**TUGAS PENGGANTI UTS**

**DATA SAINS**

**ANALISIS TERHADAP VEHICLE CO2 EMSSIONS DATASET**

****

Disusun Oleh:

Muhammad Naufal Afif (1103210089)

Mochamad Arief Dermawan(1103210128)

Mochamad Qussay Alhindi Achmadi (1103213087)

Rangga Khalid (1103210027)

**PROGRAM STUDI S1 TEKNIK KOMPUTER**

**FAKULTAS TEKNIK ELEKTRO**

**UNIVERSITAS TELKOM**

**2025**

1. **Dataset**

* **Nama Dataset: Vehicle CO2 Emssions Dataset**
* **Sumber dataset:** [**https://www.kaggle.com/datasets/brsahan/vehicle-co2-emissions-%20dataset/data**](https://www.kaggle.com/datasets/brsahan/vehicle-co2-emissions-%20dataset/data)

1. **Deskripsi Datset**

Dataset ini memuat spesifikasi teknis dari 7.385 kendaraan yang bertujuan untuk menganalisis dan memprediksi emisi karbon dioksida (CO₂) berdasarkan konsumsi bahan bakar dan karakteristik mesin. Data mencakup atribut seperti merek dan model kendaraan, ukuran mesin, jumlah silinder, jenis transmisi, jenis bahan bakar, serta konsumsi bahan bakar pada kondisi mengemudi di dalam kota, jalan tol, dan kombinasi keduanya. Variabel target utama adalah CO2 Emissions(g/km), sehingga dataset ini sangat cocok digunakan untuk pemodelan regresi dalam rangka memahami dampak spesifikasi kendaraan terhadap emisi gas rumah kaca.

1. **Penjalasan Fittur**

 **Brand**: Merek atau manufaktur kendaraan (misalnya: Toyota, Ford, BMW).

 **Vehicle Class**: Klasifikasi kendaraan berdasarkan ukuran dan penggunaannya (misalnya: SUV, Sedan, Pickup Truck).

 **Engine Size (L)**: Volume kapasitas mesin dalam satuan liter.

 **Cylinders**: Jumlah silinder dalam mesin kendaraan, yang mempengaruhi performa dan konsumsi bahan bakar.

 **Transmission**: Jenis transmisi yang digunakan, seperti transmisi otomatis (Automatic) atau manual (Manual).

 **Fuel Type**: Jenis bahan bakar kendaraan, dikodekan sebagai:

* **X** : Bensin Biasa (Regular Gasoline)
* **Z** : Bensin Premium (Premium Gasoline)
* **D** : Diesel
* **E** : Ethanol (E85)
* **N** : Gas Alam (Natural Gas)

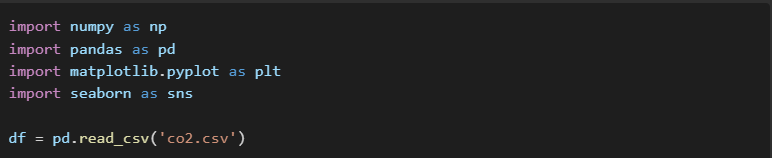
 **Fuel Consumption (City, Hwy, Combined)**: Konsumsi bahan bakar yang diukur dalam liter per 100 kilometer, dibagi berdasarkan:

* **City**: Penggunaan di jalan dalam kota
* **Hwy**: Penggunaan di jalan bebas hambatan
* **Combined**: Rata-rata kombinasi keduanya

 **Fuel Consumption (mpg)**: Konsumsi bahan bakar dalam satuan mil per galon (semakin tinggi = lebih hemat).

 **CO₂ Emissions (g/km)**: Emisi karbon dioksida yang dikeluarkan kendaraan per kilometer. Ini merupakan **variabel target utama** yang ingin diprediksi dalam analisis regresi.

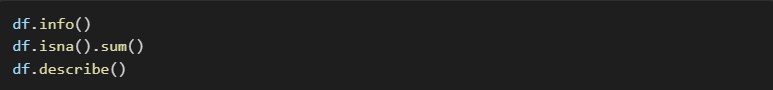
1. **Penjelasan Code**
   1. **Liblary**

****Kode ini mengimpor pustaka-pustaka penting:

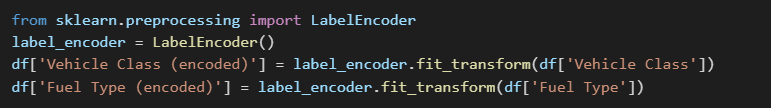
* numpy dan pandas untuk manipulasi data.
* matplotlib dan seaborn untuk visualisasi. Kemudian dataset co2.csv dibaca ke dalam variabel df.
  1. **Eksplorasi Awal Dataset**

****

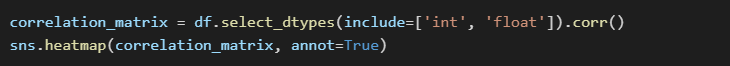
Menampilkan isi awal dari dataset.



* info() menunjukkan tipe data dan jumlah non-null.
* isna().sum() menghitung jumlah nilai kosong per kolom.
* describe() menampilkan statistik deskriptif (min, max, mean, dll).
  1. Label Encoding & Korelasi



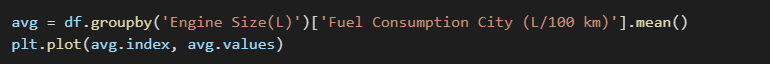
* Mengubah data kategorikal menjadi angka agar bisa digunakan dalam analisis korelasi atau machine learning.



* Membuat matriks korelasi dari semua kolom numerik dan divisualisasikan dalam bentuk heatmap.
  1. Rename Kolom Make Menjadi Brand



* Kolom Make diubah namanya menjadi Brand untuk penamaan yang lebih deskriptif
  1. Line Plot: Engine Size vs Fuel Consumption



* Mengelompokkan data berdasarkan ukuran mesin dan menghitung rata-rata konsumsi BBM dalam kota.
* Menampilkan hasil dalam bentuk grafik garis.
  1. Visualisasi Distribusi Brand, Vehicle Class, Fuel Typer



* + - Menampilkan proporsi brand dalam pie chart.
    - Menampilkan distribusi frekuensi menggunakan countplot untuk Brand, Vehicle Class, dan Fuel Type.
    - df.hist() membuat histogram untuk semua kolom numerik.