

Vehicle Product Research for Better Emission

Data Analyst

Swipe





Our Team



GreatEdu
Home for All Education Needs

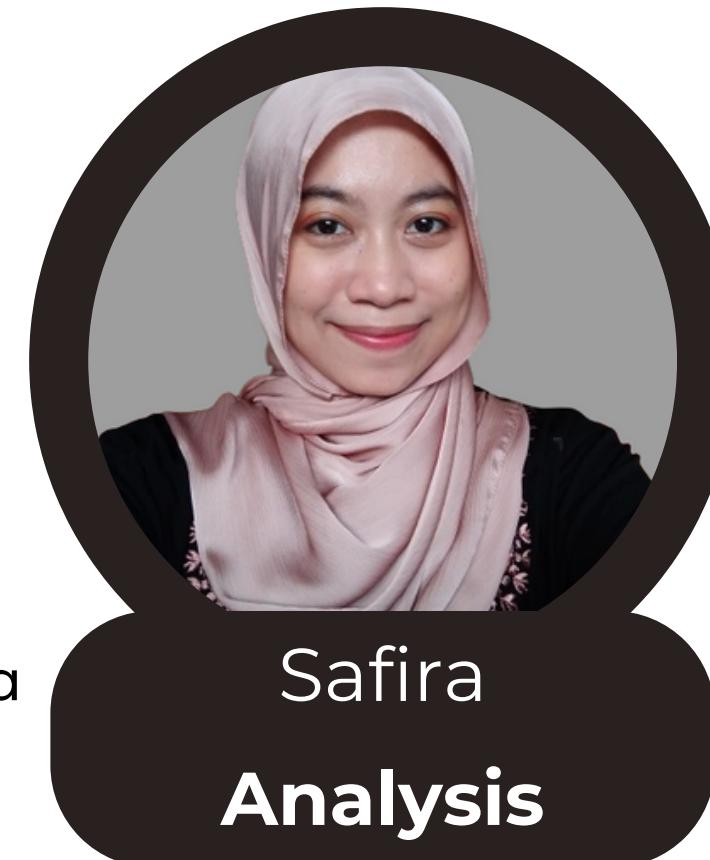
MSIB
magang dan studi independen bersertifikat

**Kampus
Merdeka**
INDONESIA JAYA



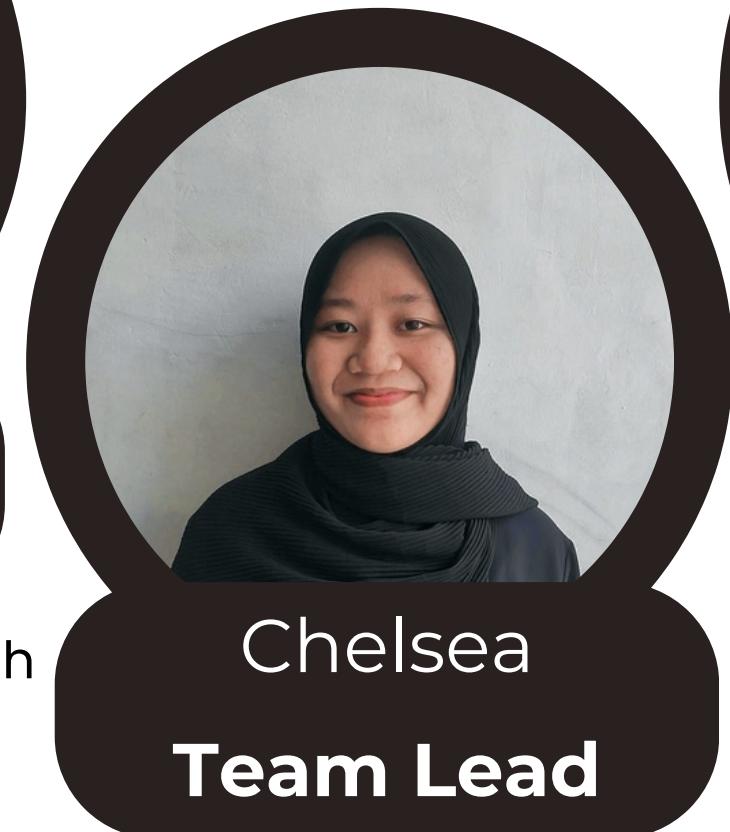
Ikhmal
Visualisasi

Universitas Negeri Jakarta
Tidak Berkontribusi



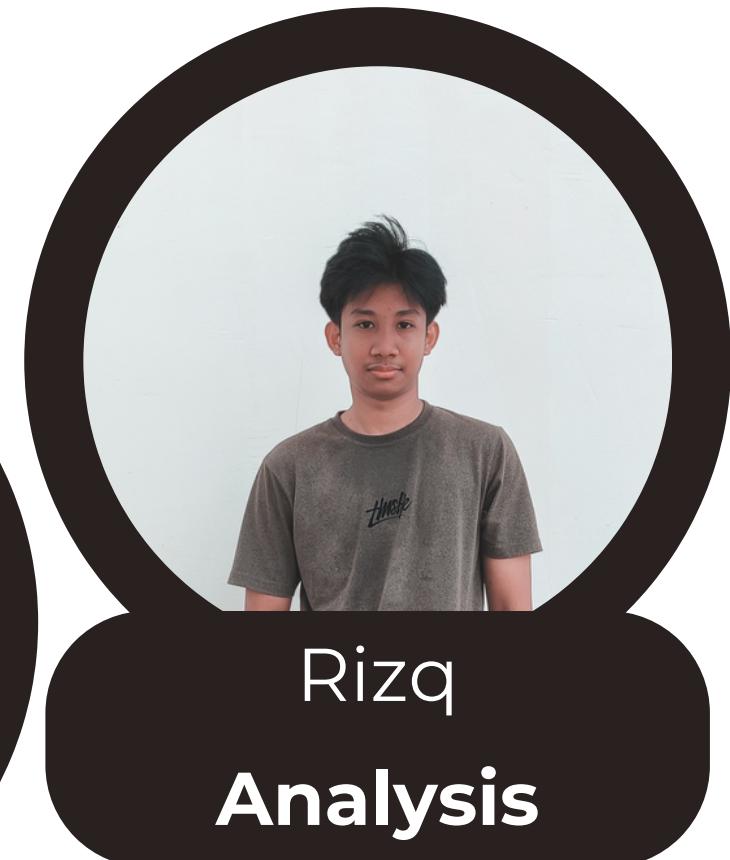
Safira
Analysis

Universitas Muhammadiyah
Prof Dr. Hamka



Chelsea
Team Lead

Universitas Pembangunan Nasional
“Veteran” Jawa Timur



Rizq
Analysis

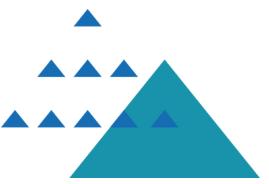
Universitas Mulawarman



Dwi
Visualisasi

Universitas Jember

Project Overview



01 Business Understanding

02 Data Understanding

03 Data Preparation

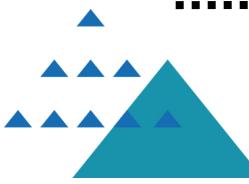
04 Modelling

05 Evaluation

06 Deployment



Business Understanding



FutureDrive Research Group adalah lembaga penelitian transportasi yang ada di Indonesia. Dengan fokus utama pada dampak lingkungan, mereka menyediakan wawasan mendalam terkait pengeluaran emisi gas dari berbagai merek dan kelas kendaraan di pasar Indonesia.

Problem Statement :

- Keluhan pelanggan terkait konsumsi bahan bakar yang boros
- Keluhan pelanggan terkait emisi CO2 yang tinggi

Objektif :

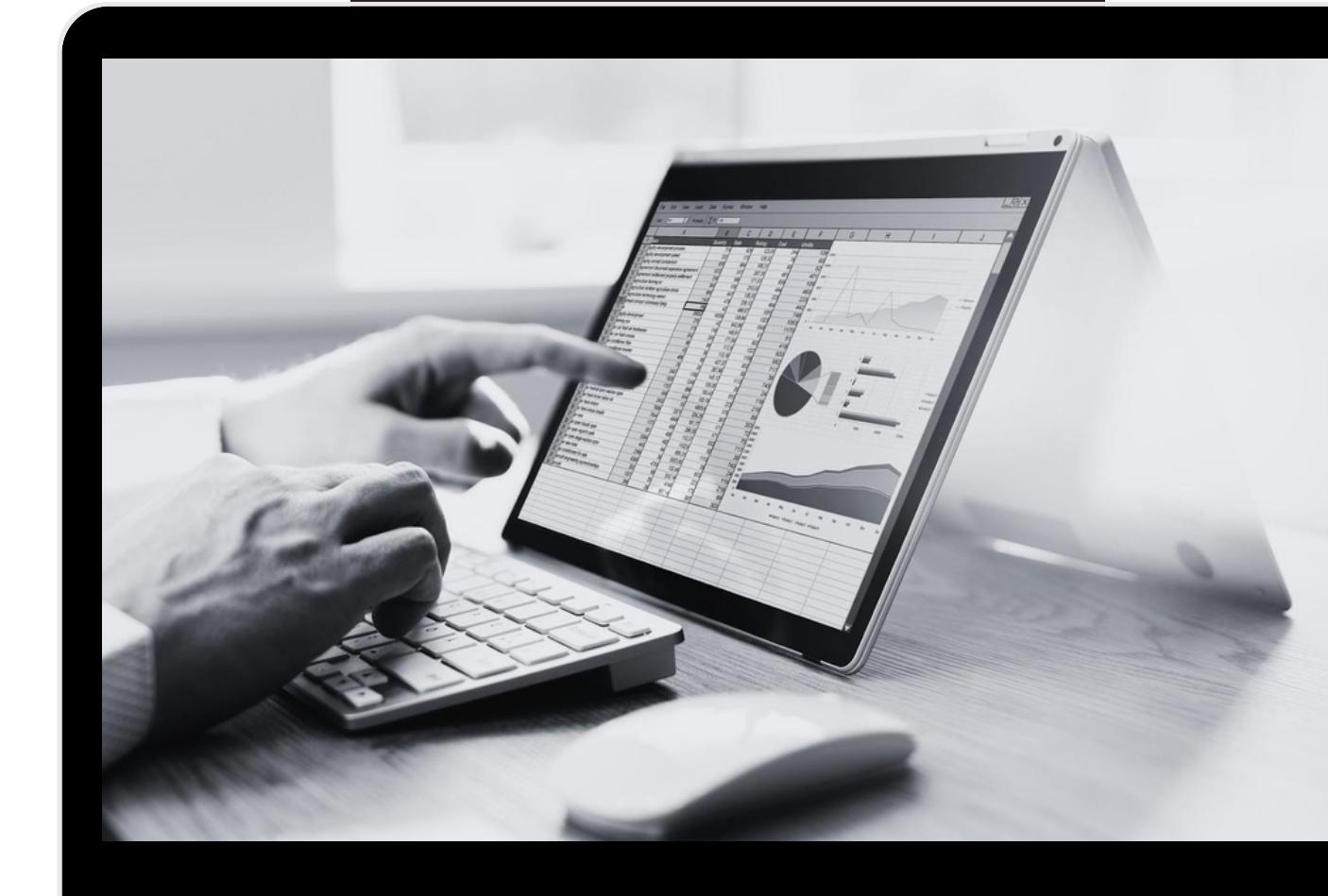
Membuat prediksi emisi CO2 pada tahun 2023 dan 2024





Data Understanding

Dataset yang kami gunakan berasal dari kaggle dan dapat diakses [disini](#).



The screenshot shows the Kaggle interface. On the left, there is a sidebar with navigation links: kaggle (selected), Create, Home, Competitions, Datasets (highlighted in blue), Models, Code, Discussions, Learn, and More. The main content area displays a dataset titled "2022 Fuel Consumption Ratings". It includes a search bar, a user profile for RINI CHRISTY, a download button for a notebook (14 kB), and a preview image of a car. Below the title, it says "Fuel consumption dataset containing Co2 emissions, CO2 ratings and smog ratings". There are tabs for Data Card, Code (14), and Discussion (1). Under "About Dataset", it states: "Dataset provides model-specific fuel consumption ratings and estimated carbon dioxide emissions for new light-duty vehicles for retail sale in Canada in 2022." It also includes sections for Usability (10.00), License (CC0: Public Domain), and Expected update frequency.



Data Understanding

Metadata

Kolom	Keterangan	Catatan	Tipe Data
Model year	Tahun keluaran model mobil	Hanya berisi kendaraan yang diproduksi pada tahun 2022	Integer
Make	Merk mobil		Object
Model	Model dari mobil	<ul style="list-style-type: none">• 4WD/4X4 = four wheel drive• AWD = All wheel drive• CNG = Compressed Natural Gas• FFV = Flexible fuel vehicle• NGV = Natural gas vehicle	Object
Vehicle Class	Kelas kendaraan		Object





Data Understanding

Metadata

Kolom	Keterangan	Catatan	Tipe Data
Engine Size	Kapasitas tangki bensin	Dalam satuan liter	Float
Cylinders	Banyaknya silinder pada mobil		Integer
Transmission	Jenis transmisi mobil	<ul style="list-style-type: none">• A = Automatic• AM = Automated Manual• AS = Automatic with selected shift• AV = Continuously variabel• M = Manual	Object
Fuel Type	Jenis bensin yang digunakan	<ul style="list-style-type: none">• X = Pertalite• Z = Pertamax• D = Diesel• E = Ethanol	Object





Data Understanding

Metadata

Kolom	Keterangan	Catatan	Tipe Data
Fuel Consumption (City (L/100 Km))	Konsumsi bahan bakar penggunaan di dalam kota	Dalam satuan liter/100 Km	Float
Fuel Consumption (Hwy (L/100 Km))	Konsumsi bahan bakar penggunaan di dalam tol	Dalam satuan liter/100 Km	Float
Fuel Consumption (Comb (L/100 Km))	Konsumsi bahan bakar penggunaan kombinasi dengan 55% di dalam kota dan 45% di dalam tol	Dalam satuan liter/100 Km	Float





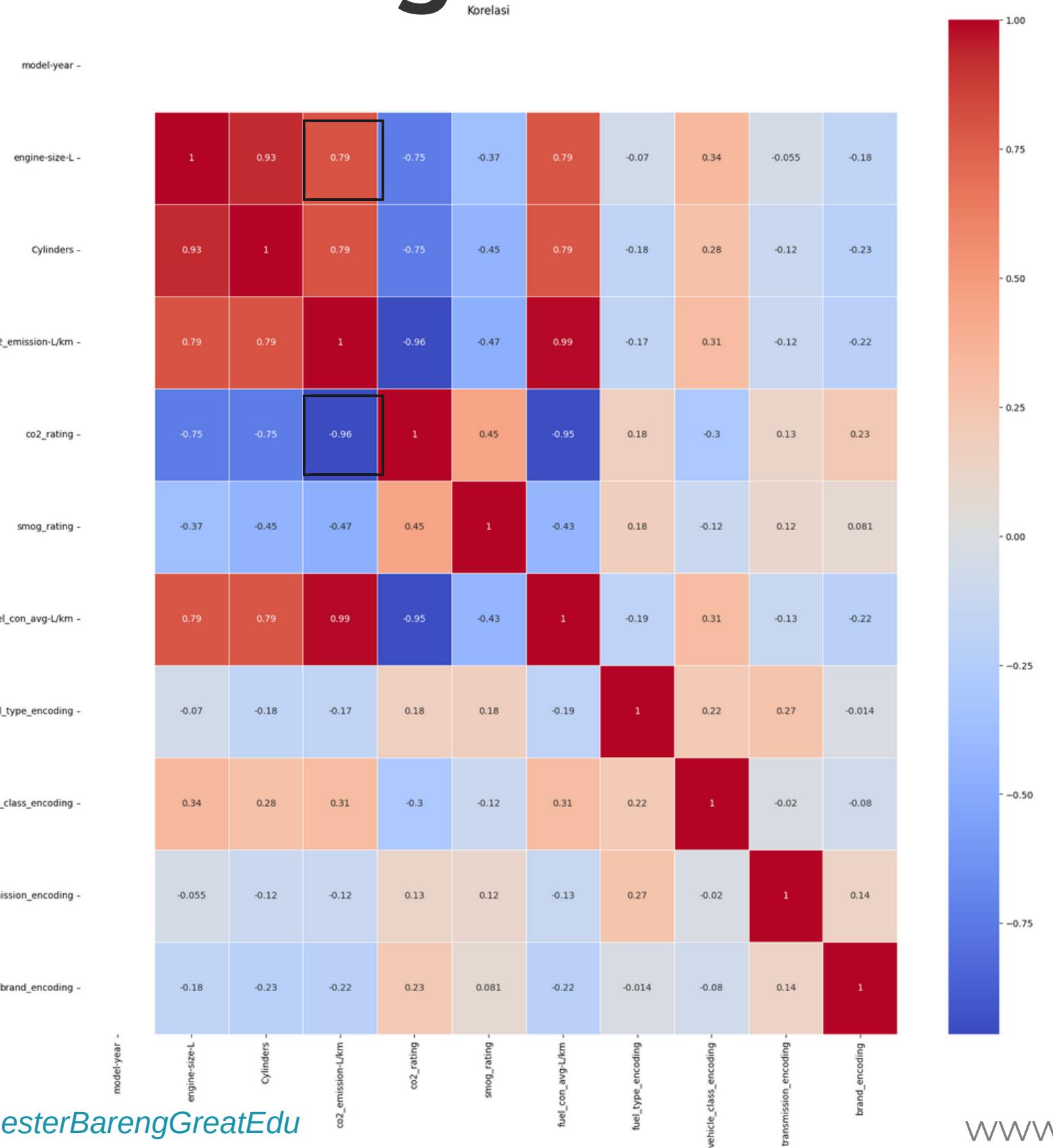
Data Understanding

Metadata

Kolom	Keterangan	Catatan	Tipe Data
Fuel Consumption (Comb (mpg))	Konsumsi bahan bakar penggunaan kombinasi dengan 55% di dalam kota dan 45% di dalam tol	Dalam satuan mpg	Integer
CO2 Emission (g/km)	Emisi karbon dioksida yang dihasilkan	Dalam satuan g/km	Integer
CO2 Rating	Rating emisi CO2 yang dihasilkan	1 = Buruk 10 = Baik	Integer
Smog rating	Rating kabut asap yang dihasilkan	1 = Buruk 10 = Baik	Integer

Data Understanding

Hubungan antar variabel



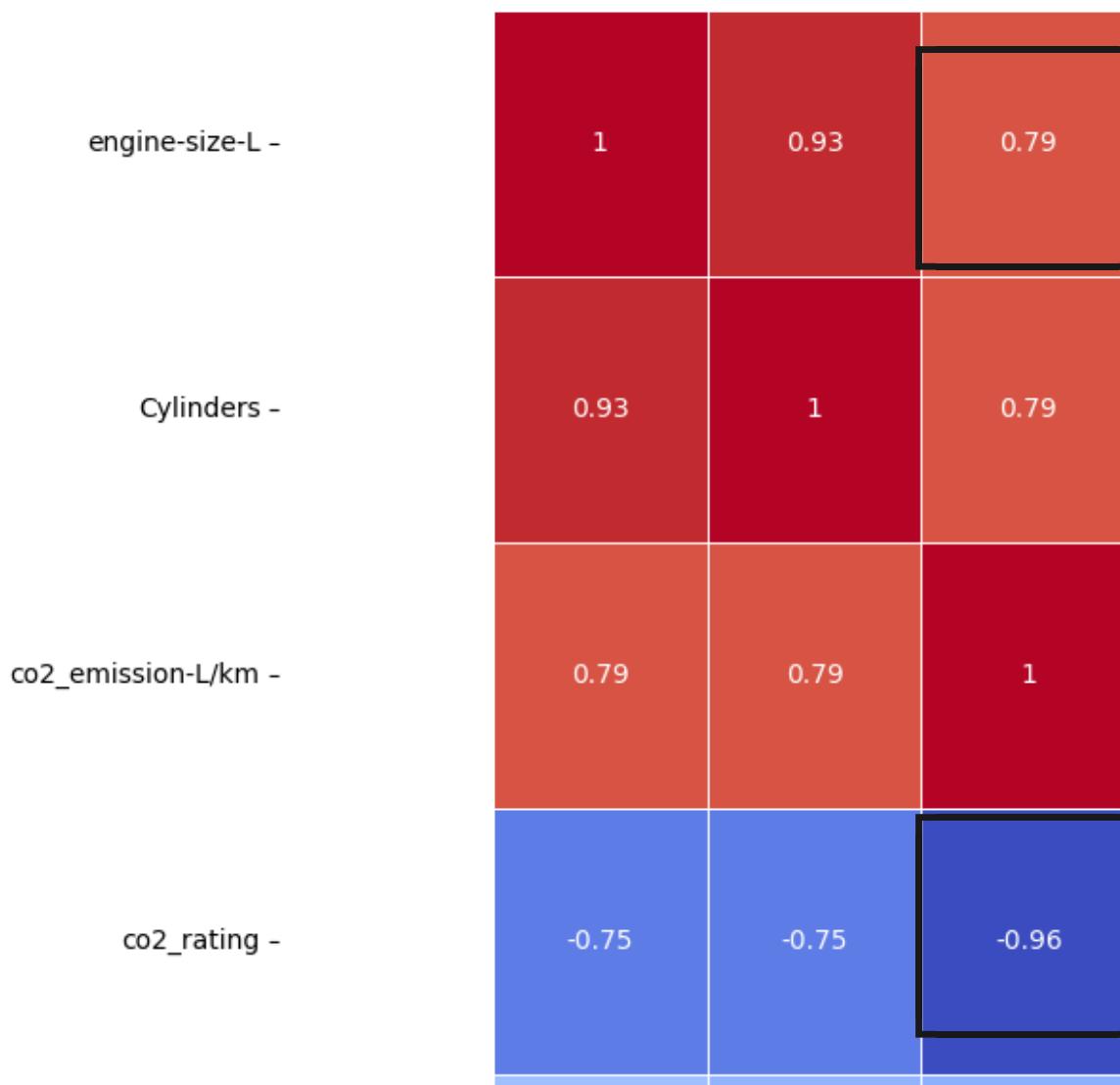
Dari tabel korelasi dapat dilihat bahwasanya :

- Warna gelap menunjukkan korelasi yang kuat
- Warna terang menunjukkan korelasi yang lemah
- Korelasi positif = jika suatu kolom meningkat, kolom lainnya cenderung meningkat
- Korelasi negatif = jika suatu kolom meningkat, kolom lainnya cenderung menurun



Data Understanding

Hubungan antar variabel

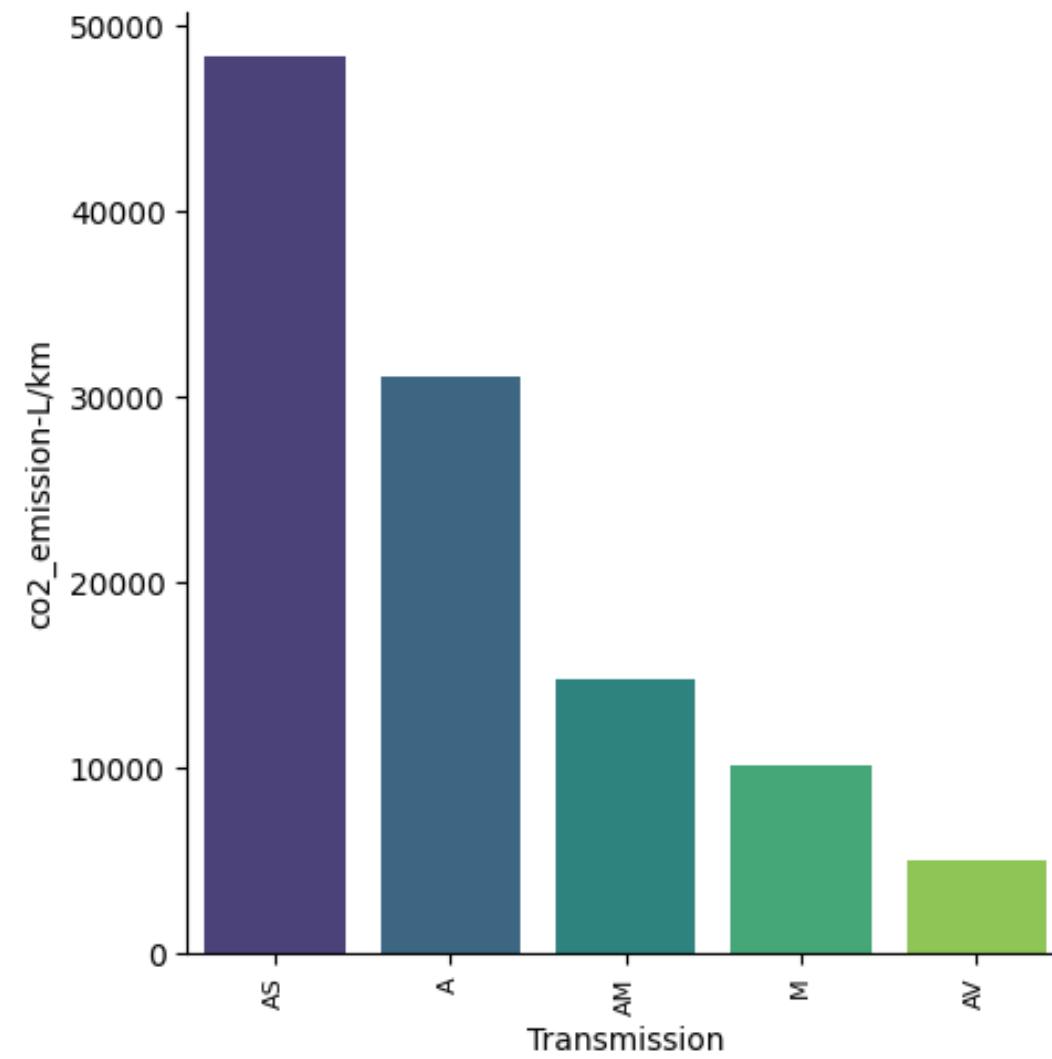


Contoh :

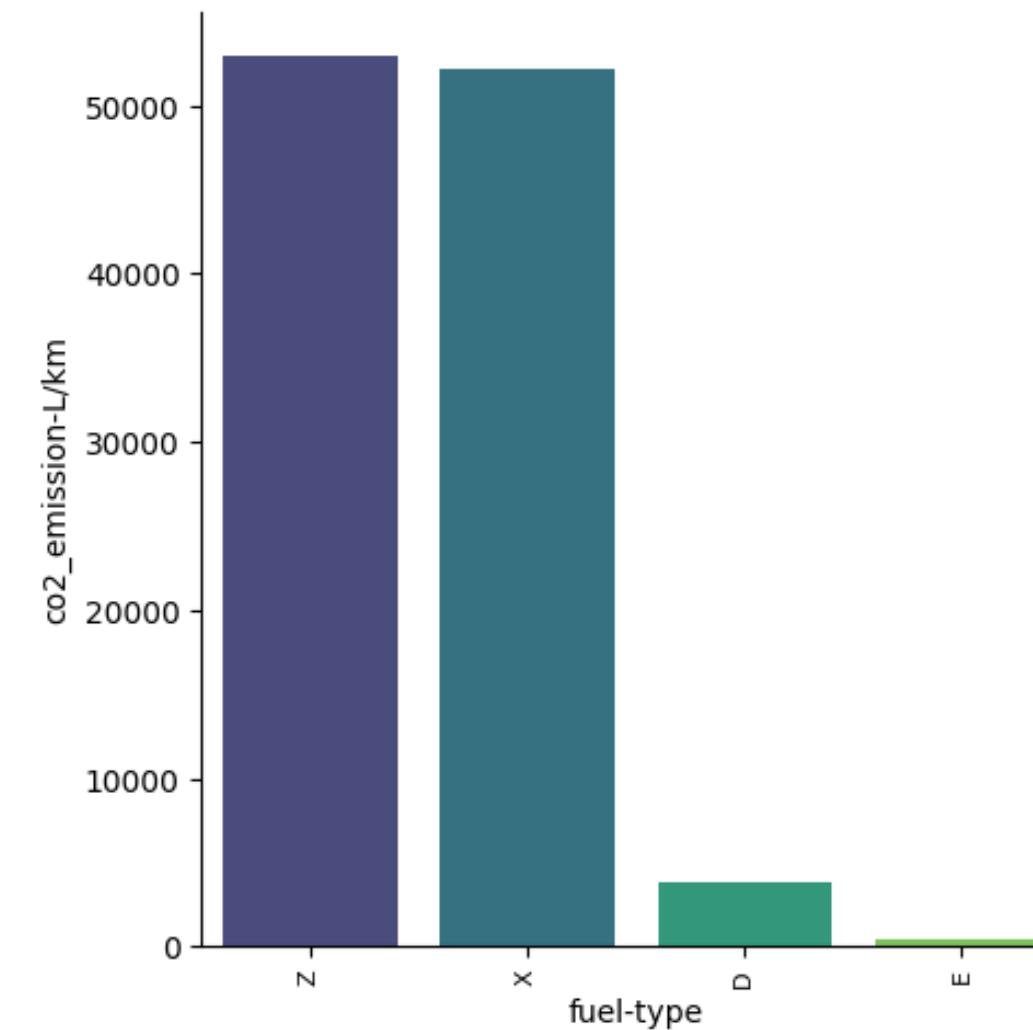
CO2 emission memiliki korelasi yang kuat dengan engine size dan memiliki korelasi yang lemah dengan CO2 rating. Jika CO2 emission bertambah, maka engine size akan ikut bertambah. Tetapi ketika CO2 bertambah, maka CO2 rating menurun

Data Understanding

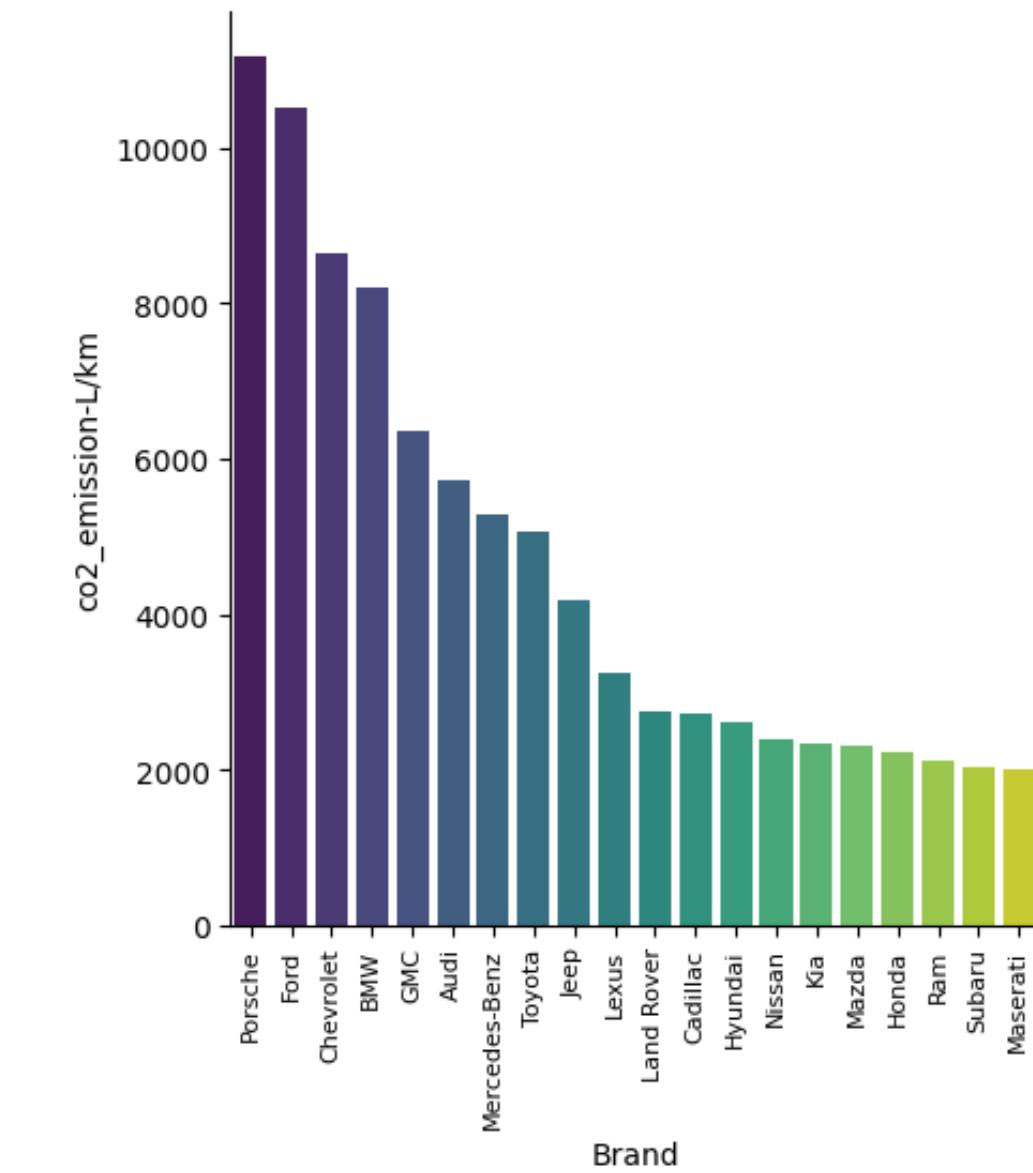
Visualisasi tiap kolom berdasarkan jumlah emisi CO2 yang dihasilkan



- A = Automatic
- AM = Automated Manual
- AS = Automatic with selected shift
- AV = Continuously variabel
- M = Manual



- X = Pertalite (RON 92)
- Z = Pertamax Green (RON 95)
- D = Diesel
- E = Ethanol (E5)





Data Preparation

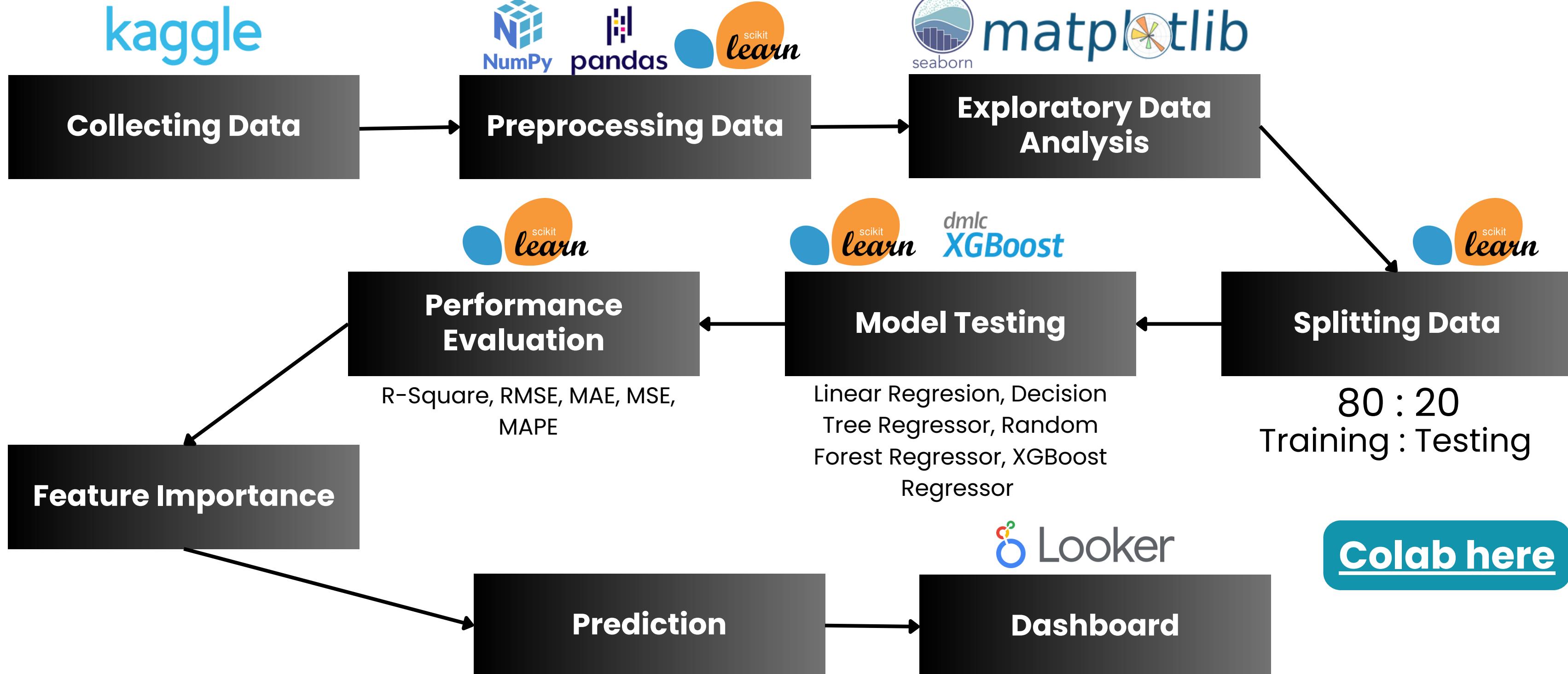
Pada tahap ini akan dilakukan proses pembersihan data hingga data siap untuk dilakukan pemodelan dan visualisasi dashboard



Data Preparation

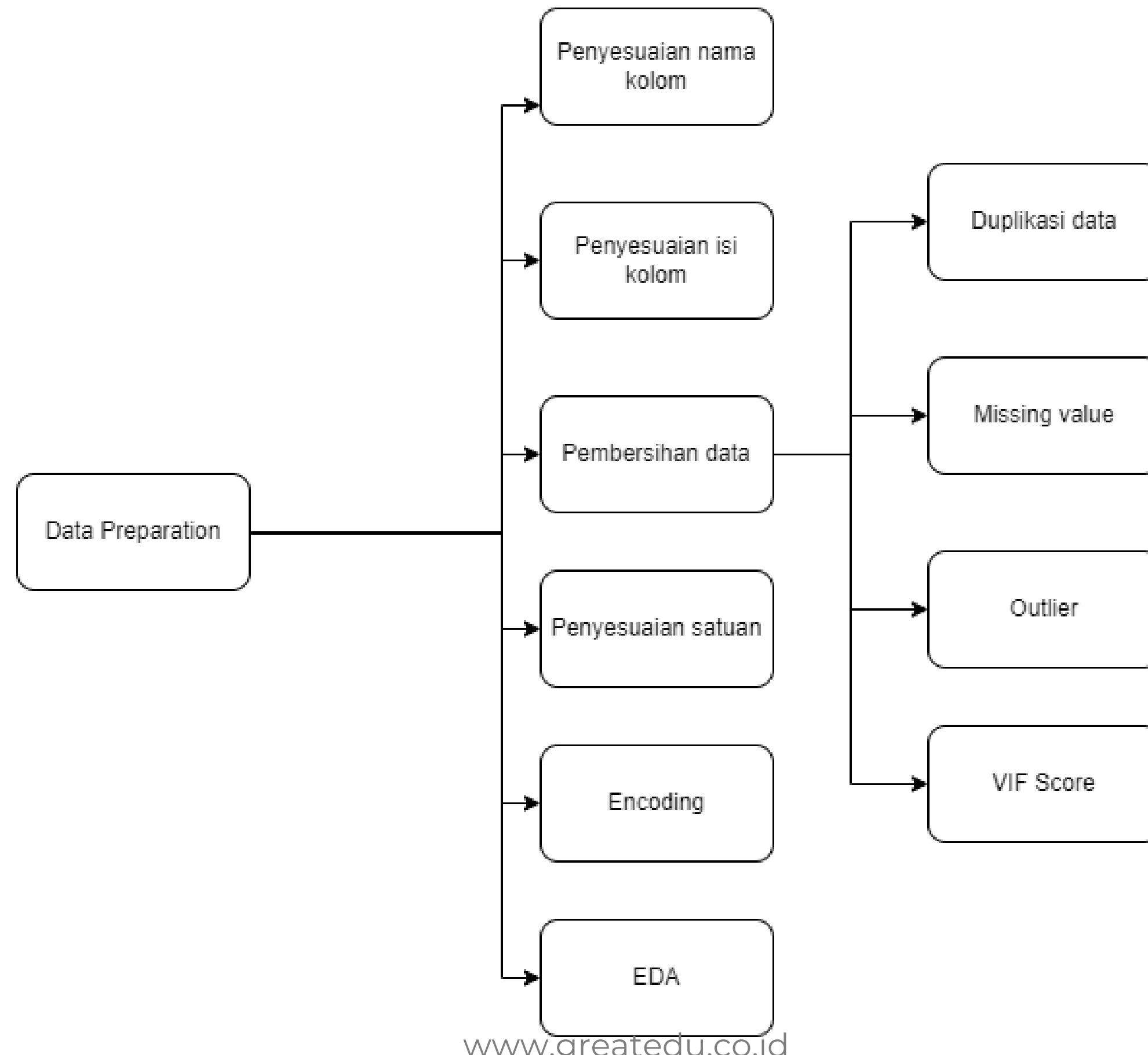


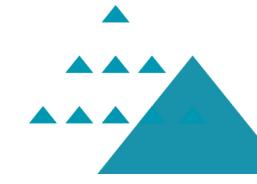
Workflow CRISP-DM





Data Preparation





Data Modelling

Dalam tahap pemodelan, kami menjalankan beberapa langkah untuk mendapatkan tingkat akurasi dari setiap model, yang mencakup:

- Memisahkan data menjadi variabel dependen dan independen.
- Melakukan pembagian data menjadi data latih dan data uji dengan rasio 80% dan 20%.
- Menggunakan alat untuk melakukan standarisasi pada setiap kolom yang akan dimodelkan.
- Membuat daftar yang berisi model-model yang akan dievaluasi.



Data Modeling

Terdapat beberapa 4 model yang digunakan dalam analisis regresi pada CO2 Emission yaitu :

1

Linear Regression

Linear Regression, sebagai metode statistik sederhana, digunakan untuk menemukan hubungan linear antara variabel independen dan variabel dependen.

2

Decision Tree Regression

Decision Tree Regression membangun model dalam bentuk pohon keputusan, memungkinkan penanganan hubungan non-linear dan interaksi antar variabel.

3

Random Forest Regression

Random Forest Regression, sebagai metode ensemble, menggunakan banyak pohon keputusan untuk mengurangi overfitting dan meningkatkan akurasi

4

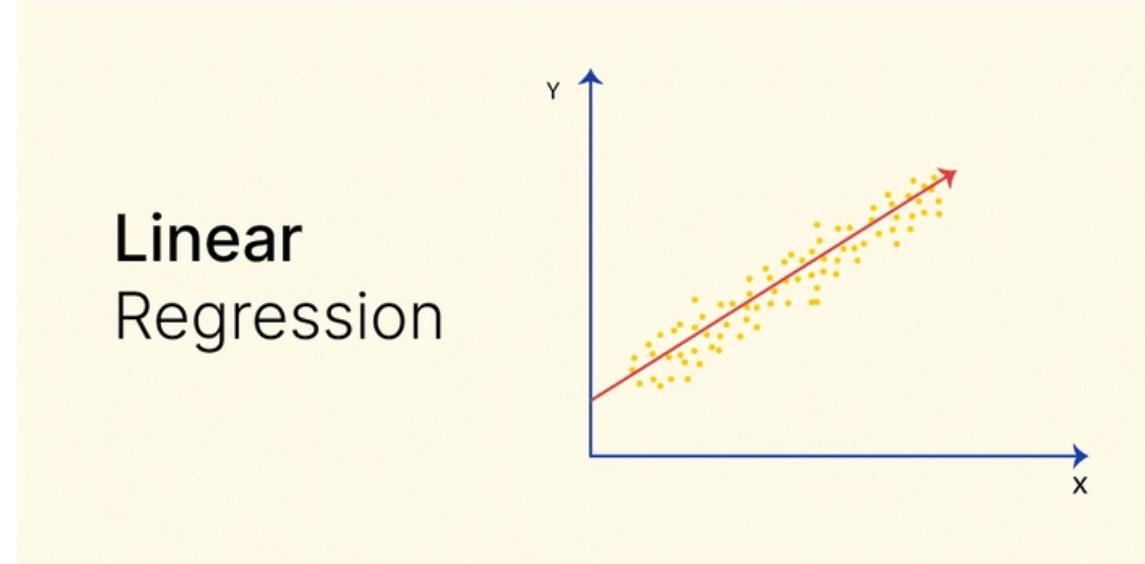
XGBoost Regression

XGBoost Regression, yang merupakan algoritma boosting, menggabungkan pohon keputusan untuk menciptakan model yang kuat, namun memerlukan tuning parameter yang cermat

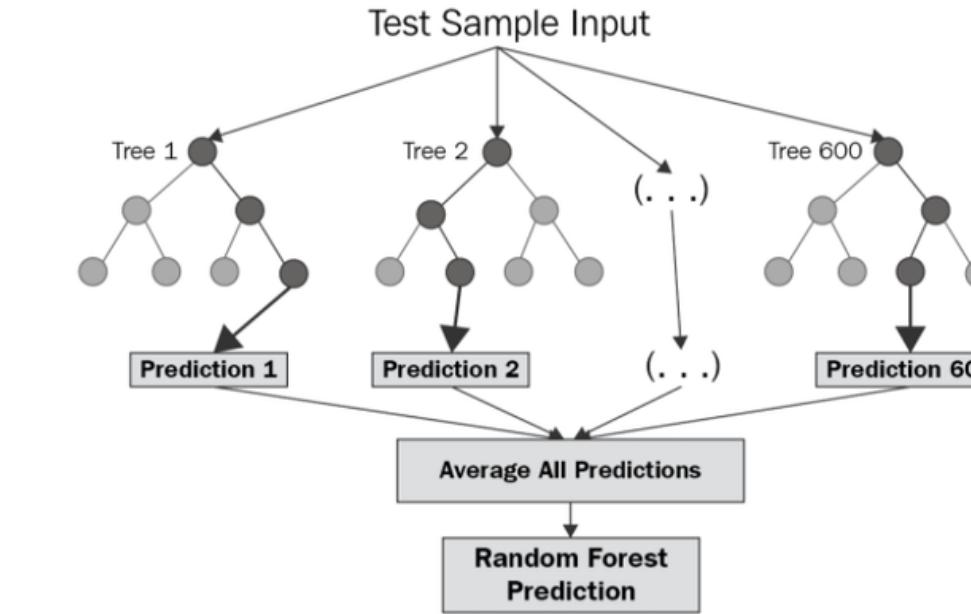
Modelling



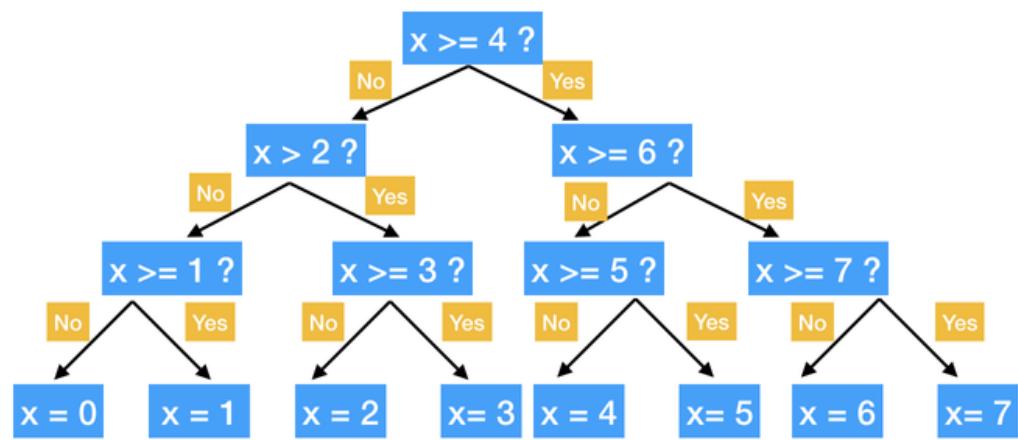
Linear Regression



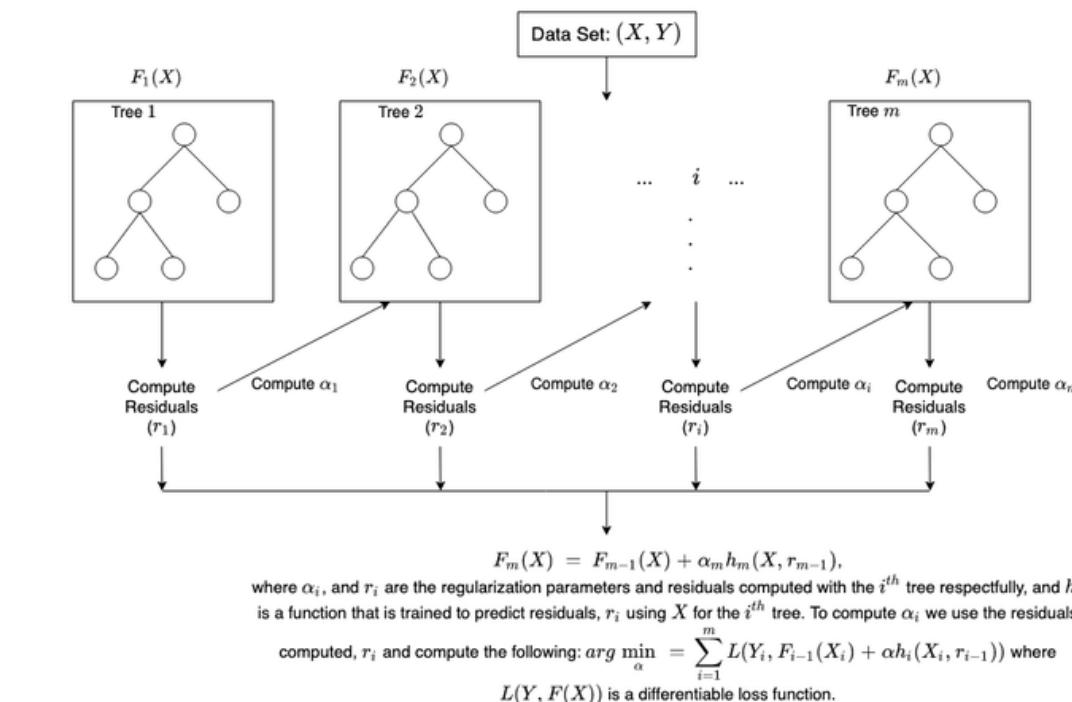
Random Forest Regression



Decision Tree Regression



XGBoost Regression





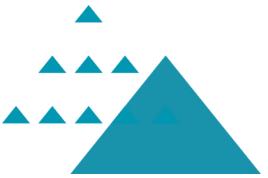
Performance Evaluation



GreatEdu
Home for All Education Needs

MSIB
magang dan studi independen bersertifikat

Kampus
Merdeka
INDONESIA JAYA



	Model	MAE	MSE	RMSE	MAPE	R-squared
0	Linear Regression	1.948836	26.072051	5.106080	0.015420	0.962143
1	Decision Tree Regression	0.973624	15.334036	3.915870	0.007991	0.977735
2	Random Forest Regression	0.960500	10.685022	3.268795	0.007626	0.984485
3	XGBoost Regression	1.040179	22.910116	4.786451	0.008012	0.966734

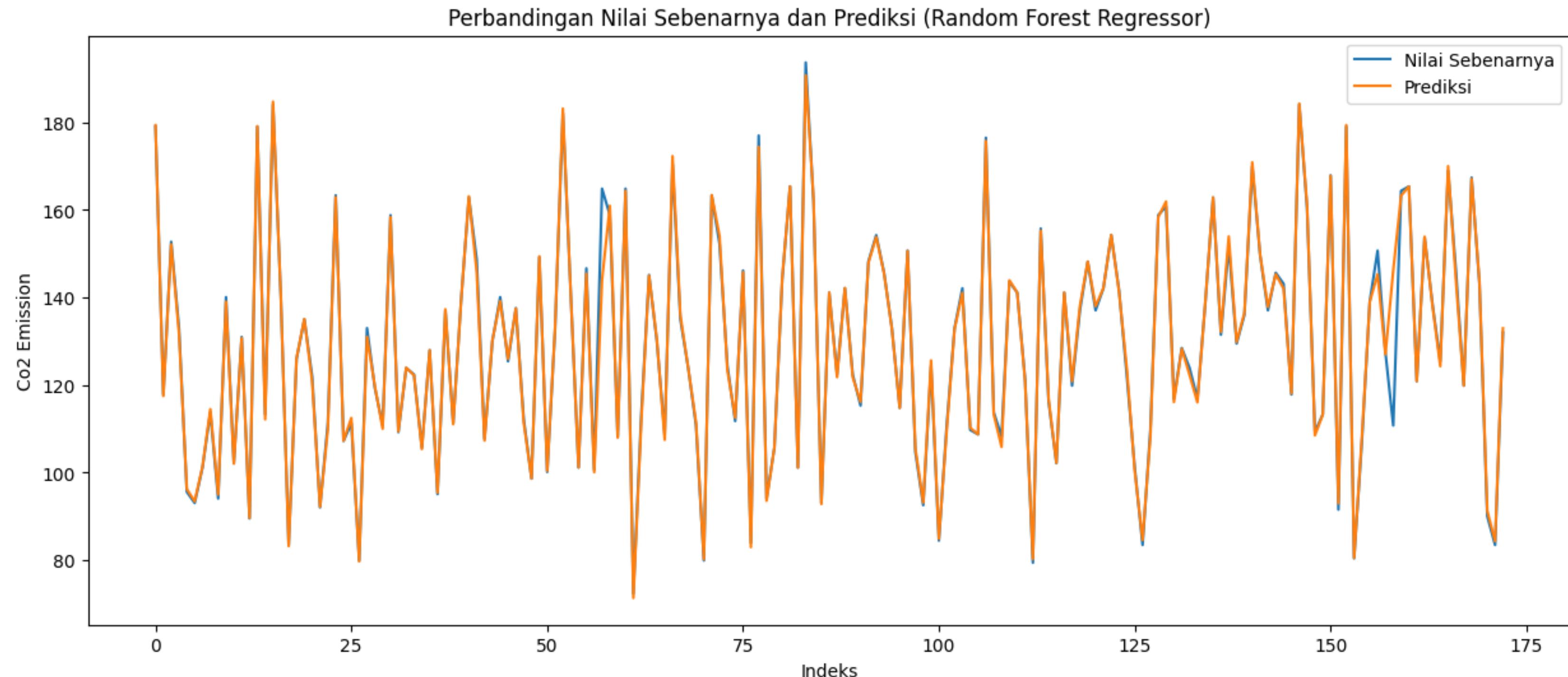


Best Model :

Random Forest
Regression



Visualization Model





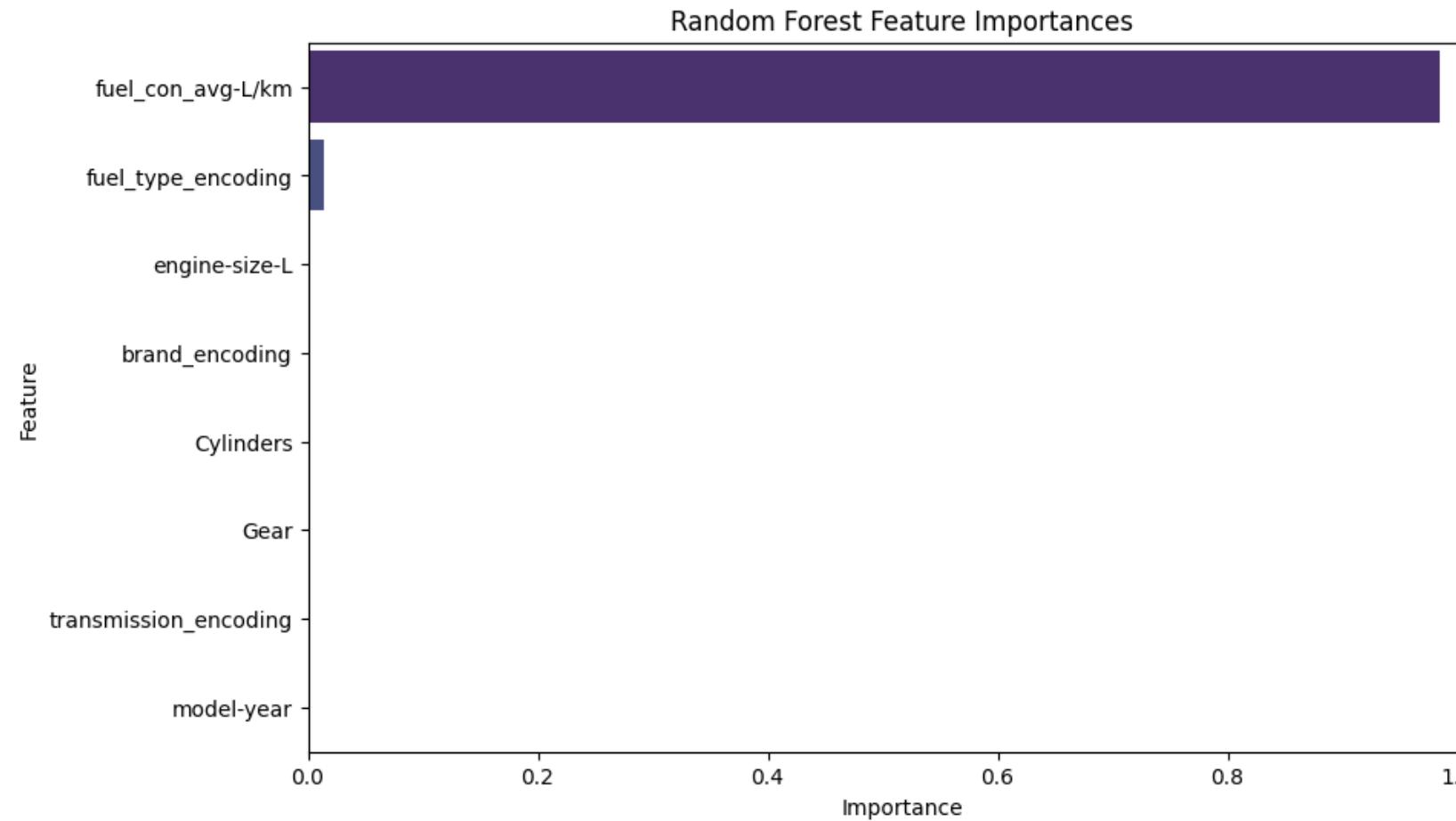
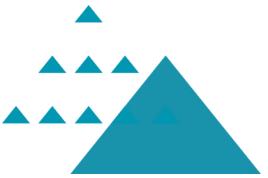
Feature Importance



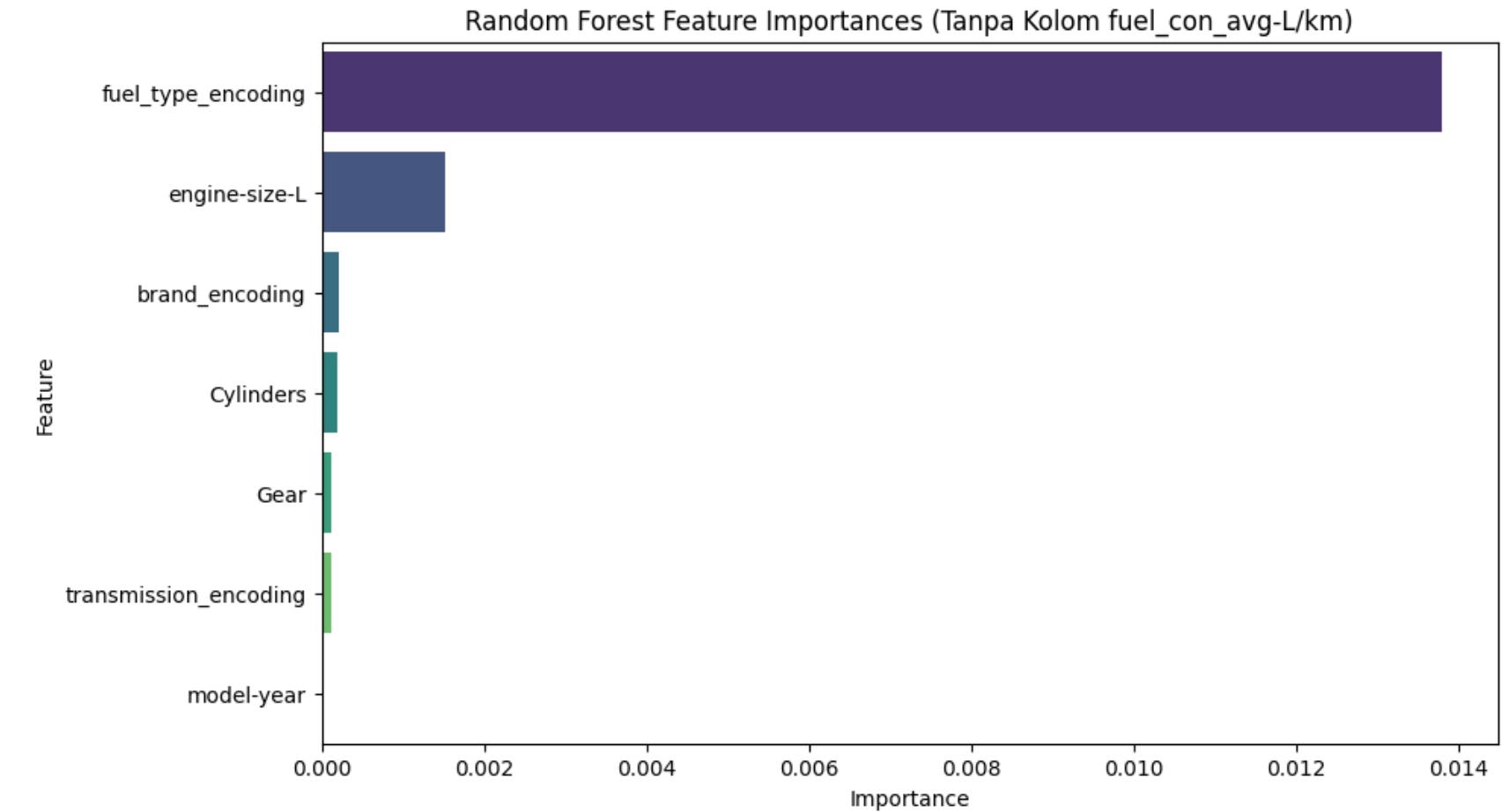
GreatEdu
Home for All Education Needs

MSIB
magang dan studi independen bersertifikat

Kampus
Merdeka
INDONESIA JAYA



Feature Importance



Feature Importance
tanpa (fuel_con_avg-L/km)

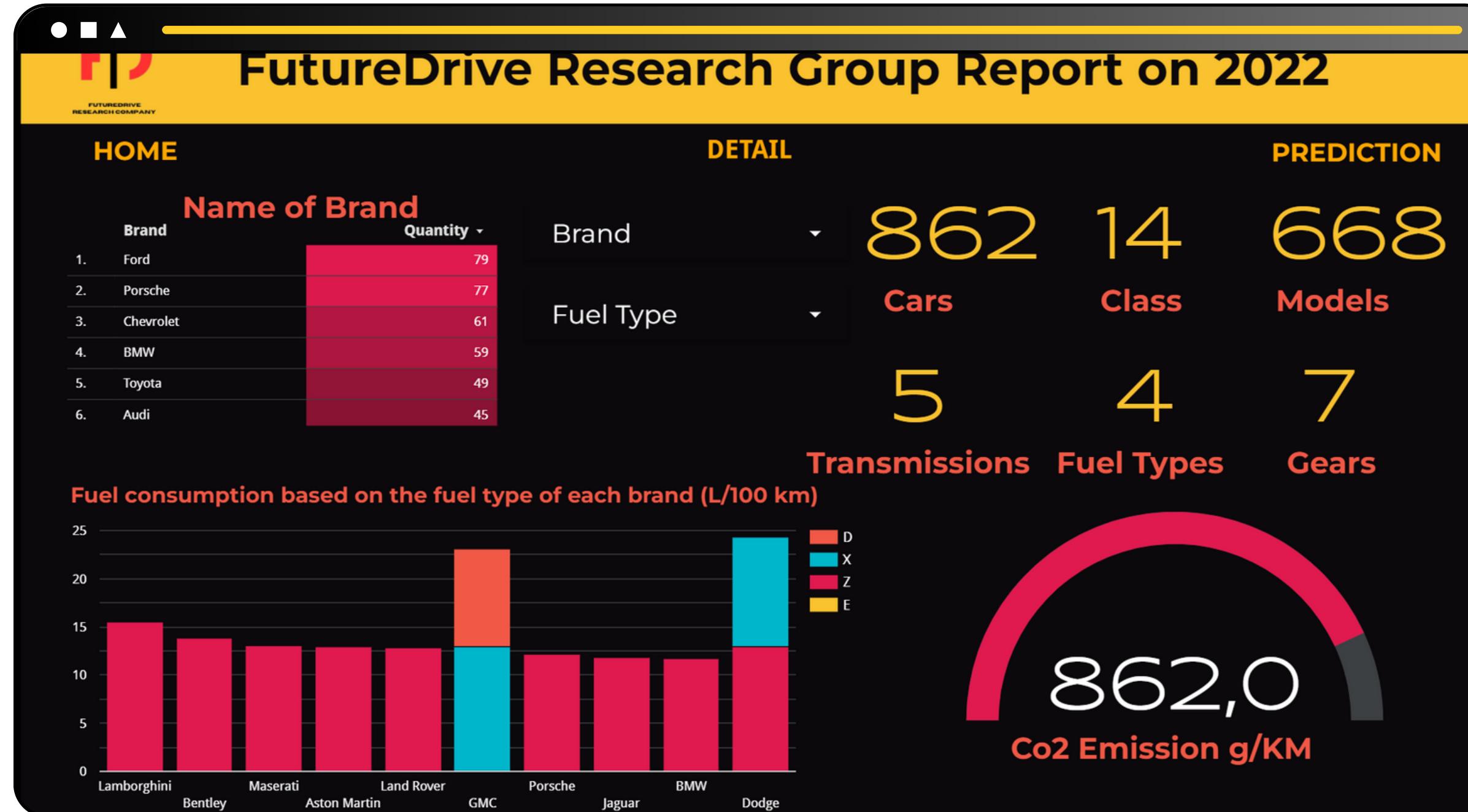
Deployment



Berdasarkan prediksi data yang kami peroleh, kami merekomendasikan kepada para produsen mobil untuk meningkatkan produksi mobil bertenaga baterai. Sedangkan untuk para pengguna mobil, disarankan untuk beralih ke tipe bahan bakar pertamax green (RON 95) atau tipe bahan bakar lainnya dengan RON diatas 95.

Dashboard

Kami menyediakan visualisasi dashboard yang lebih interaktif. Dashboard dapat dikunjungi pada link [berikut](#).





**Terima
Kasih.**