

KELOMPOK 10

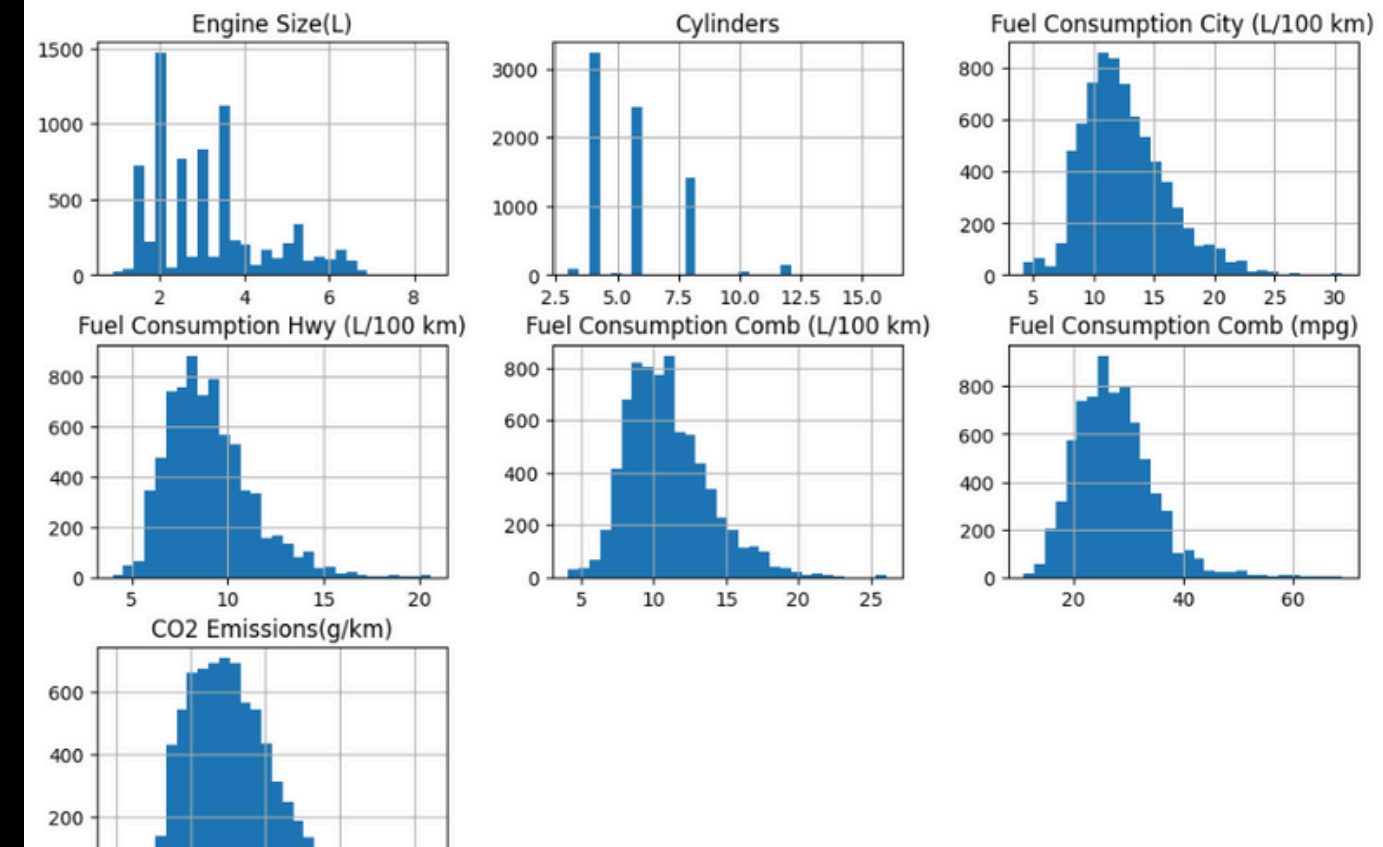
Linear & Polynomial Regression



Deskripsi Dataset

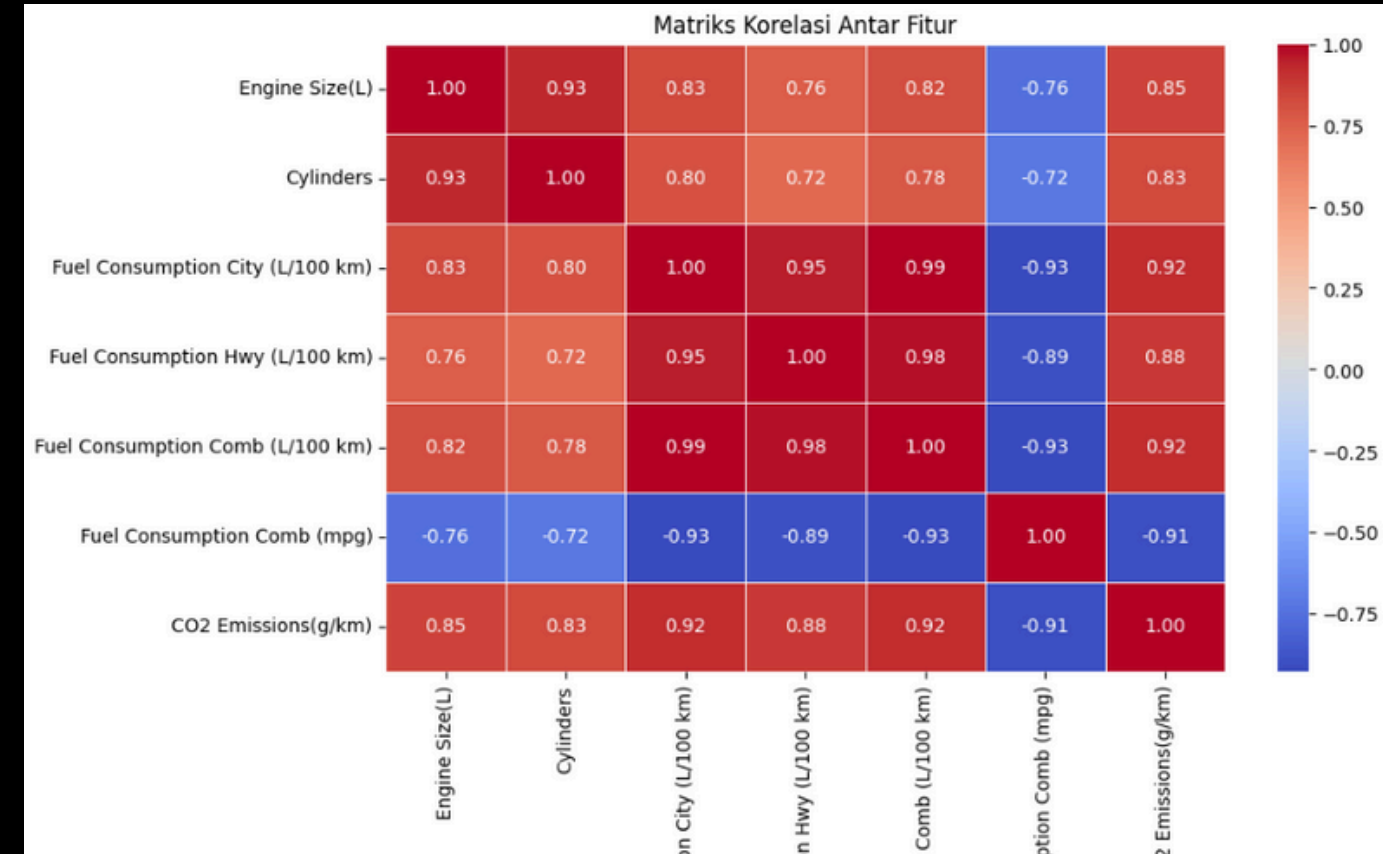
Dataset dengan judul "**CO2 Emissions by Vehicles**", yang berisi informasi tentang emisi karbon dioksida (CO₂) yang dihasilkan oleh kendaraan serta berbagai faktor yang memengaruhinya. Mencakup **7.385** data kendaraan dari pemerintah Kanada selama tujuh tahun. Terdiri dari **12** atribut, termasuk spesifikasi teknis, konsumsi bahan bakar, dan emisi CO₂, dataset ini digunakan untuk menganalisis faktor-faktor yang memengaruhi emisi kendaraan.

Analisis Distribusi Data



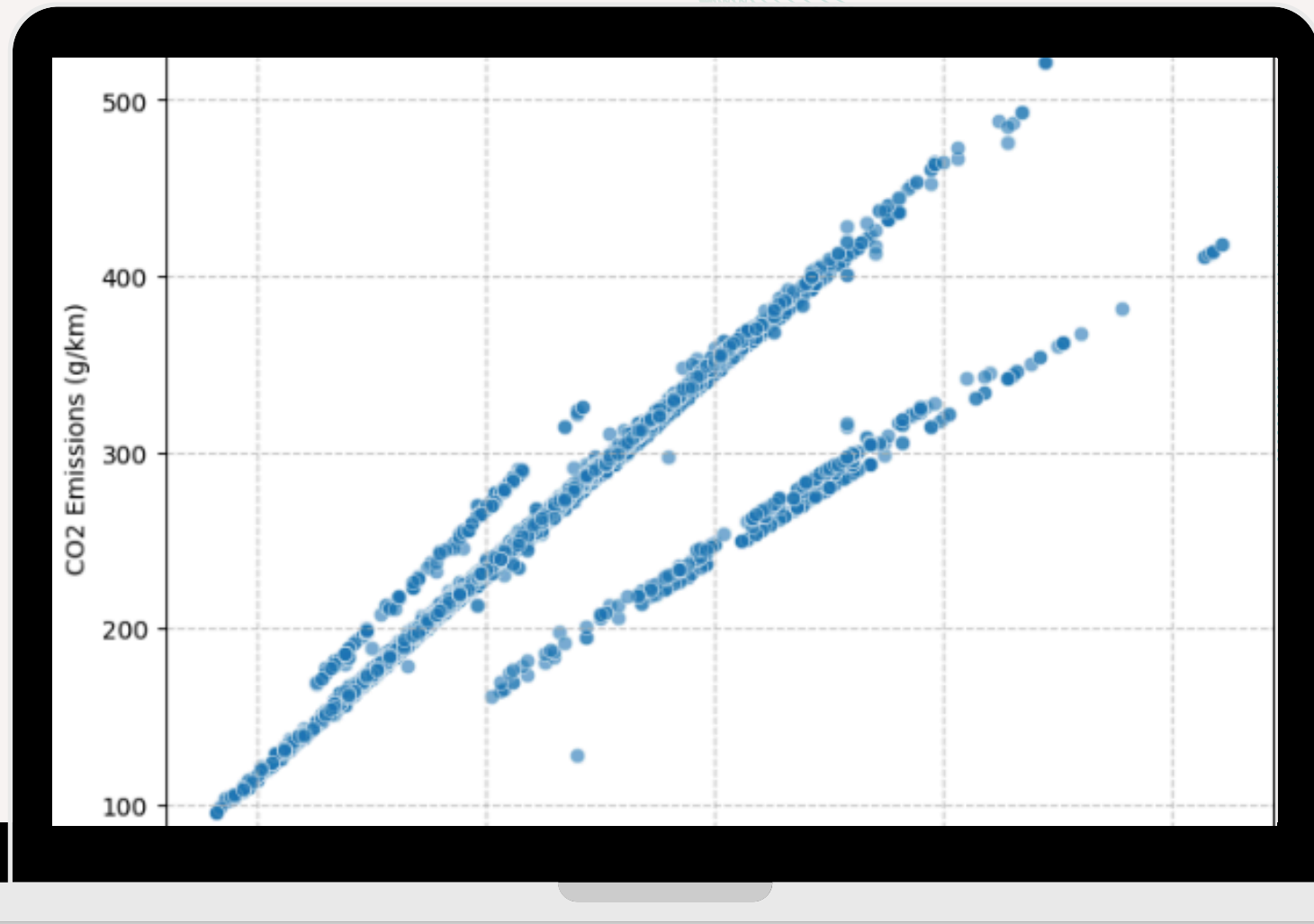
Distribusi data dianalisis untuk melihat bagaimana variabel-variabel utama, seperti konsumsi bahan bakar dan emisi CO₂, tersebar dalam dataset. Dari hasil analisis distribusi, ditemukan bahwa sebagian besar kendaraan memiliki konsumsi bahan bakar kombinasi antara 5 hingga 15 L/100 km dan emisi CO₂ berkisar antara 100 hingga 400 g/km.

Analisis Matriks Korelasi antar fitur



Hasilnya menunjukkan bahwa terdapat korelasi positif yang cukup kuat antara ukuran mesin (Engine Size) dan jumlah silinder (Cylinders) terhadap konsumsi bahan bakar dan emisi CO2. Artinya, semakin besar ukuran mesin dan jumlah silinder, semakin tinggi konsumsi bahan bakar dan emisi CO2 yang dihasilkan.

Scatter Plot Hubungan Fuel Consumption & CO2 Emissions



Dari hasil visualisasi, terlihat bahwa konsumsi bahan bakar yang lebih tinggi berkorelasi dengan peningkatan emisi CO₂. Hal ini sejalan dengan ekspektasi bahwa kendaraan dengan konsumsi bahan bakar lebih boros akan menghasilkan lebih banyak emisi karbon dioksida.

Membangun Model: Linear & Polynomial Regression

01

Linear regression

Regresi Linier digunakan untuk memodelkan hubungan sederhana antara variabel independen, seperti ukuran mesin dan konsumsi bahan bakar, dengan emisi CO₂. Model ini mengasumsikan hubungan linier di mana kenaikan satu unit pada variabel input menghasilkan perubahan tetap pada emisi CO₂.

02

Polynomial Regression

Regresi Polinomial memperluas model linier dengan menambahkan komponen pangkat lebih tinggi untuk menangkap hubungan non-linear yang lebih kompleks. Dalam kasus ini, model polinomial lebih akurat dalam memprediksi emisi CO₂ karena dapat menyesuaikan pola data yang tidak sepenuhnya linier.

01

Linear regression

Model ini mencari hubungan linier antara fitur kendaraan dan emisi CO₂. Dataset dibagi menjadi data latih dan data uji, lalu model diinisialisasi dengan **LinearRegression()** dan dilatih menggunakan metode **fit()**. Setelah pelatihan, bobot koefisien regresi diperoleh untuk prediksi emisi CO₂.

```
LinearRegression ⓘ ⓘ  
LinearRegression()
```

02

Polynomial regression

Model ini menangkap hubungan non-linear antara fitur kendaraan dan emisi CO₂. Fitur awal dikonversi menjadi bentuk polinomial derajat 2 menggunakan **PolynomialFeatures(degree=2)**, lalu model **LinearRegression()** dilatih ulang dengan fitur baru. Evaluasi menggunakan **MSE** dan **R²** menunjukkan bahwa metode ini meningkatkan akurasi prediksi dibandingkan regresi linier.

```
LinearRegression ⓘ ⓘ  
LinearRegression()
```

Perbandingan Evaluasi model Linear & Polynomial Regression

	Metric	Regresi Linear	Regresi Polinomial
0	MSE	377.2062	91.8307
1	R2	0.8903	0.9733
2	MAE	12.9623	5.6609
3	MAPE (%)	5.3619%	2.3281%

MSE

(377.2062 vs. 91.8307)

Regresi Polinomial memiliki kesalahan lebih kecil.

R²

(0.8903 vs. 0.9733)

Regresi Polinomial lebih baik dalam menjelaskan variasi data.

MAE

(12.9623 vs. 5.6609)

Rata-rata kesalahan prediksi lebih rendah pada Regresi Polinomial.

MAPE

(5.3619% vs. 2.3281%)

Regresi Polinomial lebih presisi dibanding Regresi Linier.

Kesimpulannya, Regresi Polinomial memberikan hasil prediksi yang lebih akurat dibandingkan Regresi Linier.

Analisis Hasil: Koefisien Linear & Polynomial

01

Koefisien Linear Regression

- **Intercept (18.93)** → Emisi awal saat semua fitur nol.
- **Fuel Type (10.7371)** → Jenis bahan bakar menambah emisi CO₂.
- **Fuel Consumption Comb (14.9993)** → Setiap kenaikan 1 L/100 km meningkatkan emisi CO₂.
- **Engine Size (10.1408)** → Mesin lebih besar menghasilkan lebih banyak emisi.

	Fitur	Koefisien
0	Intercept	18.932011
1	Fuel Type	10.737114
2	Fuel Consumption Comb (L/100 km)	14.999313
3	Engine Size(L)	10.140795

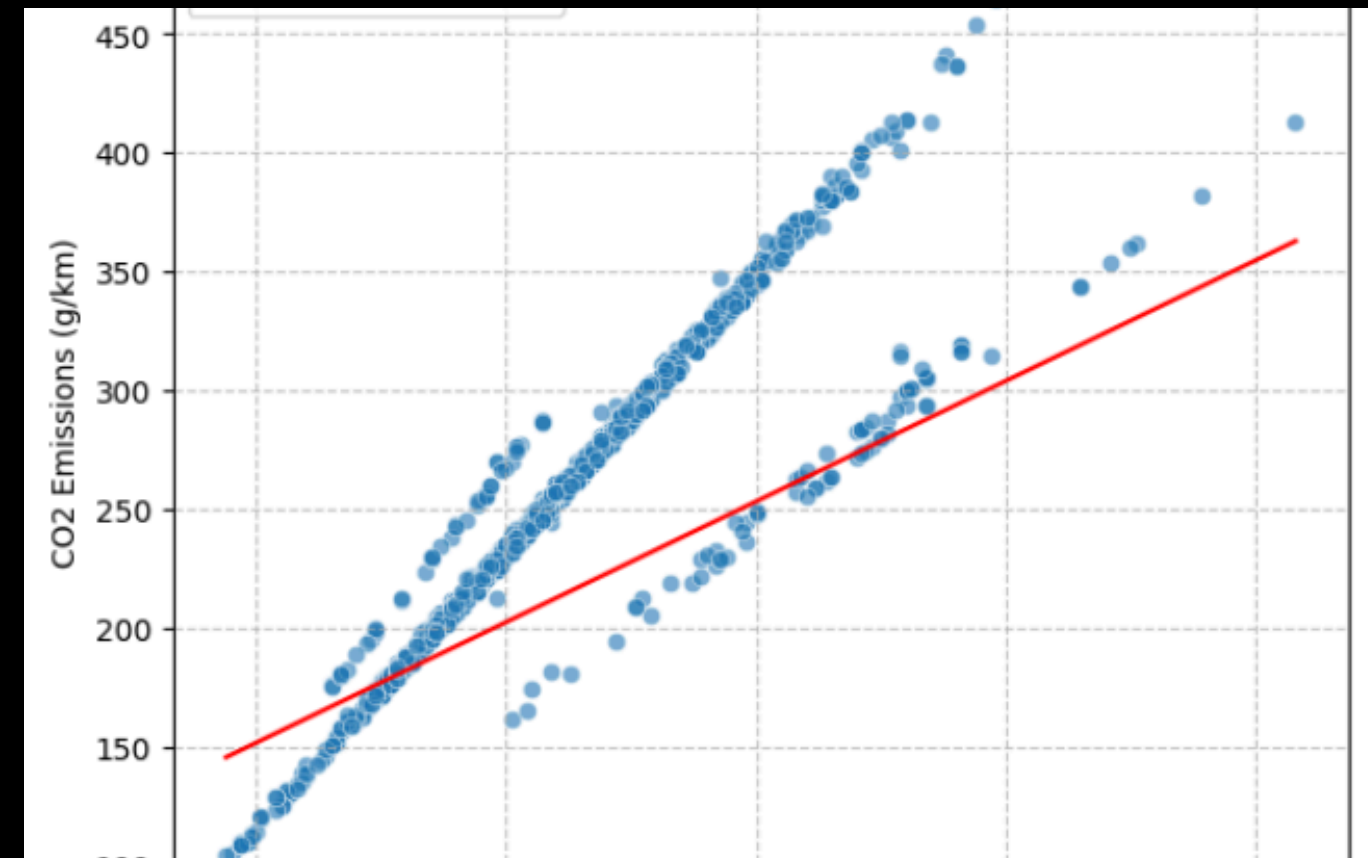
02

Koefisien Polynomial Reagression

- **Intercept (144.43)** → Emisi awal saat semua fitur nol.
- **Fuel Type (-44.46)** → Jenis bahan bakar berpengaruh negatif pada emisi.
- **Fuel Consumption Comb (-0.03)** → Pengaruh kecil secara linear, lebih signifikan dalam interaksi polinomial.
- **Engine Size (35.26)** → Mesin lebih besar meningkatkan emisi CO₂.
- **Koefisien polinomial** menunjukkan interaksi antar fitur, menangkap pola data lebih akurat dibanding regresi linier.

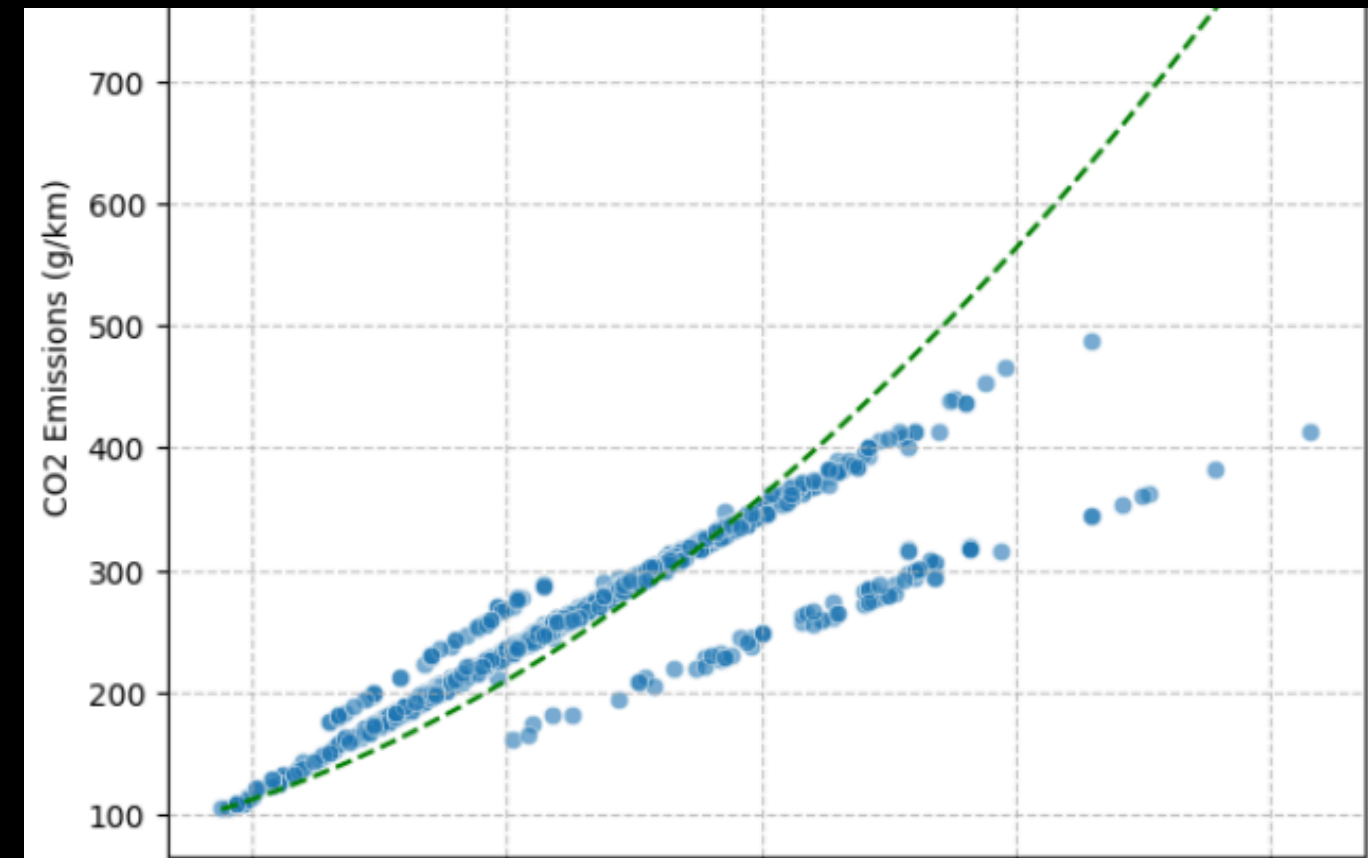
	Fitur	Koefisien
0	Intercept	144.431428
1	1	0.000000
2	Fuel Type	-44.457472
3	Fuel Consumption Comb (L/100 km)	-0.034179
4	Engine Size(L)	35.260627
5	Fuel Type^2	-0.608633
6	Fuel Type Fuel Consumption Comb (L/100 km)	7.244580
7	Fuel Type Engine Size(L)	-8.734772
8	Fuel Consumption Comb (L/100 km)^2	0.043506
9	Fuel Consumption Comb (L/100 km) Engine Size(L)	-1.001662
10	Engine Size(L)^2	1.057230

Scatter Plot: Linear Regression



Grafik ini menunjukkan hubungan antara konsumsi bahan bakar dan emisi CO₂. Titik biru merepresentasikan data aktual, sementara garis merah adalah model regresi linier. Terlihat bahwa model tidak sepenuhnya menangkap pola data karena adanya distribusi non-linear, menunjukkan bahwa hubungan antar variabel tidak sepenuhnya linier.

Scatter Plot: Polynomial Regression



Grafik ini menunjukkan hubungan antara konsumsi bahan bakar dan emisi CO₂ menggunakan regresi polinomial derajat 2. Titik biru merepresentasikan data aktual, sementara garis hijau putus-putus menunjukkan model prediksi. Regresi polinomial menangkap pola non-linear dengan lebih baik, terlihat dari kurva yang melengkung ke atas, menunjukkan bahwa emisi CO₂ meningkat lebih cepat pada konsumsi bahan bakar yang lebih tinggi. Model ini lebih akurat pada konsumsi bahan bakar rendah hingga menengah.