Machine Learning di PT Microsoft Indonesia. (Skripsi, Universitas Dinamika). <https://repository.dinamika.ac.id/id/eprint/6479/9/19410200020-2022-UNIVERSITASDINAMIKA.pdf>
- Fengki, F. (2018). Implementasi Regresi Linier untuk Memprediksi Lama Waktu Pengiriman Catering Kepada Konsumen Studi Kasus Home Catering Malang (Skripsi, Universitas Islam Negeri Maulana Malik Ibrahim). <http://etheses.uin-malang.ac.id/12531/>
- Ganesha, H. (2018). Perancangan Data Warehouse untuk Kebutuhan Sistem Penunjang Keputusan Divisi Revenue Assurance Studi Kasus: PT. XXX. *InfoTekJar (Jurnal Nasional Informatika dan Teknologi Jaringan)*, 3(1), 74–80. <https://doi.org/10.30743/infotekjar.v3i1.491>
- Gowthami, K., & Kumar, M. R. P. (2017). Study On Business Intelligence Tools For Enterprise Dashboard Development. *International Research Journal of Engineering and Technology (IRJET)*, 04(04).
- Hasibuan, E., & Karim, A. (2022). Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web. *Jurnal Ilmiah Komputasi*, 21(4). <https://doi.org/10.32409/jikstik.21.4.3327>
- Hasibuan, L. H., & Musthofa, S. (2022). Penerapan Metode Regresi Linear Sederhana Untuk Prediksi Harga Beras di Kota Padang. *JOSTECH: Journal of Science and Technology*, 2(1), 85–95. <https://doi.org/10.15548/jostech.v2i1.3802>
- Herawati, I. M., & Azahra, D. (2024). Evaluasi Usability Website Jasuda.net Menggunakan System Usability Scale (SUS). *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 9(2), 994–1000. <https://doi.org/10.29100/jipi.v9i2.4328>

- Hidayat, R., & Dewi Lubis, D. I. (2017). Pengaruh Citra Merek dan Harga terhadap Keputusan Pembelian pada Sekolah Tinggi Ilmu Manajemen Sukma Medan. *Jurnal Ilman*, 5(1).
- Ihzaniah, L. S., Setiawan, A., & Wijaya, R. W. N. (2023). Perbandingan Kinerja Metode Regresi K-Nearest Neighbor dan Metode Regresi Linear Berganda pada Data Boston Housing. *Jambura Journal of Probability and Statistics*, 4(1), 17–29. <https://doi.org/10.34312/jjps.v4i1.18948>
- Indarwati, T., Irawati, T., & Rimawati, E. (2019). Penggunaan Metode Linear Regression Untuk Prediksi Penjualan Smartphone. *Jurnal Teknologi Informasi dan Komunikasi IKomSiN*, 6(2). <https://doi.org/10.30646/tikomsin.v6i2.369>
- Jusuf, D. N. B. (2023). Implementasi Intelijen Bisnis Untuk Visualisasi Pola Pembelian Dengan Algoritma FP-Growth. *Edusaintek*, 483-507. <https://doi.org/10.47668/edusaintek.v11i2.1057>
- Kriswantara, B., & Sadikin, R. (2022). Used Car Price Prediction with Random Forest Regressor Model. *Journal of Information Systems*.
- Kumar, P., Manish, & Vatsalya, K. (2023). Is India's Used Car Market Larger than Previously Thought? *International Journal for Research in Applied Science and Engineering Technology*, 11(9), 87–91. <https://doi.org/10.22214/ijraset.2022.43459>
- Kusumawati, N., Marisa, F., & Wijaya, I. D. (2017). Prediksi Kurs Rupiah Terhadap Dollar Amerika Dengan Menggunakan Metode Regresi Linear. *JIMP - Jurnal Informatika Merdeka Pasuruan*, 2(3). <https://doi.org/10.37438/jimp.v2i3.79>
- Lewis, J. R., & Sauro, J. (2018). Item Benchmarks for the System. *Journal of Usability Studies*, 13(3), 158–167.

- Malik, M. H. A. (2023). Analisis Visualisasi Data Prestasi Mahasiswa Universitas Lampung. (Skripsi, Universitas Lampung). <http://digilib.unila.ac.id/73407/3/SKRIPSI%20TANPA%20BAB%20PEMBAHASAN.pdf>
- Marbun, M., Sihotang, H. T., & Nababan, M. A. (2018). Perancangan Sistem Peramalan Jumlah Wisatawan Asing. *Jurnal Mantik Penusa*, 2(1).
- Maryanto, H., Putranto, B. P. D., Kartadie, R., Guntara, M., & Saptoto, R. (2023). Analysis And Design Of Data Warehouse And Data Mart Budget. *Journal of Intelligent Software Systems*, 2(1), 6. <https://doi.org/10.26798/jiss.v2i1.927>
- Muhaemin, M. (2019). Desain Arsitektur Teknologi Data Warehouse Untuk Mendukung Manajemen Pengawasan Dan Pengendalian PNS Studi Kasus: Badan Kepegawaian Negara. *Just IT : Jurnal Sistem Informasi, Teknologi Informasi Dan Komputer*, 9(2), 163–168. <https://doi.org/10.24853/justit.9.2.163-168>
- Munte, C. (2022). Implementasi Data Mining Untuk Memprediksi Prestasi Siswa/Siswi Menggunakan Metode Regresi Linear Berganda (Studi Kasus: SMK Negeri 1 Percut Sei Tuan). (Skripsi, Universitas Medan Area). <https://repositori.uma.ac.id/jspui/bitstream/123456789/17428/1/178160080%20-%20Casma%20Munte%20-%20Fulltext.pdf>
- Muriyatmoko, D. (2018). Analisa Volume Terhadap Sitasi Menggunakan Regresi Linier Pada Jurnal Bereputasi Di Indonesia. *Jurnal Ilmiah SimanteC*, 6.
- Nasution, S. L., Limbong, C. H., & Ramadhan, D. A. (2020). Pengaruh Kualitas Produk, Citra Merek, Kepercayaan, Kemudahan, Dan Harga Terhadap Keputusan Pembelian Pada E-Commerce Shopee (Survei Pada Mahasiswa S1 Fakultas Ekonomi Jurusan Manajemen Universitas Labuhan Batu). *Ecobisma (Jurnal Ekonomi, Bisnis dan Manajemen)*, 7(1), 43–53. <https://doi.org/10.36987/ecobi.v7i1.1528>.

- Ofela, H., & Agustin, S. (2016). Pengaruh Harga, Kualitas Produk dan Kualitas Pelayanan Terhadap Kepuasan Konsumen Kebab Kingabi. *Jurnal Ilmu dan Riset Manajemen*, 5.
- Oslan, Y., & Kristanto, H. (2019). Proses ETL (Extract, Transformation, Loading) Data Warehouse untuk Peningkatan Kinerja Biodata dalam Menyajikan Profil Mahasiswa dari Dimensi Asal Sekolah. *Research Fair Unisri 2019* 3(1).
- Pohan, D. A., & Dar, M. H. (2022). Penerapan Data Mining untuk Prediksi Penjualan Produk Sepatu Terlaris Menggunakan Metode Regresi Linier Sederhana. *InfoTekjar : Jurnal Nasional Informatika dan Teknologi Jaringan*, 6(2).
- Ponniah, P. (2001). *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*. Boston: Wiley.
- Ponniah, P. (2004). *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*. John Wiley & Sons.
- Prasetyowati, E. (2018). Aplikasi Penentuan Harga Pokok Produksi Batik Madura dengan Metode Activity Based Costing dan Analisis Regresi Linier. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 16(1), 48. <https://doi.org/10.12962/j24068535.v16i1.a690>
- Puspita Dewi, Y., & Indra Sensuse, D. (2013). Pengembangan Data Warehouse dengan Konsep Slowly Changing Dimension untuk Menunjang Pengambilan Keputusan dalam Bidang Distribusi BBM: Studi Kasus PT. Sinar Pedoman Abadi (SPA). *Jurnal Telematika Mkom*, 5.
- Rahmawati, D., Kristanto, T., Setya Pratama, B. F., & Abiansa, D. B. (2022). Prediksi Pelaku Perjalanan Luar Negeri Di Masa Pandemi COVID-19 Menggunakan Metode Regresi Linier Sederhana. *Journal of Information System Research (JOSH)*, 3(3), 338–343. <https://doi.org/10.47065/josh.v3i3.1507>

- Randles, B. M., Pasquetto, I. V., Golshan, M. S., & Borgman, C. L. (2017). Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 1–2. Toronto, ON, Canada: IEEE. <https://doi.org/10.1109/JCDL.2017.7991618>
- Rinaldo Fernandes, A. A., & Solimun. (2022). *Metode Analisis Data Penelitian: Pendekatan Regresi*. UB Press.
- Saputra, R. (2020). Pengaruh Harga dan Kualitas Produk Terhadap Keputusan Pembelian pada Distro Label Store Pekanbaru (Skripsi, Universitas Islam Riau). <https://repository.uir.ac.id/6149/1/RIDHO%20SAPUTRA.pdf>
- Sidik, A. (2018). Penggunaan System Usability Scale (SUS) Sebagai Evaluasi Website Berita Mobile. *Technologia: Jurnal Ilmiah*, 9(2), 83. <https://doi.org/10.31602/tji.v9i2.1371>
- Siswanto, I. A., & Asmunin. (2016). Aplikasi Visualisasi Data Mahasiswa dan Dosen dengan Memanfaatkan Highcharts. *Jurnal Manajemen Informatika*, 5, 93–98.
- Sri Mulyani, E. D., Mulady, F., Ramadhan, D., Ariyantono, A., Ramdani, D., Wahyundana, R., & Gilang, M. (2020). Estimasi Harga Jual Mobil Bekas Menggunakan Metode Regresi Linier Berganda. *e-Jurnal JUSITI (Jurnal Sistem Informasi dan Teknologi Informasi)*, 9, 1–8. <https://doi.org/10.36774/jusiti.v9i1.649>
- Srinath, K. R. (2017). Python – The Fastest Growing Programming Language. *International Research Journal of Engineering and Technology (IRJET)*, 04(12).
- Susiana. (2015). Analisis Peramalan Penjualan Minyak Kelapa Sawit (MKS) pada PT. Perkebunan Nusantara IV Unit Kebun Pabatu. (*Kumpulan Artikel Ilmiah, Informatika, Statistik, Matematika dan Aplikasi*, 1, 12).
- Syafruddin, M., Hakim, L., & Despa, D. (2014). Metode Regresi Linier untuk Prediksi Kebutuhan Energi Listrik Jangka Panjang (Studi Kasus Provinsi Lampung) *Jurnal Informatika dan Teknik Elektro Terapan*, 2(2). <https://doi.org/10.23960/jitet.v2i2.237>

- Tan, W. (2020). Analisis Faktor-Faktor yang Mempengaruhi Keputusan Pembelian Mobil Bekas pada PT Ong Citra Nusa (Skripsi, Universitas Putera Batam). <http://repository.upbatam.ac.id/2230/1/cover%20s.d%20bab%20III.pdf>
- Tatachar, A. V. (2021). Comparative Assessment of Regression Models Based On Model Evaluation Metrics. *International Research Journal of Engineering and Technology (IRJET)*, 08(09).
- Tindriyani, N. A. (2017). Implementasi Neural Network pada MATLAB untuk Peramalan Konsumsi Beban Listrik di Kabupaten Ponorogo, Jawa Timur. *Jurnal Teknik Elektro*, 7-12.
- Tjung, J., Hugeng, & Sutrisno, T. (2023). Dashboard Tingkat Kepuasan dan Penjualan CV. Plus Lestari Rakyat. *Jurnal Ilmu Komputer dan Sistem Informasi*, 11(1). <https://doi.org/10.24912/jiksi.v11i1.24135>
- Umam, Q., Agung Wicaksono, S., & Purnomo, W. (2019). Analisis Dan Perancangan Data Warehouse Menggunakan Pendekatan (Studi Pada Dinas Komunikasi Dan Informatika Kabupaten Sidoarjo) | PDF. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3, 1824–1833.
- Utama, R. (2019). *Buku Manajemen Operasi Full*. Um Jakarta Press
- Wandi, S., & Abaharis, H. (2017). Pengaruh Kualitas Produk, Promosi, dan Harga Terhadap Keputusan Pembelian Mobil Bekas Merek Avanza di Kota Padang. *OSF (Open Science Framework)*
- Wijoyo, R. H. S. (2021). Exploratory Data Analysis (EDA) Tinggi Muka Air di Jakarta (Skripsi, Universitas Teknologi Digital Indonesia Yogyakarta). https://eprints.utdi.ac.id/9464/1/1_145410006_HALAMAN_DEPAN.pdf
- Zikra, A. A., & Idris, M. (2022). Implementasi Business Intelligence pada ACCAbsensi Menggunakan Aplikasi Power BI. *Automata*, 3.

LAMPIRAN

Lampiran tidak perlu diberi nomor halaman. Dokumen apa saja yang dimasukkan dalam lampiran cukup diberi judul dengan kata ‘LAMPIRAN’ yang dilanjutkan dengan huruf abjad besar untuk penomoran. Cukup judul ‘LAMPIRAN’ saja yang dimasukkan dalam daftar isi. Judul-judul lampiran, seperti Lampiran A, Lampiran B dan seterusnya, tidak perlu dimasukkan dalam daftar isi.