

CO2 Emissions by Vehicles Prediction

Dimas G – Final Project – dibimbing.id

Factors Affecting The Emissions by Vehicles In The Canada



Summary



Project ini bertujuan untuk memahami faktor-faktor yang mempengaruhi emisi CO2 dan membangun model untuk memprediksi emisi tersebut. Fokus kami adalah mengembangkan kendaraan dengan emisi karbon minimal dan mendukung lingkungan yang lebih bersih.

Dalam rangka mencapai tujuan ini, kami mengidentifikasi beberapa faktor kunci, seperti efisiensi bahan bakar, jumlah silinder, merek dan kelas kendaraan, serta jenis bahan bakar dan transmisi yang berdampak pada emisi CO2.

Dengan memahami faktor-faktor ini, kami merekomendasikan langkah-langkah seperti penelitian dan pengembangan kendaraan ramah lingkungan, pemilihan jenis transmisi yang tepat, dan penghargaan untuk efisiensi, yang dapat meningkatkan daya saing di pasar otomotif, memberikan kontribusi pada lingkungan, dan mendorong inovasi dan teknologi baru di industri otomotif.



Table of contents



01

Business Understanding

02

Data Understanding

03

Exploratory Data
Analysis

04

Data Preprocessing

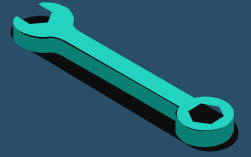
05

Machine Learning

06

Conclusion &
Recommendation





01

Business Understanding





Background

- Kekhawatiran global tentang perubahan iklim dan dampaknya terhadap lingkungan.
- Memahami emisi sangat penting untuk persaingan di pasar otomotif



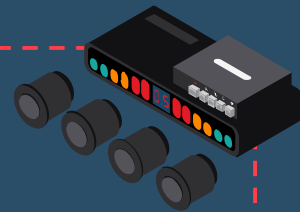
Problem

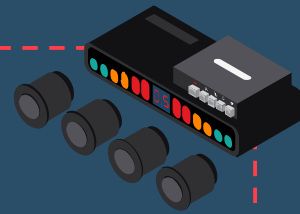
- Mengembangkan kendaraan dengan emisi karbon minimal
- Mengidentifikasi faktor-faktor yang mempengaruhi emisi CO2



Goals

- Memprediksi emisi CO2 dan memahami faktor-faktor yang mempengaruhi emisi CO2
- Optimalkan manufaktur dan desain untuk kendaraan ramah lingkungan





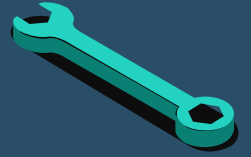
Objective

- membangun model untuk memprediksi emisi CO2
- Mengidentifikasi faktor-faktor kunci yang berdampak pada emisi



Benefits

- Meningkatkan daya saing di industri otomotif
- Mendorong inovasi dan berkontribusi pada lingkungan yang lebih bersih

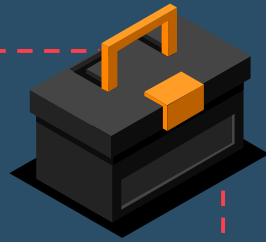


02

Data Understanding



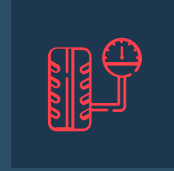
Data Information



There are total 7385 rows and 12 columns.



24 Feature

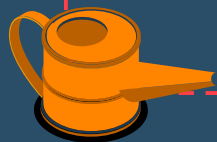


1 Target





| Columns | Description |
|---|--|
| Make | perusahaan otomotif kendaraan |
| Model | jenis atau varian spesifik dari kendaraan |
| Vehicle Class | Merupakan kategori atau tipe kendaraan berdasarkan ukuran, kapasitas, dan fungsi |
| Engine Size (L) | Ukuran mesin kendaraan yang diukur dalam liter (L). Ukuran mesin biasanya berkaitan dengan kapasitas dan tenaga kendaraan. |
| Cylinders | Jumlah silinder dalam mesin kendaraan. Semakin banyak silinder, umumnya semakin besar daya mesinnya. |
| Transmission | Jenis transmisi kendaraan |
| Fuel Type | Jenis bahan bakar yang digunakan oleh kendaraan |
| Fuel consumption in city roads (L/100 km) | Rata-rata konsumsi bahan bakar kendaraan dalam liter per 100 kilometer saat digunakan di area perkotaan. |
| Fuel consumption in highways (L/100 km) | Rata-rata konsumsi bahan bakar kendaraan dalam liter per 100 kilometer saat digunakan di jalan tol atau jalan raya. |
| fuel consumption Comb | Rata-rata kombinasi konsumsi bahan bakar kendaraan dalam liter per 100 kilometer, dihitung berdasarkan perbandingan penggunaan di area perkotaan (55%) dan jalan tol/jalan raya (45%). |
| Fuel Consumption Comb (mpg) | Ukuran efisiensi bahan bakar kendaraan dalam mil per galon (miles per gallon). |
| CO2 Emissions | Emisi karbon dioksida yang dihasilkan oleh kendaraan, diukur dalam gram per kilometer, berdasarkan penggunaan kombinasi di area perkotaan dan jalan tol/jalan raya. |



Meta Data

Model

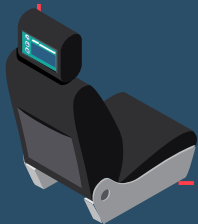
4WD/4X4 = Four-wheel drive
AWD = All-wheel drive
FFV = Flexible-fuel vehicle
SWB = Short wheelbase
LWB = Long wheelbase
EWB = Extended wheelbase

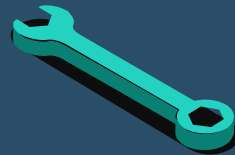
Fuel Type

X = regular gasoline
Z = premium gasoline
D = diesel
E = ethanol (E85)
N = natural gas

Transmission

A = automatic
AM = automated manual
AS = automatic selectshift
AV = continuously variable
M = manual

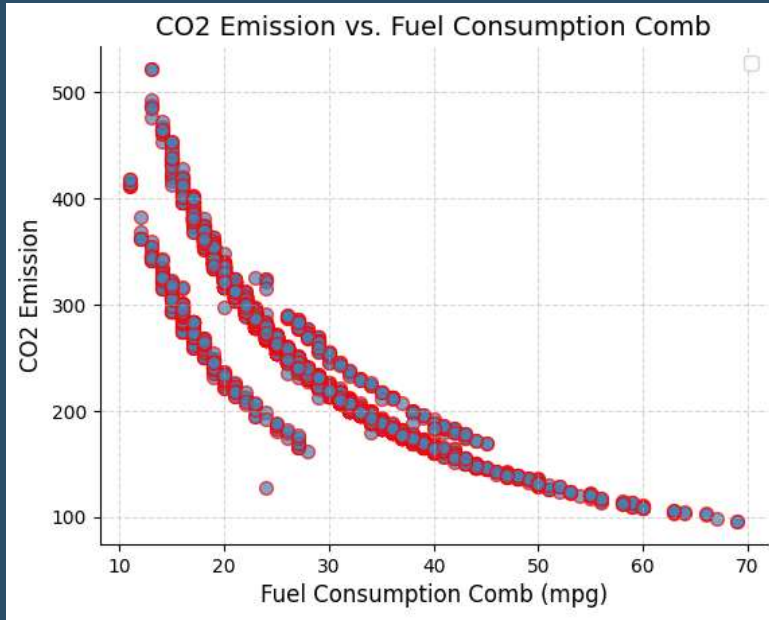




03

Exploratory Data Analysis



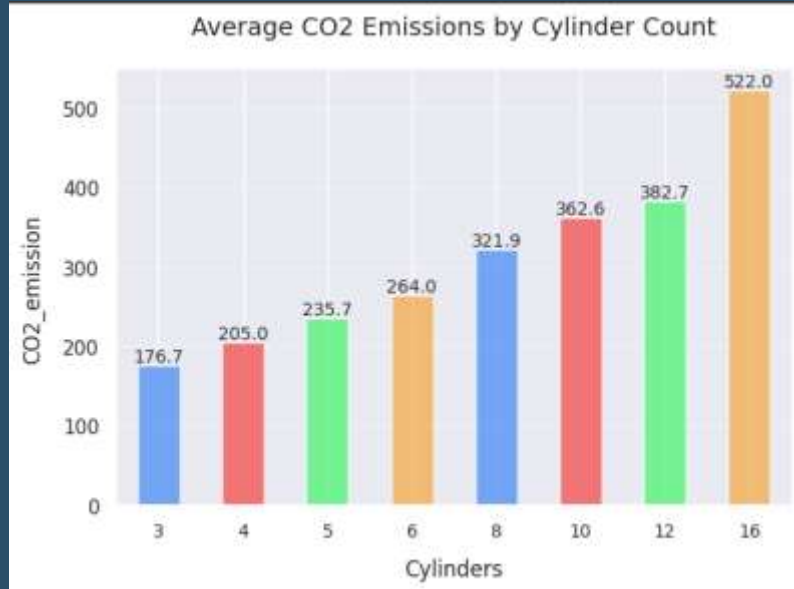


Terdapat hubungan berbanding terbalik antara Fuel Consumption Comb (mpg) dan emisi CO₂. Semakin tinggi nilai Fuel Consumption Comb (mpg), semakin rendah emisi CO₂ yang dihasilkan.

Kendaraan yang lebih efisien dalam hal konsumsi bahan bakar cenderung menghasilkan emisi CO₂ yang lebih rendah.

Peningkatan efisiensi bahan bakar dapat menjadi salah satu strategi penting dalam mengurangi emisi CO₂ dan menjaga lingkungan yang lebih bersih

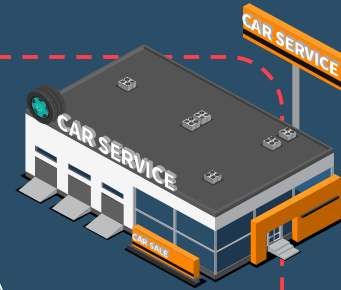


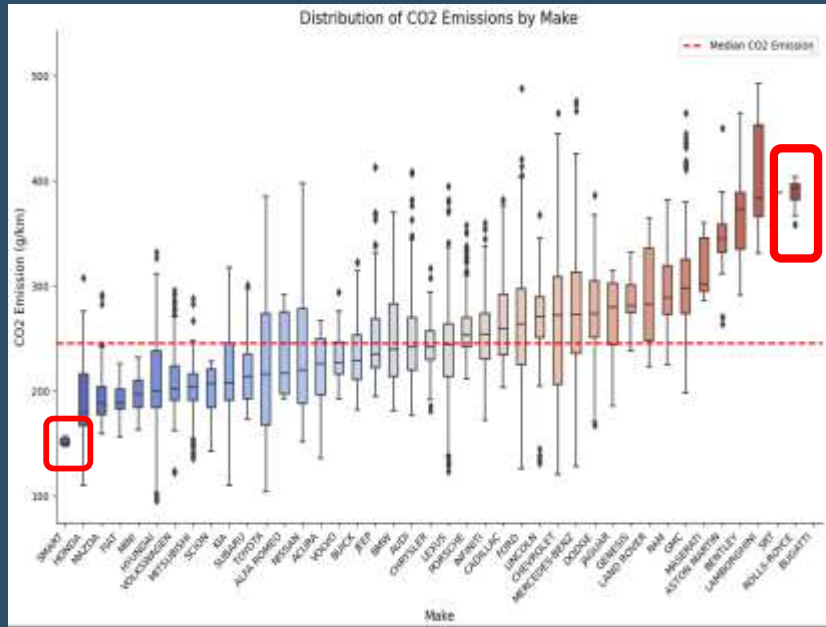


Semakin banyak jumlah silinder, semakin besar pula emisi CO2 yang dihasilkan.

Kendaraan dengan lebih banyak silinder cenderung menghasilkan emisi CO2 yang lebih tinggi dibandingkan dengan kendaraan dengan jumlah silinder yang lebih sedikit.

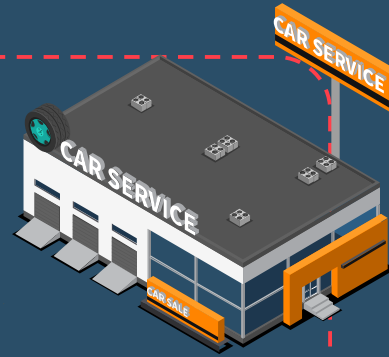
Pentingnya mempertimbangkan jumlah silinder dalam upaya mengurangi emisi CO2 dan menjaga lingkungan yang lebih bersih.

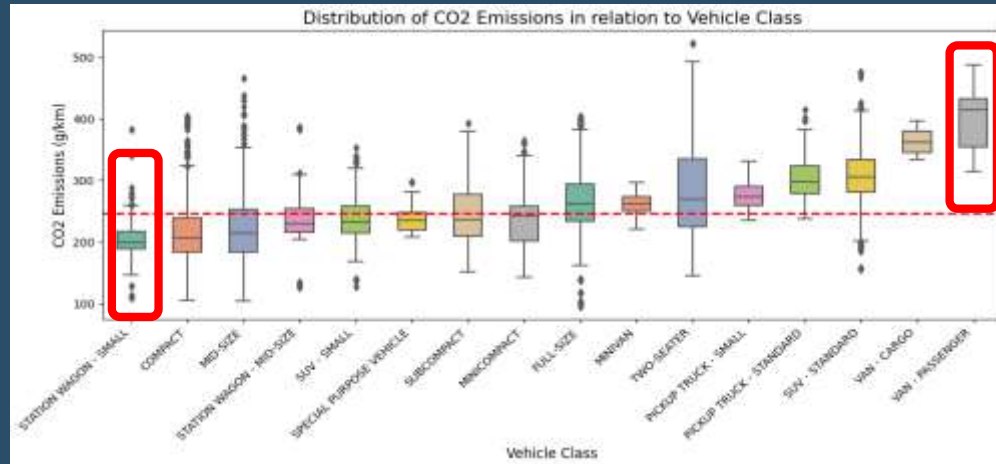




Bugatti memiliki rata-rata emisi CO2 tertinggi, sedangkan SMART memiliki emisi terendah.

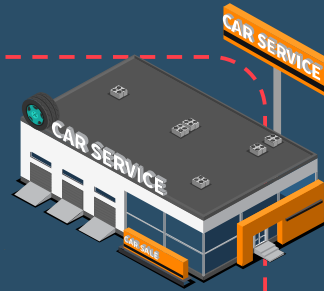
Hampir setengah dari merek memiliki emisi CO2 median lebih rendah dari median emisi CO2 keseluruhan.

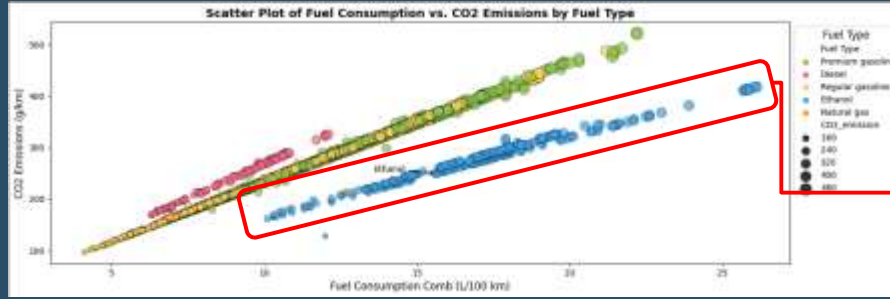




Kendaraan VAN-Passenger dan VAN-Cargo memiliki emisi CO2 tertinggi. Sedangkan, kendaraan Station Wagon-Small memiliki emisi CO2 terendah.

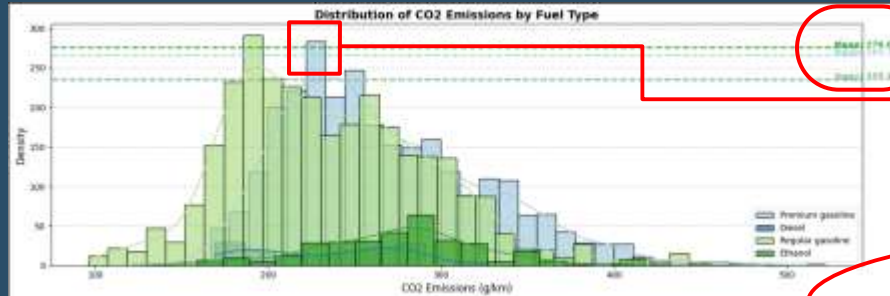
Dari 16 kelas kendaraan, setengahnya memiliki emisi CO2 di atas median sampel, dan setengahnya memiliki emisi di bawah median tersebut





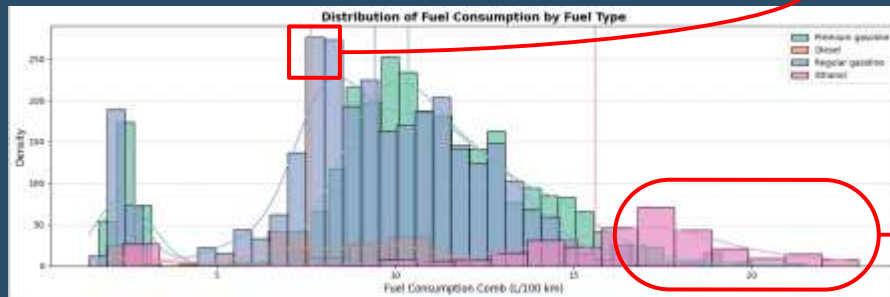
Hubungan Konsumsi Bahan Bakar dan Emisi CO2:

- Lereng yang curam menunjukkan peningkatan emisi CO2 untuk konsumsi bahan bakar yang tinggi.
- Bahan bakar berbasis etanol memiliki kemiringan terendah, menghasilkan emisi CO2 lebih rendah.



Distribusi Emisi CO2 berdasarkan Jenis Bahan Bakar:

- Bahan bakar berbasis etanol rata-rata menghasilkan emisi CO2 tertinggi.
- Diesel memiliki beberapa puncak dalam distribusinya, menunjukkan kelompok kendaraan yang berbeda.



Distribusi Konsumsi Bahan Bakar berdasarkan Jenis Bahan Bakar:

- Bensin biasa memiliki frekuensi kendaraan tertinggi dengan konsumsi bahan bakar lebih rendah.
- Bahan bakar berbasis etanol memiliki konsumsi bahan bakar secara keseluruhan lebih besar.

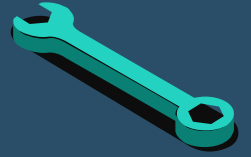
Perbedaan Bensin Biasa dan Bensin Premium:

- Kendaraan bensin biasa memiliki frekuensi lebih tinggi dengan emisi CO2 lebih rendah.
- Kendaraan bensin premium memiliki frekuensi lebih tinggi dengan emisi CO2 lebih tinggi, karena sifat kinerja yang berbeda.



04

Data Preprocessing



Preprocessing Timeline



Data Cleansing

Missing Value
Duplicated Columns
Checking data types



Encoding

Frequently Encoding
data
One Hot Encoding



Handling Outliers

Log transform from
7385 rows to 6282
rows



Train Test Split

Split dataset into Train
and test data



Handling Multicollinearity

Using VIF Score to
eliminate high
correlation variable



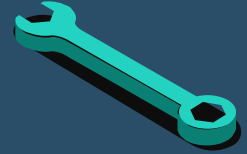
Standardization

Min Max Scaler for
Data with bigger
range



05

Machine Learning



Modelling with Recursive Feature Elimination



| Model | MSE | MAE | R2 |
|-------------------------|-------|-------|-------|
| Linear Regression | 0.491 | 0.329 | 0.758 |
| Random Forest Regressor | 0.113 | 0.076 | 0.987 |
| XGBoost Regressor | 0.106 | 0.072 | 0.988 |

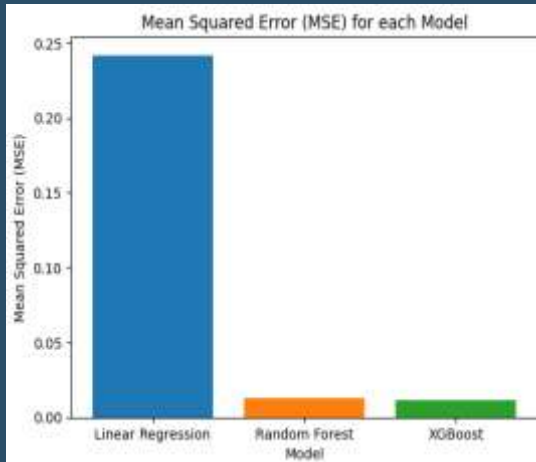
XGBoost Regressor have the best R2 score compared than other models.



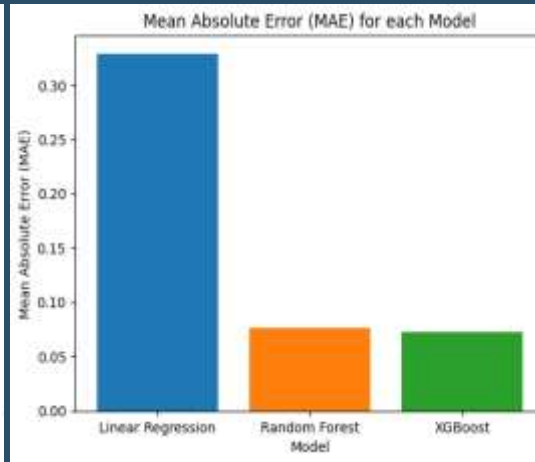
Model Evaluation



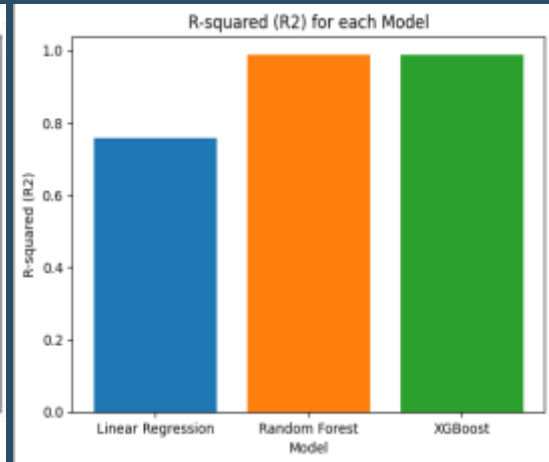
MSE



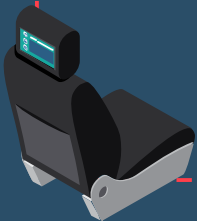
MAE



R2



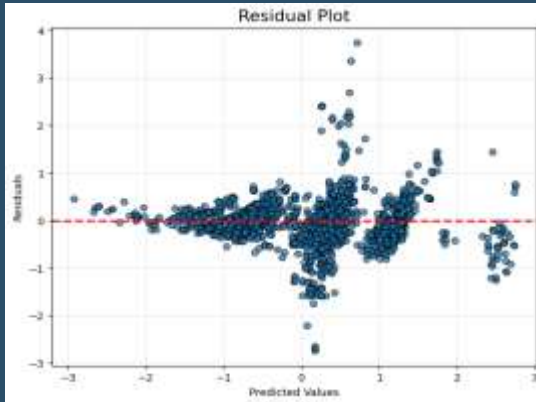
Random Forest memiliki R2 tertinggi dan kesalahan prediksi (MAE dan MSE) yang paling kecil karena kemampuan Random Forest dalam menangani kompleksitas data, mengurangi overfitting, dan memberikan hasil yang stabil serta generalisasi yang baik pada data baru



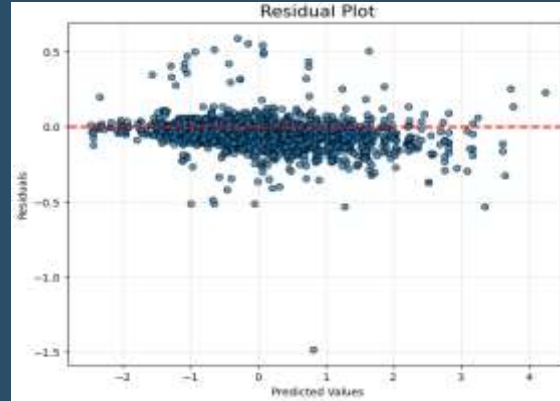
Residual



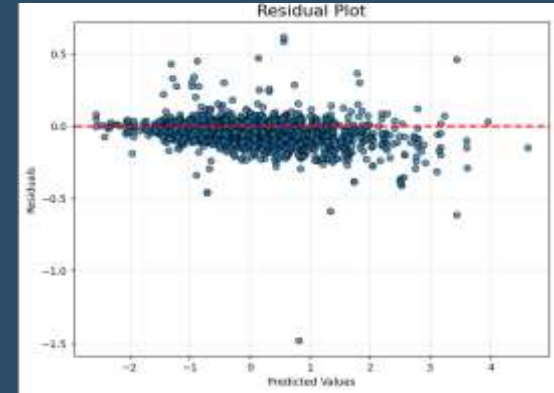
Linear Regression



Random Forest Regressor



XGBoost Regressor



residu (perbedaan antara nilai prediksi dan nilai aktual) menyebar secara merata di sepanjang sumbu $y = 0$

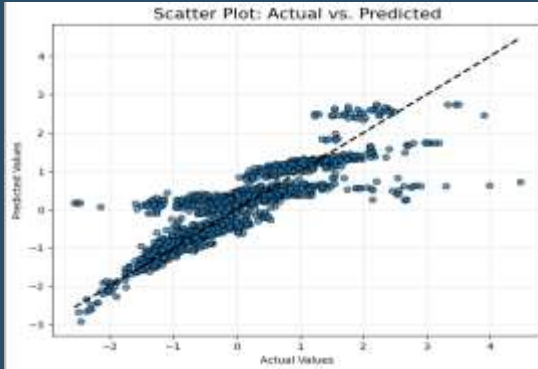
Residu memiliki varian konstan



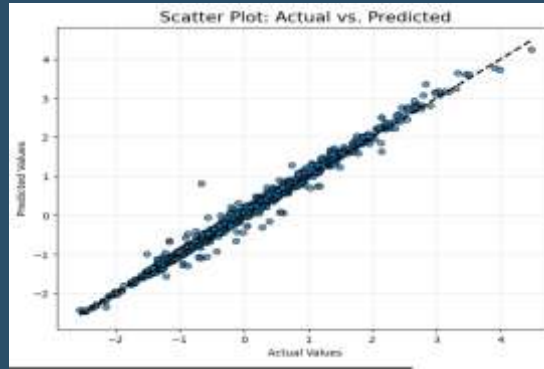
Scatter Plot Prediksi vs Aktual



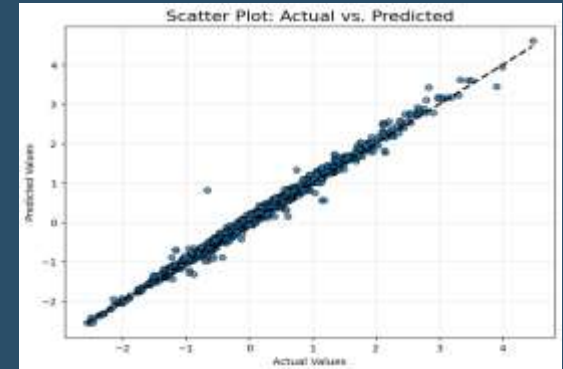
Linear Regression



Random Forest Regressor



XGBoost Regressor



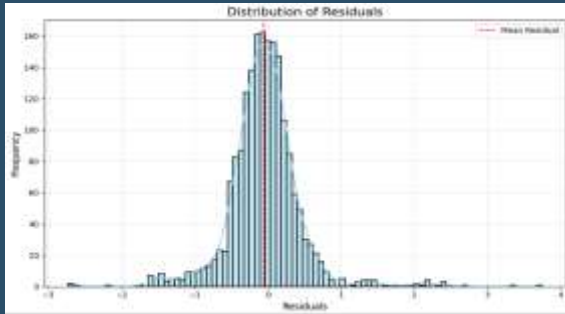
titik-titik pada scatterplot berada di sekitar garis $y = x$.



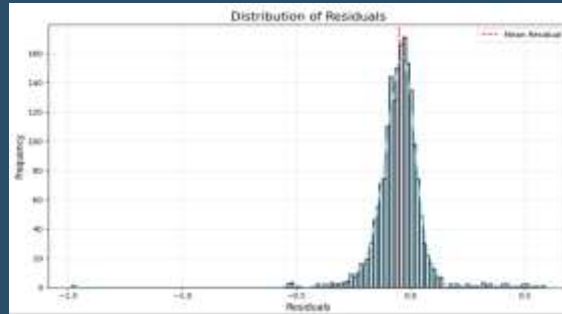
Distribution Residual



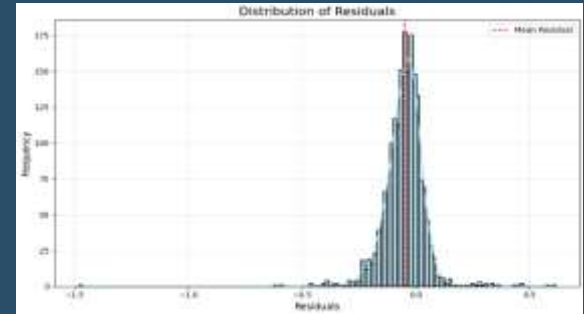
Linear Regression



Random Forest Regressor

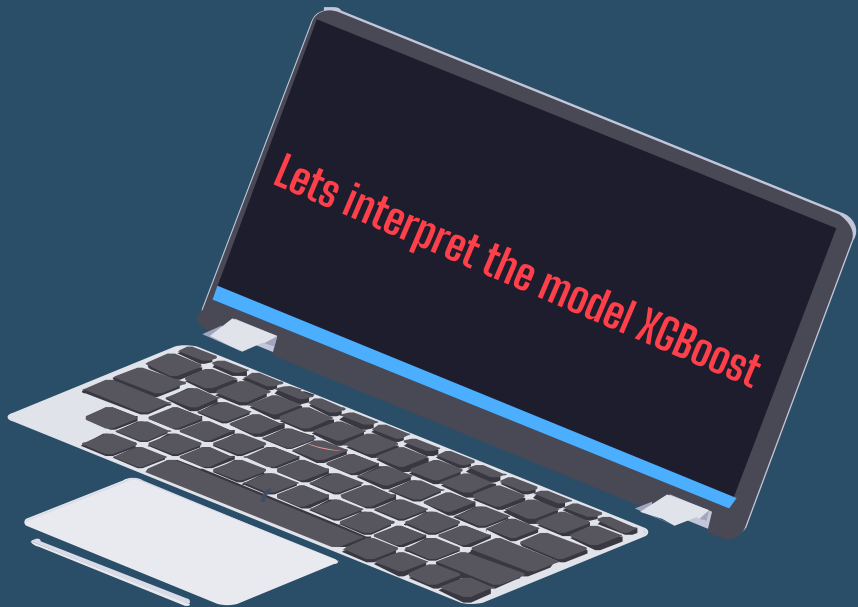
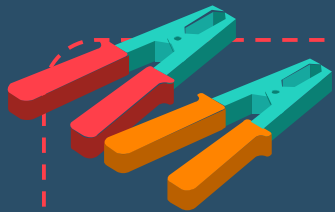


XGBoost Regressor



Residuals are normally distributed

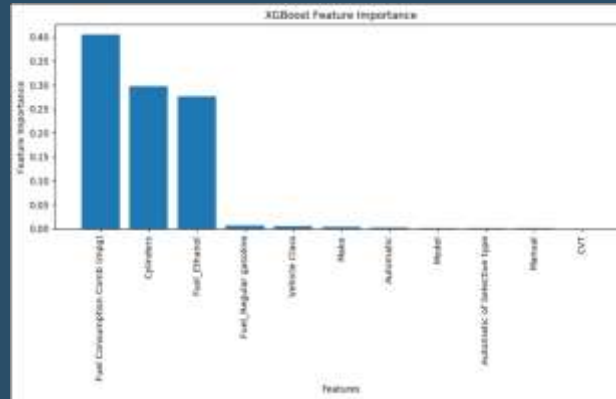




Feature Importance



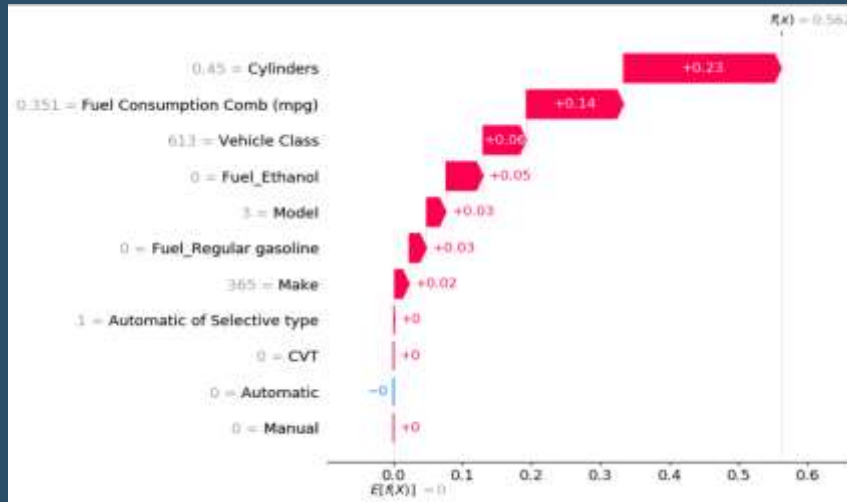
- Feature importance gain mengukur seberapa penting atau berkontribusinya setiap fitur terhadap proses pembelajaran model XGBoost. Gain importance dihitung berdasarkan seberapa sering dan seberapa signifikan suatu fitur digunakan dalam membangun pohon keputusan dalam model XGBoost.



- Fuel Consumption Comb (mpg), Cylinders, Fuel Ethanol memiliki pengaruh yang paling besar terhadap prediksi model. Semakin tinggi nilai fitur tersebut, semakin berpengaruh terhadap hasil prediksi model.
- Fitur-fitur lainnya seperti Fuel Regular gasoline, Vehicle Class, Make, Automatic, Model, Automatic of Selective type, Manual, dan CVT juga memiliki gain importance yang lebih rendah. Meskipun memiliki pengaruh yang lebih kecil, tetap penting untuk mempertimbangkan kontribusi mereka terhadap prediksi model.

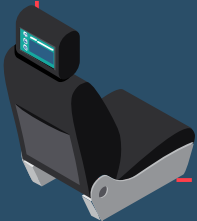


Waterfall Plot



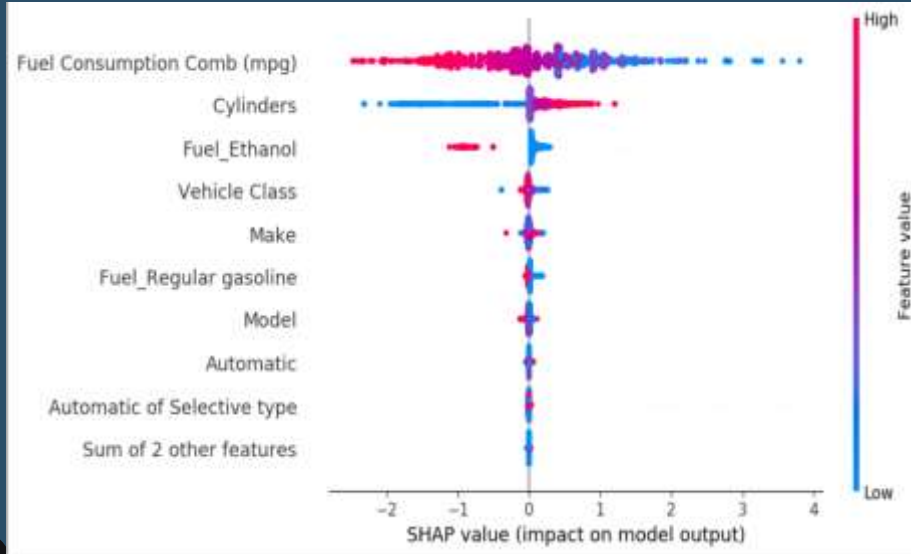
melihat sumbangan masing-masing variabel terhadap hasil prediksi secara terisolasi untuk satu observasi saja. Pada plot tersebut, jika terdapat variabel yang memiliki sumbangan negatif terhadap hasil prediksi, biasanya ini berarti semakin rendah nilai variabel tersebut, maka semakin tinggi nilai prediksi menjadi. Sebaliknya, jika terdapat variabel yang memiliki sumbangan positif, maka semakin tinggi nilai variabel tersebut, semakin tinggi pula nilai prediksi

- Semakin besar nilai fitur, semakin besar pula kemungkinan nilai prediksi akan meningkat.
- Fitur 'CVT', 'Automatic', dan 'Manual' memiliki kontribusi negatif (-0.0, -0.0, dan -0.0, masing-masing). Hal ini menunjukkan bahwa ketiga variabel tersebut memiliki pengaruh kecil atau bahkan tidak berpengaruh terhadap nilai prediksi pada data indeks 0



Beeswarm Plot

Beeswarm plot pada SHAP (SHapley Additive exPlanations) digunakan untuk memvisualisasikan pengaruh positif atau negatif dari setiap fitur terhadap prediksi model secara individu.

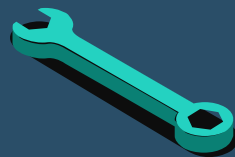


- MPG (Miles Per Gallon) adalah ukuran efisiensi bahan bakar pada kendaraan. Semakin tinggi nilai MPG, semakin sedikit dampak terhadap emisi yang dikeluarkan.
- Jumlah Cylinders pada mesin kendaraan juga memengaruhi emisi CO2. Dengan mengurangi jumlah Cylinders, emisi CO2 yang dihasilkan kendaraan akan berkurang.
- Kandungan Etanol dalam bahan bakar juga berdampak pada emisi. Semakin tinggi kandungan Ethanol, semakin sedikit dampak pada emisi yang dikeluarkan.
- Transmisi Manual (manual transmission) dan CVT (Continuous Variable Transmission) adalah dua jenis transmisi yang umum digunakan pada kendaraan. Dalam hal dampak terhadap emisi, keduanya memiliki pengaruh yang relatif kecil.



05

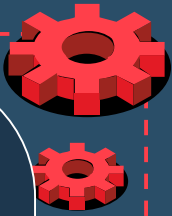
Conclusion Recommendation



Memahami faktor-faktor membantu mengurangi dampak emisi CO₂ dan mendukung upaya menjaga lingkungan yang lebih bersih melalui inovasi kendaraan yang lebih ramah lingkungan.



1. Efisiensi Bahan Bakar dan Emisi CO₂: Terdapat hubungan berbanding terbalik antara konsumsi bahan bakar (mpg) dan emisi CO₂. Kendaraan yang lebih efisien dalam bahan bakar cenderung menghasilkan emisi CO₂ yang lebih rendah.
1. Jumlah Silinder: Semakin banyak silinder kendaraan, semakin tinggi emisi CO₂ yang dihasilkan. Perhatikan jumlah silinder untuk mengurangi emisi CO₂ dan mendukung lingkungan bersih.
1. Perbedaan Merek dan Kelas Kendaraan: Merek dan kelas kendaraan memiliki tingkat emisi CO₂ yang berbeda. Identifikasi merek dan kelas dengan emisi rendah untuk inovasi kendaraan yang ramah lingkungan.
1. Bahan Bakar dan Transmisi: Bahan bakar etanol dan transmisi otomatis terkait dengan emisi CO₂ yang lebih rendah, sedangkan CVT membantu mengurangi emisi. Terapkan bahan bakar dan transmisi ramah lingkungan untuk mengurangi emisi CO₂.





Insight Bisnis

1. Fokus Efisiensi Bahan Bakar: Kunci untuk kurangi emisi CO2, kembangkan teknologi bahan bakar efisien dan kendaraan listrik.
1. Optimal Desain Kendaraan: Perhatikan silinder dan transmisi untuk kurangi emisi CO2, ciptakan produk lebih kompetitif.
1. Investasi Bahan Bakar Ramah Lingkungan: Tingkatkan penggunaan etanol dan CVT pada kendaraan untuk capai target emisi rendah.

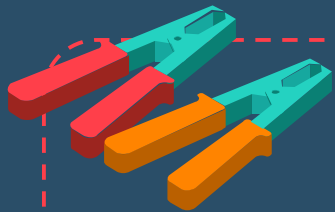
Rekomendasi Bisnis

1. FR&D pada Kendaraan Ramah Lingkungan: Mendorong penelitian dan pengembangan kendaraan ramah lingkungan dengan fokus pada efisiensi bahan bakar dan penggunaan bahan bakar alternatif.
2. Pemilihan Jenis Transmisi: Kendaraan dengan transmisi otomatis atau CVT harus dipilih untuk mengurangi emisi CO2 dan memberikan pilihan yang lebih baik bagi konsumen.
3. Penghargaan untuk Efisiensi: Memberikan insentif dan penghargaan bagi produsen kendaraan dengan emisi rendah dapat mendorong inovasi dalam industri otomotif.

Dampak Bisnis

1. Daya Saing di Pasar Otomotif: Hadirkan kendaraan ramah lingkungan untuk meningkatkan daya saing dan menarik konsumen peduli lingkungan.
1. Kontribusi pada Lingkungan: Kurangi emisi CO2 untuk melawan perubahan iklim dan menciptakan lingkungan lebih bersih.
1. Inovasi dan Teknologi Baru: Fokus pada kendaraan ramah lingkungan mendorong inovasi di industri otomotif.





Thanks!



