

BAB 3

METODOLOGI PENELITIAN

Penelitian ini menggunakan metode penelitian kuantitatif dengan pendekatan eksperimen. Pendekatan kuantitatif dipilih karena penelitian ini mengolah data numerik yang dianalisis menggunakan metode matematis dan statistik untuk membangun serta mengevaluasi model prediksi harga kendaraan listrik. Pendekatan eksperimen dilakukan dengan membangun, melatih, dan mengevaluasi model machine learning berbasis regresi linear menggunakan dataset kendaraan listrik yang merepresentasikan pasar Eropa.

Eksperimen bertujuan untuk menganalisis hubungan antara spesifikasi teknis kendaraan listrik dan harga pasar kendaraan listrik bekas. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari dataset publik kendaraan listrik yang tersedia pada platform Kaggle. Dataset tersebut berisi informasi spesifikasi teknis kendaraan listrik, karakteristik performa, serta harga kendaraan dalam satuan mata uang Euro. Penggunaan data sekunder dipilih karena dataset telah melalui proses pengumpulan dan dokumentasi secara sistematis, serta relevan dengan tujuan penelitian.

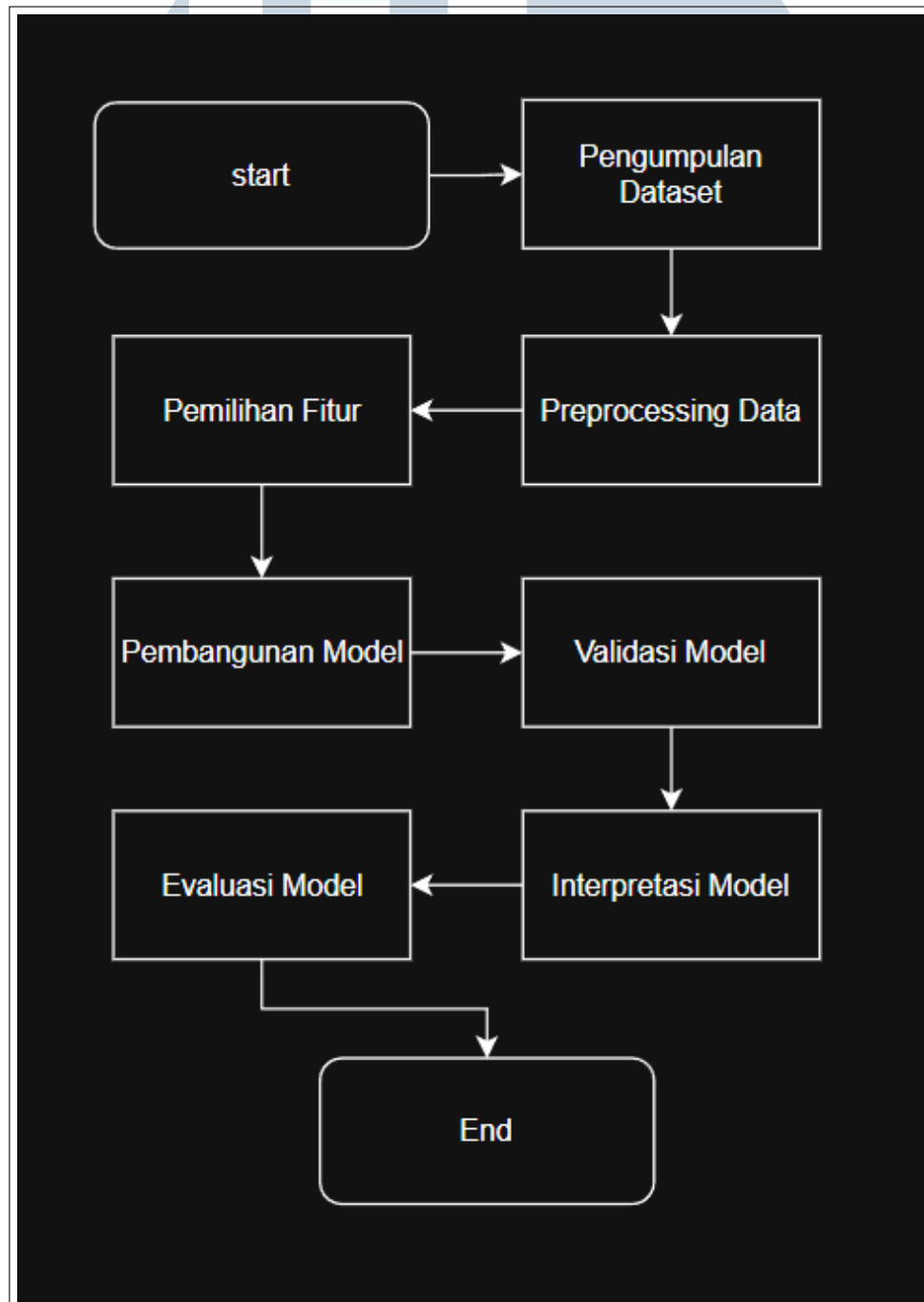
Proses pengolahan data dalam penelitian ini dilakukan melalui beberapa tahapan, yaitu:

- Pembersihan data (*data cleaning*) untuk menangani nilai kosong dan inkonsistensi data,
- Transformasi dan encoding fitur kategorikal menggunakan metode one-hot encoding,
- Normalisasi data numerik menggunakan standardisasi untuk menyamakan skala data,
- Pemilihan fitur yang relevan berdasarkan karakteristik teknis kendaraan dan tujuan prediksi harga.

Metode pemodelan yang digunakan dalam penelitian ini adalah regresi linear berbasis machine learning dengan pendekatan Ordinary Least Squares (OLS). Regresi linear dipilih karena memiliki interpretabilitas yang baik serta mampu menunjukkan hubungan linier antara variabel independen berupa spesifikasi teknis kendaraan dan variabel dependen berupa harga kendaraan listrik. Model dilatih

menggunakan data latih (*training set*) dan dievaluasi menggunakan data uji (*testing set*) untuk mengukur performa prediksi secara objektif.

3.1 Pipeline Penelitian



Gambar 3.1. Tampilan Pipeline Penelitian

Pipeline penelitian pada gambar 3.1 dirancang untuk memvisualisasikan alur penelitian secara sistematis, mulai dari pengumpulan data hingga interpretasi hasil prediksi harga kendaraan listrik. Setiap tahapan di dalam pipeline memiliki peran penting untuk memastikan kualitas data, kestabilan model, dan keakuratan prediksi. Tahapan-tahapan tersebut dijelaskan sebagai berikut:

1. **Pengumpulan Dataset**

Tahap awal penelitian dimulai dengan pengumpulan dataset mobil listrik dari sumber terpercaya. Dataset ini memuat informasi teknis kendaraan serta harga dalam satuan Euro, yang menjadi variabel target penelitian.

2. **Pra-pemrosesan Data (Data Preprocessing)**

Dataset yang diperoleh masih mengandung nilai tidak valid, perbedaan format data, serta variasi skala antar fitur. Oleh karena itu, dilakukan pra-pemrosesan meliputi:

- Pembersihan data untuk menangani nilai hilang dan format tidak valid.
- Transformasi variabel, termasuk encoding variabel kategorikal menjadi numerik.
- Penanganan outlier menggunakan metode Z-score untuk menjaga distribusi data tetap representatif.
- Standarisasi fitur numerik agar setiap variabel memiliki kontribusi proporsional terhadap model.

3. **Pemilihan Fitur (Feature Selection)**

Fitur-fitur numerik yang memiliki hubungan langsung dengan harga kendaraan dipilih sebagai variabel independen. Pemilihan fitur dilakukan dengan mempertimbangkan analisis korelasi antar variabel dan pengujian multikolinearitas menggunakan Variance Inflation Factor (VIF). Fitur yang dipilih mencakup akselerasi, kecepatan maksimum, jarak tempuh, efisiensi energi, kecepatan pengisian daya, dan jumlah kursi.

4. **Pembangunan Model Multiple Linear Regression (MLR)**

Setelah data siap, tahap pembangunan model dilakukan menggunakan algoritma Multiple Linear Regression. Data dibagi menjadi subset latih dan uji untuk menguji kemampuan generalisasi model terhadap data baru. Model dilatih untuk mempelajari hubungan linear antara spesifikasi teknis kendaraan listrik dengan harga.

5. Validasi Asumsi MLR

Asumsi-asumsi penting dalam regresi linear diuji untuk memastikan model dapat dipercaya:

- **Linearity:** Hubungan antara variabel independen dan harga diasumsikan linier, divalidasi melalui scatter plot dan grafik residual.
- **Homoscedasticity:** Variansi residual diasumsikan konstan, diperiksa menggunakan grafik residual terhadap prediksi.
- **Normality of Residuals:** Distribusi residual diperiksa menggunakan histogram dan Q-Q plot.
- **Independence of Errors:** Residual diuji menggunakan Durbin-Watson untuk mendeteksi autokorelasi.

6. Evaluasi Model

Performa model diukur menggunakan metrik: MAE, MSE, RMSE, dan MAPE. Hasil evaluasi ini digunakan untuk menilai akurasi prediksi model serta memahami kesalahan prediksi yang terjadi. Dan confidence interval (CI) yang memungkinkan peneliti mengetahui seberapa stabil prediksi model ketika diterapkan pada data baru yang belum pernah dilihat,

7. Interpretasi Hasil Model

Koefisien regresi diinterpretasikan untuk mengetahui pengaruh masing-masing fitur terhadap harga. Intersep model juga dianalisis sebagai titik awal persamaan regresi. Hasil evaluasi dan interpretasi ini menjadi dasar untuk melakukan prediksi harga kendaraan listrik baru.

Dengan pipeline ini, alur penelitian dari pengolahan data hingga estimasi harga dapat dipahami secara menyeluruh. Setiap tahap saling berhubungan dan membentuk rangkaian proses yang terstruktur, memastikan prediksi harga kendaraan listrik dilakukan secara akurat dan dapat dipertanggungjawabkan.

3.2 Variabel Penelitian

Data yang digunakan dalam penelitian ini bersumber dari dataset kendaraan listrik (*Electric Vehicles*) yang tersedia pada platform Kaggle. Dataset tersebut memuat informasi mengenai berbagai atribut teknis dan spesifikasi kendaraan listrik yang dipasarkan di wilayah Eropa, termasuk kapasitas baterai, performa kendaraan,

efisiensi energi, serta harga kendaraan dalam satuan Euro yang digunakan sebagai variabel target.

Tabel 3.1. Deskripsi variabel penelitian

| Variabel | Tipe Data | Deskripsi |
|-----------------|-----------|--|
| AccelSec | Numerik | Waktu akselerasi kendaraan listrik dari 0 hingga 100 km/jam (dalam detik). |
| TopSpeed_KmH | Numerik | Kecepatan maksimum kendaraan listrik (km/jam). |
| Range_Km | Numerik | Jarak tempuh maksimum kendaraan listrik dalam satu kali pengisian daya (km). |
| Efficiency_WhKm | Numerik | Konsumsi energi kendaraan listrik dalam watt-hour per kilometer (Wh/km). |
| FastCharge_KmH | Numerik | Kecepatan pengisian daya cepat kendaraan listrik (km/jam). |
| Seats | Numerik | Jumlah kursi yang tersedia pada kendaraan listrik. |
| Price_EUR | Numerik | Harga kendaraan listrik yang telah dikonversi ke dalam Rupiah (IDR). |

Fitur-fitur yang digunakan dalam penelitian ini dipilih berdasarkan relevansinya terhadap tujuan penelitian, yaitu membangun model regresi linear untuk merepresentasikan hubungan antara spesifikasi teknis kendaraan dan harga mobil listrik bekas. Oleh karena itu, penelitian ini memfokuskan pada fitur-fitur numerik yang secara langsung merepresentasikan performa dan karakteristik teknis kendaraan. Beberapa fitur kategorikal yang tersedia dalam dataset, seperti *PowerTrain*, *PlugType*, *BodyStyle*, dan *Segment*, tidak digunakan dalam pemodelan. Hal ini disebabkan oleh sifat fitur-fitur tersebut yang bersifat nominal dan tidak memiliki hubungan linier secara langsung dengan harga kendaraan. Pada regresi linear, penggunaan fitur kategorikal berpotensi menurunkan interpretabilitas model serta tidak sejalan dengan asumsi hubungan linier antara variabel independen dan variabel dependen.

Selain itu, pengaruh fitur kategorikal tersebut terhadap harga kendaraan listrik umumnya bersifat tidak langsung dan telah terwakili oleh fitur numerik lain yang digunakan dalam penelitian ini, seperti kapasitas baterai, jarak

tempuh maksimum, efisiensi energi, dan performa kendaraan. Oleh karena itu, untuk menjaga kesederhanaan dan interpretabilitas model, fitur-fitur kategorikal tersebut tidak dimasukkan dalam proses pemodelan regresi linear. Penelitian ini menggunakan satu variabel dependen dan beberapa variabel independen.

Variabel dependen dalam penelitian ini adalah harga mobil listrik dalam satuan Euro, yang merepresentasikan harga pasar kendaraan listrik di wilayah Eropa dan digunakan sebagai target prediksi. Variabel independen yang digunakan merupakan atribut teknis kendaraan, meliputi akselerasi kendaraan (*AccelSec*), kecepatan maksimum (*TopSpeed_KmH*), jarak tempuh maksimum dalam sekali pengisian daya (*Range_Km*), efisiensi energi kendaraan (*Efficiency_WhKm*), kecepatan pengisian daya cepat (*FastCharge_KmH*), kemampuan pengisian cepat (*RapidCharge*), serta jumlah kursi (*Seats*).

Dataset yang digunakan dalam penelitian ini tidak menyediakan informasi mengenai tahun produksi atau usia kendaraan secara eksplisit. Oleh karena itu, variabel usia kendaraan tidak digunakan dalam pemodelan dan diasumsikan konstan sesuai dengan keterbatasan data yang tersedia.

3.3 Pra-pemrosesan Data (Data Preprocessing)

Tahap pra-pemrosesan data merupakan langkah krusial dalam penelitian ini karena kualitas data secara langsung memengaruhi performa dan stabilitas model regresi linear yang dibangun. Dataset kendaraan listrik yang digunakan masih mengandung nilai tidak valid, perbedaan format data, serta variasi skala antar fitur, sehingga perlu dilakukan serangkaian proses pra-pemrosesan agar data siap digunakan dalam pemodelan.

Seluruh tahapan pra-pemrosesan yang dilakukan pada penelitian ini disesuaikan dengan karakteristik dataset *ElectricCarData_Clean* dan diimplementasikan secara langsung menggunakan bahasa pemrograman Python, sebagaimana ditunjukkan pada Bab IV. Tahapan pra-pemrosesan meliputi pembersihan data, transformasi variabel, penanganan outlier, serta standarisasi fitur numerik.

3.3.1 Pembersihan dan Transformasi Data

Pada tahap awal, dilakukan pembersihan data untuk menangani nilai tidak valid dan memastikan seluruh fitur yang digunakan memiliki format numerik.

Salah satu fitur yang memerlukan perhatian khusus adalah *FastCharge_KmH*, yang pada dataset awal mengandung simbol non-numerik berupa tanda strip (-) untuk menandai data yang tidak tersedia. Nilai tersebut dikonversi menjadi nilai kosong (*missing value*) dan selanjutnya ditangani menggunakan metode imputasi median.

Pemilihan nilai median sebagai metode imputasi dilakukan karena median lebih robust terhadap pengaruh nilai ekstrem dibandingkan nilai rata-rata, sehingga dapat menjaga distribusi data tetap representatif. Proses ini bertujuan agar fitur *FastCharge_KmH* tetap dapat digunakan dalam pemodelan tanpa mengurangi jumlah data secara signifikan.

Selain itu, variabel *RapidCharge* yang semula bersifat kategorikal dengan nilai *Yes* dan *No* dikonversi ke dalam bentuk numerik biner, yaitu 1 untuk *Yes* dan 0 untuk *No*. Transformasi ini dilakukan agar variabel tersebut dapat diproses oleh algoritma regresi linear tanpa mengubah makna informasinya.

3.3.2 Pemilihan Fitur Numerik

Setelah proses pembersihan dan transformasi data, dilakukan pemilihan fitur numerik yang akan digunakan sebagai variabel independen dalam pemodelan regresi linear. Pemilihan fitur difokuskan pada atribut teknis kendaraan listrik yang memiliki hubungan langsung dengan harga kendaraan dan dapat direpresentasikan secara numerik.

Berdasarkan analisis awal terhadap struktur dataset dan kesesuaian dengan tujuan penelitian, fitur-fitur yang digunakan meliputi waktu akselerasi kendaraan (*AccelSec*), kecepatan maksimum (*TopSpeed_KmH*), jarak tempuh maksimum (*Range_Km*), efisiensi energi (*Efficiency_WhKm*), kecepatan pengisian daya (*FastCharge_KmH*), serta jumlah kursi (*Seats*). Fitur-fitur ini dipilih karena merepresentasikan aspek performa, efisiensi, dan utilitas kendaraan yang relevan terhadap pembentukan harga pasar kendaraan listrik.

3.3.3 Penanganan Outlier

Untuk menjaga kestabilan model regresi linear, dilakukan penanganan terhadap data outlier yang berpotensi mendistorsi hasil pemodelan. Deteksi outlier dilakukan menggunakan metode *Z-score*, di mana nilai *Z-score* dihitung untuk setiap fitur numerik yang digunakan dalam model.

Data dengan nilai *Z-score* lebih besar dari tiga dianggap sebagai outlier

ekstrem dan dikeluarkan dari dataset. Ambang batas ini dipilih karena secara statistik mencerminkan data yang berada jauh dari distribusi utama. Penghapusan outlier bertujuan untuk menjaga distribusi data tetap representatif serta mengurangi pengaruh nilai ekstrem terhadap estimasi koefisien regresi.

3.3.4 Standarisasi Fitur

Tahap akhir pra-pemrosesan adalah standarisasi fitur numerik menggunakan metode *StandardScaler*. Standarisasi dilakukan karena fitur-fitur yang digunakan memiliki satuan dan rentang nilai yang berbeda-beda, seperti detik, kilometer per jam, dan watt-hour per kilometer.

Melalui proses standarisasi, setiap fitur ditransformasikan sehingga memiliki nilai rata-rata nol dan simpangan baku satu. Proses ini bertujuan untuk mencegah dominasi fitur dengan skala besar terhadap proses pelatihan model serta memastikan bahwa seluruh fitur berkontribusi secara proporsional dalam pembentukan model regresi linear.

Hasil dari tahap pra-pemrosesan ini adalah dataset yang telah bersih, terstandarisasi, dan siap digunakan dalam tahap pemodelan regresi linear sebagaimana dijelaskan pada bagian selanjutnya.

3.4 Pemilihan Fitur (Feature Selection)

Pemilihan fitur dilakukan untuk memastikan bahwa variabel independen yang digunakan dalam model regresi linear memiliki relevansi terhadap variabel target serta tidak menimbulkan permasalahan multikolinearitas. Tahap ini bertujuan untuk meningkatkan stabilitas model, menjaga interpretabilitas koefisien regresi, serta menghindari redundansi informasi antar fitur.

Pada penelitian ini, proses pemilihan fitur dilakukan secara bertahap dengan mempertimbangkan karakteristik dataset kendaraan listrik dan tujuan utama penelitian, yaitu memprediksi harga kendaraan listrik berdasarkan spesifikasi teknis kendaraan. Oleh karena itu, hanya fitur numerik yang secara langsung merepresentasikan performa dan karakteristik kendaraan yang dipertimbangkan dalam pemodelan.

3.4.1 Analisis Korelasi

Tahap awal pemilihan fitur dilakukan dengan menganalisis hubungan antara masing-masing variabel independen dan variabel dependen harga kendaraan listrik. Analisis ini bertujuan untuk mengidentifikasi fitur-fitur yang memiliki kontribusi terhadap perubahan harga kendaraan.

Fitur dengan tingkat korelasi yang sangat rendah terhadap harga dipertimbangkan untuk tidak digunakan karena kontribusinya yang minimal terhadap kemampuan prediksi model. Sebaliknya, fitur dengan korelasi yang lebih tinggi dipertahankan karena berpotensi memberikan informasi yang relevan dalam pembentukan model regresi linear.

3.4.2 Pengujian Multikolinearitas menggunakan Variance Inflation Factor (VIF)

Setelah analisis korelasi, dilakukan pengujian multikolinearitas antar variabel independen menggunakan Variance Inflation Factor (VIF). Multikolinearitas dapat menyebabkan ketidakstabilan estimasi koefisien regresi serta menurunkan kejelasan interpretasi model.

Dalam penelitian ini, digunakan ambang batas VIF sebesar lima. Variabel independen dengan nilai VIF lebih besar dari lima dianggap memiliki korelasi tinggi dengan variabel lain dan berpotensi dikeluarkan dari model. Hasil pengujian menunjukkan bahwa seluruh fitur yang digunakan memiliki nilai VIF di bawah ambang batas tersebut, sehingga tidak terdapat indikasi multikolinearitas yang signifikan.

3.4.3 Fitur Akhir yang Digunakan

Berdasarkan hasil analisis korelasi dan pengujian multikolinearitas, diperoleh himpunan fitur akhir yang digunakan dalam pemodelan regresi linear. Fitur-fitur tersebut merepresentasikan aspek performa, efisiensi, dan utilitas kendaraan listrik yang relevan terhadap pembentukan harga pasar.

Fitur akhir yang digunakan dalam penelitian ini adalah:

- AccelSec
- TopSpeed.KmH

- Range_Km
- Efficiency_WhKm
- FastCharge_KmH
- Seats

Pemilihan fitur ini memastikan bahwa model regresi linear dibangun menggunakan variabel yang relevan, tidak saling bergantung secara berlebihan, serta tetap memiliki interpretabilitas yang baik dalam konteks analisis harga kendaraan listrik.

3.5 Pembangunan Model Multiple Linear Regression

Setelah melalui tahap pra-pemrosesan data dan pemilihan fitur, langkah selanjutnya adalah membangun model prediksi harga kendaraan listrik menggunakan metode Multiple Linear Regression (MLR). Model ini digunakan untuk merepresentasikan hubungan linear antara beberapa variabel independen berupa spesifikasi teknis kendaraan dan satu variabel dependen berupa harga kendaraan listrik.

Metode regresi linear dipilih karena memiliki tingkat interpretabilitas yang tinggi dan mampu menjelaskan pengaruh masing-masing variabel independen terhadap variabel target secara kuantitatif. Selain itu, regresi linear memungkinkan analisis koefisien regresi untuk memahami kontribusi relatif setiap fitur dalam pembentukan harga kendaraan listrik.

3.5.1 Pembagian Data Latih dan Data Uji

Sebelum proses pelatihan model dilakukan, dataset dibagi menjadi dua bagian, yaitu data latih (*training set*) dan data uji (*testing set*). Pembagian data dilakukan dengan proporsi 80% untuk data latih dan 20% untuk data uji.

Data latih digunakan untuk membangun dan melatih model regresi linear, sedangkan data uji digunakan untuk mengukur kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah digunakan dalam proses pelatihan. Pembagian data dilakukan secara acak dengan nilai *random state* tertentu agar hasil penelitian dapat direproduksi.

3.5.2 Pelatihan Model Regresi Linear

Model Multiple Linear Regression dilatih menggunakan data latih yang telah melalui proses standarisasi. Standarisasi fitur dilakukan terlebih dahulu untuk memastikan bahwa perbedaan skala antar variabel tidak memengaruhi proses estimasi koefisien regresi.

Proses pelatihan bertujuan untuk memperoleh parameter model berupa koefisien regresi dan intersep yang meminimalkan selisih antara nilai harga aktual dan nilai harga prediksi. Estimasi parameter dilakukan menggunakan pendekatan Ordinary Least Squares (OLS), yang mencari nilai koefisien regresi dengan meminimalkan jumlah kuadrat galat prediksi.

Hasil dari tahap ini adalah sebuah model regresi linear yang telah mempelajari pola hubungan antara spesifikasi teknis kendaraan listrik dan harga kendaraan berdasarkan data historis yang tersedia.

3.5.3 Model Regresi Linear dalam Bentuk Matematis

Secara matematis, model Multiple Linear Regression yang dibangun dalam penelitian ini dapat dirumuskan sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (3.1)$$

dengan keterangan:

- Y adalah harga kendaraan listrik dalam satuan Rupiah,
- X_1, X_2, \dots, X_n adalah variabel independen berupa spesifikasi teknis kendaraan,
- β_0 adalah intersep model,
- $\beta_1, \beta_2, \dots, \beta_n$ adalah koefisien regresi yang menunjukkan besarnya pengaruh masing-masing variabel independen,
- ε adalah error atau selisih antara nilai prediksi dan nilai aktual.

Model regresi linear yang telah dibangun selanjutnya digunakan untuk melakukan prediksi harga kendaraan listrik serta dianalisis lebih lanjut pada bab berikutnya untuk mengevaluasi performa dan menginterpretasikan hasil prediksi yang diperoleh.

3.6 Validasi Model Multiple Linear Regression

Model Multiple Linear Regression (MLR) dibangun dengan memperhatikan asumsi-asumsi dasar regresi linear agar hasil prediksi dapat diinterpretasikan secara valid. Validasi asumsi dilakukan sebagai berikut:

- **Linearity** Hubungan antara variabel independen dan harga diasumsikan linier. Asumsi ini divalidasi menggunakan scatter plot dan grafik residual yang menunjukkan pola penyebaran acak.
- **Homoscedasticity** Variansi residual diasumsikan konstan. Hal ini diperiksa menggunakan grafik residual terhadap nilai prediksi. Tidak adanya pola tertentu menunjukkan bahwa asumsi homoscedasticity terpenuhi.
- **Normality of Residuals** Distribusi residual diperiksa menggunakan histogram residual dan Q-Q plot. Residual yang mengikuti distribusi mendekati normal menunjukkan terpenuhinya asumsi normalitas.
- **Independence of Errors** Kemandirian residual diuji menggunakan Durbin-Watson test. Nilai Durbin-Watson yang mendekati 2 mengindikasikan tidak adanya autokorelasi residual.
- **No Multicollinearity** Multikolinearitas diuji menggunakan VIF. Seluruh fitur yang digunakan memiliki nilai VIF di bawah ambang batas yang ditetapkan.

3.7 Evaluasi Model

Evaluasi model dilakukan untuk mengukur kemampuan model Multiple Linear Regression (MLR) dalam memprediksi harga kendaraan listrik berdasarkan spesifikasi teknis yang digunakan sebagai variabel independen. Tahapan evaluasi pada penelitian ini berfokus pada pengukuran selisih antara nilai prediksi dan nilai aktual harga kendaraan listrik pada data uji (*testing set*).

Proses evaluasi dilakukan setelah model selesai dilatih menggunakan data latih. Model yang telah terbentuk kemudian digunakan untuk menghasilkan nilai prediksi harga kendaraan listrik pada data uji, yang selanjutnya dibandingkan dengan nilai harga aktual. Perbandingan ini bertujuan untuk menilai tingkat akurasi dan kestabilan model dalam melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya.

Metrik evaluasi yang digunakan dalam penelitian ini mengacu pada metrik yang telah dijelaskan pada Bab II, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Mean Absolute Percentage Error (MAPE). Keempat metrik tersebut dipilih karena mampu merepresentasikan besar kesalahan prediksi secara absolut, kuadrat, akar kuadrat, serta relatif dalam bentuk persentase.

MAE digunakan untuk menggambarkan rata-rata kesalahan prediksi dalam satuan harga kendaraan listrik sehingga mudah diinterpretasikan secara langsung. MSE dan RMSE digunakan untuk memberikan penalti lebih besar terhadap kesalahan prediksi yang ekstrem, yang penting dalam konteks harga kendaraan dengan rentang nilai yang luas. Sementara itu, MAPE digunakan untuk memberikan gambaran tingkat kesalahan relatif model dalam bentuk persentase, sehingga memudahkan evaluasi performa model secara proporsional terhadap nilai aktual harga kendaraan.

Koefisien determinasi (R^2) tidak digunakan sebagai metrik evaluasi utama dalam penelitian ini karena tidak merepresentasikan besarnya kesalahan prediksi, melainkan hanya menunjukkan proporsi variasi variabel dependen yang dapat dijelaskan oleh variabel independen. Oleh karena itu, evaluasi model difokuskan pada metrik berbasis error yang lebih relevan dengan tujuan prediksi harga.

Hasil perhitungan metrik evaluasi ini selanjutnya digunakan sebagai dasar analisis pada Bab IV untuk menilai kinerja model, membandingkan kontribusi fitur yang digunakan, serta membahas tingkat keandalan model dalam memprediksi harga kendaraan listrik berdasarkan spesifikasi teknisnya.

Selain menggunakan metrik evaluasi berbasis error seperti MAE, MSE, RMSE, dan MAPE, penelitian ini juga menilai tingkat kepercayaan prediksi model melalui *confidence interval* (CI). CI memungkinkan peneliti mengetahui seberapa stabil prediksi model ketika diterapkan pada data baru yang belum pernah dilihat, serta memberikan rentang estimasi di sekitar prediksi di mana harga aktual kemungkinan besar berada dengan tingkat keyakinan tertentu, misalnya 95%.

Dengan mengombinasikan evaluasi berbasis error dan confidence interval, penelitian ini tidak hanya menilai akurasi prediksi model Multiple Linear Regression, tetapi juga tingkat keandalan prediksi terhadap data baru. Hal ini penting agar model tidak hanya sekadar memperkirakan harga, tetapi memberikan indikasi rentang harga yang realistis sehingga lebih bermanfaat bagi pengguna dan pelaku pasar mobil listrik.

3.8 Tools dan Perangkat yang Digunakan

Penelitian ini menggunakan beberapa tools dan perangkat pendukung sebagai berikut:

- Bahasa pemrograman Python 3.x,
- Jupyter Notebook sebagai lingkungan pengembangan,
- Library Python: Pandas, NumPy, Matplotlib, Seaborn, dan Scikit-learn,
- Overleaf (LaTeX) untuk penyusunan dokumen laporan penelitian,
- Kaggle sebagai sumber dataset kendaraan listrik.

