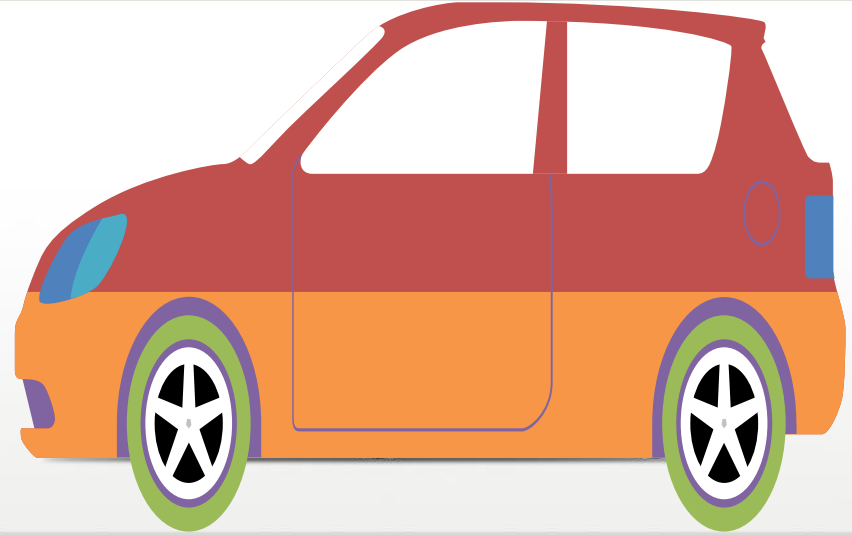


## **CAPSTONE PROJECT MODULE 3**

### **SAUDI ARABIA USED CAR MACHINE LEARNING**

**REFNO DEVIANTO - JCDS ONLINE LEARNING**





# OUTLINE

- 01 Business Problem Understanding
- 02 Data Understanding
- 03 *Data Preprocessing*
- 04 *Modeling*
- 05 *CONCLUSION*
- 06 *RECOMENDATION*

# Business Problem Understanding (1/2)

## Context

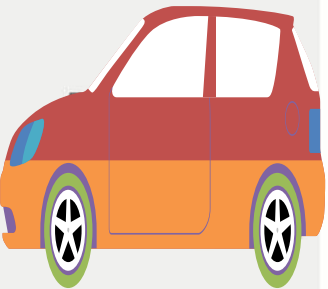
Berdasarkan Analisis Pasar Mobil Bekas Arab Saudi yang didapat dari [Mordor Intelligence](<https://www.mordorintelligence.com/industry-reports/saudi-arabia-used-car-market#:~:text=Saudi%20Arabia%20Used%20Car%20Market%20Analysis,7.36%25%20over%20the%20forecast%20period>). Pasar Mobil Bekas Arab Saudi memiliki nilai sebesar 4,91 miliar dolar AS pada tahun 2021 dan diperkirakan akan mencapai valuasi bersih sebesar 8,69 miliar dolar AS pada akhir tahun 2027, dengan pertumbuhan CAGR yang solid sebesar 7,36% selama periode perkiraan.

Pasar mobil bekas adalah pasar yang tidak menentu. Pasar mobil bekas berfluktuasi berdasarkan sejumlah faktor, baik yang nyata (atau dalam beberapa kasus yang dibayangkan) seperti halnya pasar lainnya. Setelah pasar yang booming, tampaknya Price mobil bekas akan segera turun. Jadi, masalahnya adalah bagaimana cara mendapatkan Price terbaik untuk menjual atau membeli agar kita tidak salah mengambil keputusan. Berdasarkan hal tersebut, kita mungkin harus memprediksi Price mobil bekas pada dataset yang berisi fitur (variabel) di dalamnya.

## Problem Statement

Permasalahan terbesar dari Pasar Mobil Bekas adalah bagaimana cara menjual mobil bekas di Arab Saudi sehingga mereka tidak menjualnya dengan Price yang terlalu tinggi atau terlalu rendah. Price yang terlalu tinggi akan membuat konsumen tidak tertarik untuk membeli mobil tersebut, sedangkan Price yang terlalu rendah akan membuat mereka merugi.

Marketplace pasti menginginkan keuntungan yang besar dengan menjual Price mobil bekas dengan Price yang tetap berdasarkan tipe, jarak tempuh, opsi, dll. Bagaimana cara membuat Price yang kompetitif adalah salah satu masalah untuk mendapatkan keuntungan besar pada bisnis mereka.



# Business Problem Understanding (2/2)

## GOALS

Berdasarkan permasalahan tersebut, marketplace tentu perlu memiliki toll/alat yang dapat memprediksi `Price` mobil bekas untuk menentukan Price terbaik dalam menjual mobil bekas. Hal ini akan menyulitkan para marketplace untuk mengambil keputusan sendiri karena banyaknya jenis, variasi mobil, tipe mobil, dan lain-lain. Dengan Prediksi ini, mungkin akan menjadi alat bantu bagi pemangku kepentingan / pasar untuk membuat keputusan.

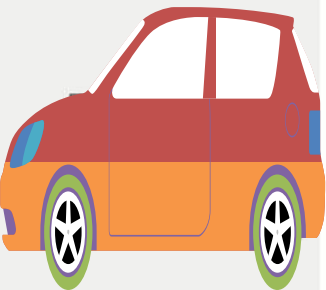
## Analytic Approach

Melalui permasalahan ini, kita akan menganalisa data untuk dapat menemukan pola-pola fitur yang ada yang membedakan satu mobil dengan mobil lainnya.

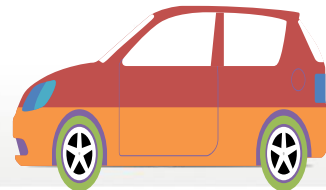
Selanjutnya, kita akan membuat model regresi untuk membantu marketplace sebagai alat untuk membuat keputusan yang lebih baik dan terbaik dalam memprediksi Price mobil bekas. Tentunya kami ingin membuat model yang baik untuk membantu marketplace dan memuaskan pelanggan untuk membeli di channel mereka.

## Metric Evaluation

Evaluasi metrik yang akan digunakan adalah RMSE, MAE, dan MAPE, di mana RMSE adalah nilai rata-rata akar kuadrat dari error, MAE adalah rata-rata nilai absolut dari error, sedangkan MAPE adalah rata-rata persentase error yang dihasilkan oleh model regresi. Semakin kecil nilai RMSE, MAE, dan MAPE yang dihasilkan, berarti model semakin akurat dalam memprediksi Price Mobil Bekas sesuai dengan limitasi fitur yang digunakan.



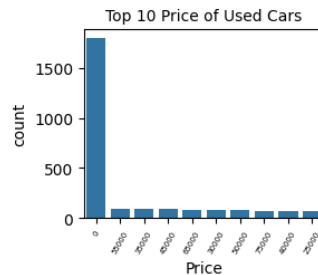
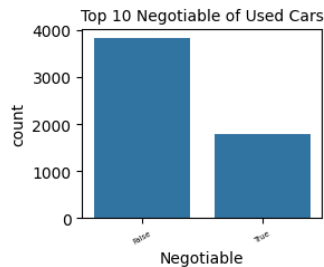
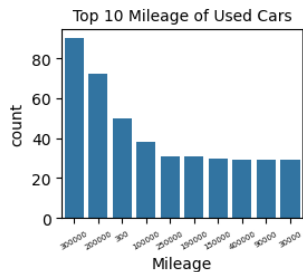
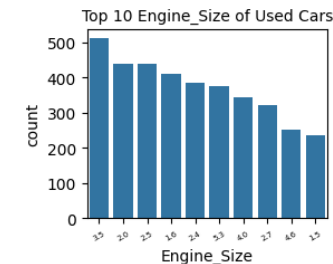
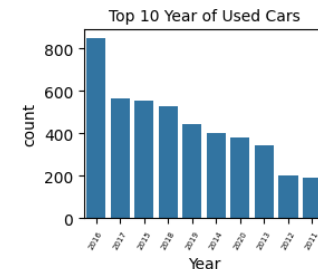
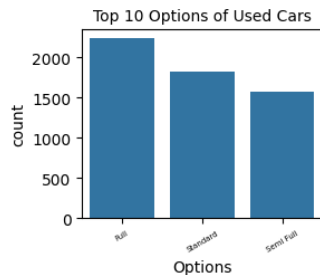
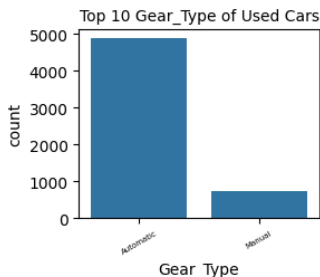
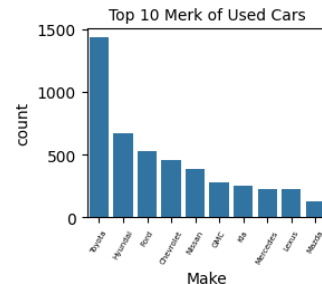
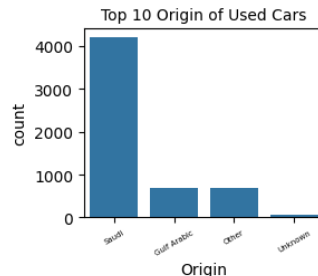
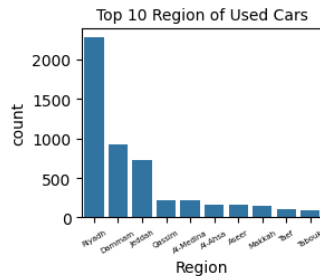
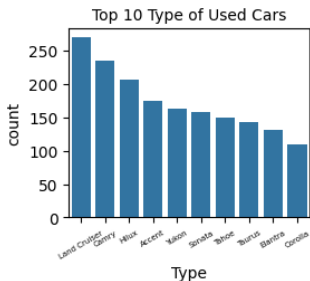
# Data Understanding (1/2)



## Attributes Information

Attribute	Data Type	Description
Type	Object	Jenis Mobil Bekas
Region	Object	Wilayah Tempat Mobil Bekas ditawarkan Untuk di Jual
Make	Object	Nama Perusahaan pembuat Mobil Bekas
Gear_Type	Object	Jenis Transmisi Mobil Bekas
Origin	Object	Asal Mobil Bekas
Option	Object	Pilihan Mobil Bekas
Year	int64	Tahun Pembuatan Mobil Bekas
Engine_Size	Float	Ukuran Mesin Mobil Bekas
Mileage	Integer	Jarak Tempuh Mobil Bekas
Negotiable	Boolean	Status negosiasi mobil bekas, Benar jika harganya 0, artinya masih bisa dinegosiasikan
Price	Integer	Harga Mobil Bekas (dalam Riyal)

# Data Understanding (2/2)



# Data Preprocessing (1/3)

Check Dataset



Missing Value  
&  
Data Duplicate



Feature Selection  
(Drop Column)



Check & Drop  
Outliers



Check Data  
Distribution



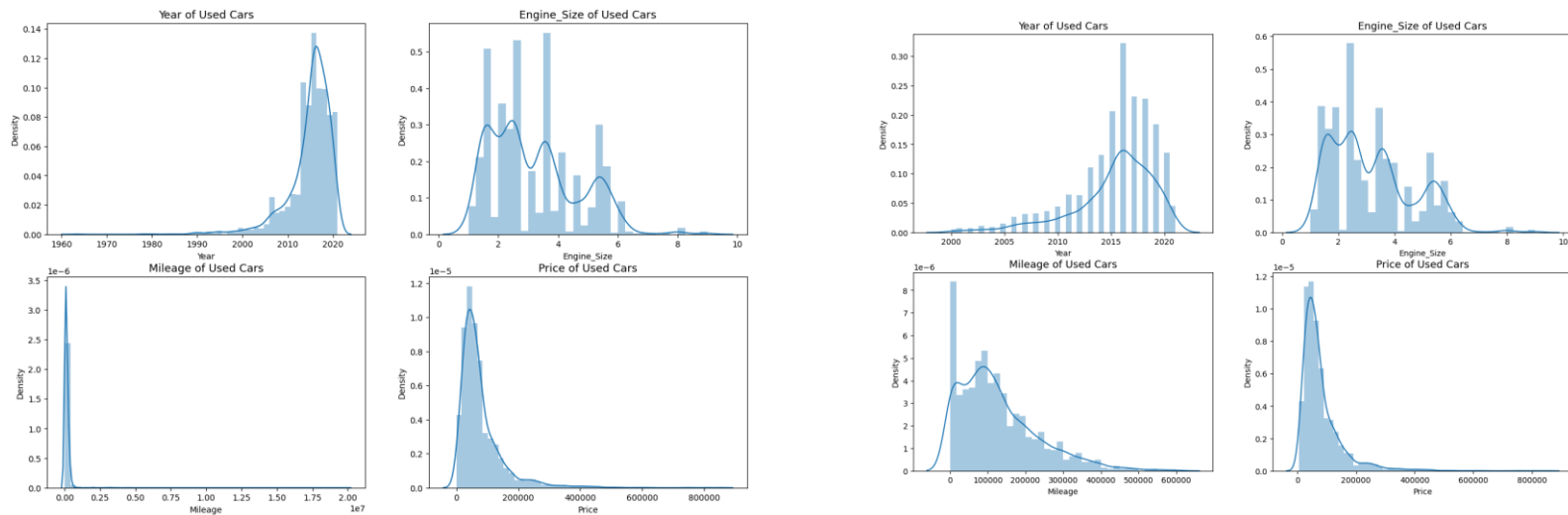
Checking Data  
Correlation



# Data Preprocessing (2/3)



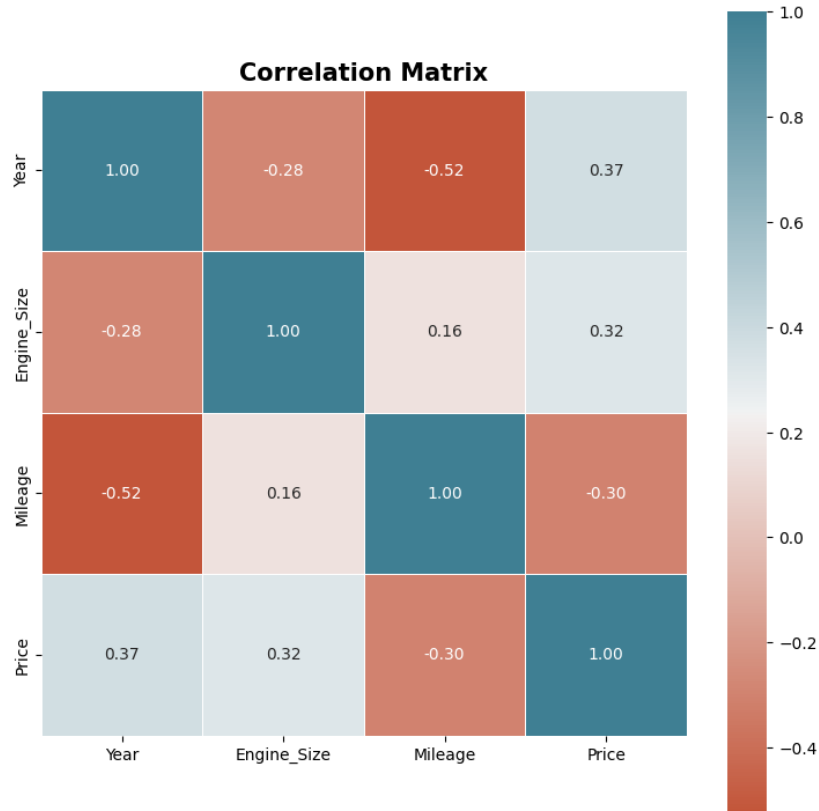
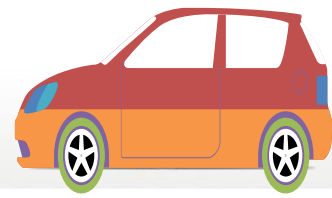
## DATA OUTLIERS



***Before Cleaning***



# Data Preprocessing (3/3)



## Insight:

- Bisa kita lihat pada Correlation Matrix Heatmap di atas, terdapat high correlation antara Year dan Mileage dimana Mileage akan meningkat dari tahun ke tahun sesuai pemakaian kendaraan.
- Kita juga bisa lihat Price bergantung pada Year dan Engine Size.



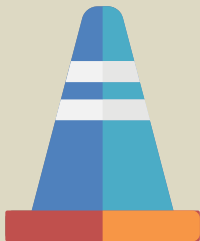
### Encoding

One-Hot Encoding:

- Gear\_Type,
- Origin, dan
- Options.

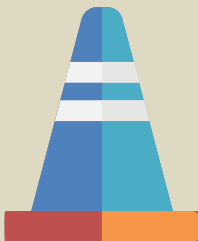
Binary Encoding:

- Type,
- Region, dan
- Make.



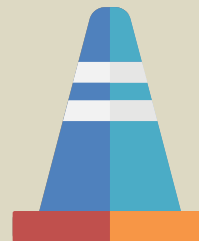
### Train and Splitting

Splitting data into training and test with propotion 70:30

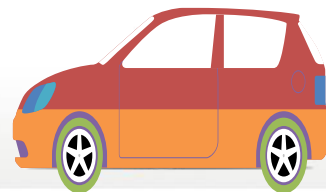


### Modeling

- Linear Regression
- KNN Regressor
- Decision Tree Regressor
- RandomForest Regressor
- XGBoost Regressor

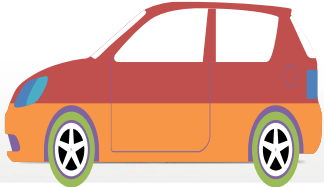


# Test Result



	Model	Mean_RMSE	Std_RMSE	Mean_MAE	Std_MAE	Mean_MAPE	Std_MAPE
0	Linear Regression	-48200.9	8074.9	-25770.9	2677.3	-0.3	0.0
1	KNN Regressor	-43083.4	7789.5	-23760.4	1692.0	-0.4	0.0
2	DecisionTree Regressor	-46633.8	7947.6	-23115.0	2667.3	-0.3	0.0
3	RandomForest Regressor	-35129.1	6639.9	-16825.1	1454.9	-0.2	0.0
4	XGBoost Regressor	-30946.1	3052.1	-15740.6	926.5	-0.2	0.0

# Hyperparameter Tuning



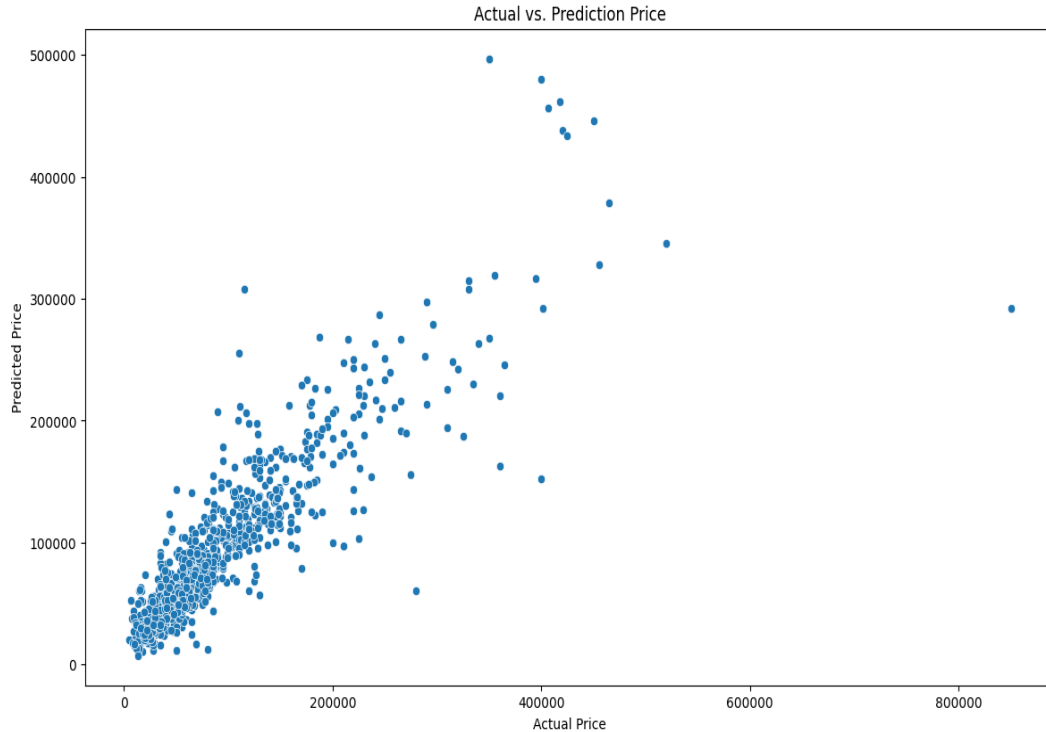
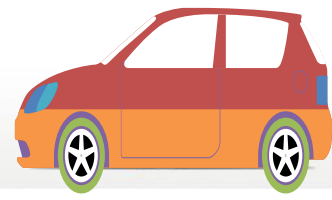
	RMSE	MAE	MAPE
XGB	37913.7	19293.7	0.3

BEFORE

	RMSE	MAE	MAPE
XGB	34743.5	18176.0	0.3

AFTER

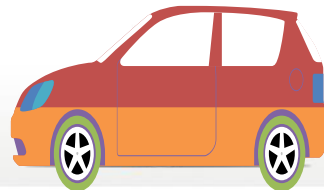
# Analysis of Model Prediction Results



## INSIGHT

- Model yang dihasilkan tergolong cukup baik dan linear, dimana terlihat dari grafik di atas serta menghasilkan nilai small error metric values, (RMSE, MAE, MAPE).
- Beberapa nilai actual price yang rendah di atas jika dibandingkan dengan nilai predicted price mengalami error, hal ini disebabkan karena banyaknya data dengan nilai **\*\*Price\*\*** yang rendah.
- Ketika Price more than 200000, distribusi plot mulai irregular. Kita dapat melihat bahwa kadang-kadang kita mempunyai harga yang diprediksi tinggi dan kadang-kadang rendah setelah 200000.

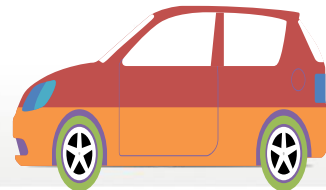
# Conclusion



Dalam proyek ini, kita telah membangun sebuah model untuk memprediksi harga mobil bekas dengan beberapa kriteria yang telah ditentukan di atas, dimana model ini dapat menjadi informasi pendukung yang berguna dalam menentukan harga jual dan beli mobil bekas bagi perusahaan maupun individu yang melakukan transaksi.

- Model terbaik yang digunakan disini adalah XGBoost Regressor.
- Hasil dari model menunjukkan bahwa fitur-fitur yang paling signifikan pengaruhnya adalah Make, Year dan Engine Size.
- Performa model regresi dievaluasi menggunakan metrik RMSE (Root Mean Square Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error) dan R-Squared . Setelah melalui proses hyperparameter tuning, model yang dihasilkan tergolong cukup baik dan linear (XGBoost) dimana menghasilkan nilai small error metric values,(RMSE, MAE, MAPE) dan nilai R2 hampir mendekati 1.
- Nilai RMSE memiliki makna bahwa ketika model digunakan untuk memprediksi harga mobil bekas, perkiraan harga rata-ratanya dapat memiliki selisih sekitar 34.743,5 Riyal (RMSE) atau 18.176 (MAE) dari harga actual. Sedangkan nilai MAPE yang dihasilkan 0.3 yang menunjukkan error absolut pada Price yang diprediksi oleh model. Semakin kecil nilai MAPE berarti nilai taksiran semakin mendekati nilai sebenarnya.
- Ketika Price more than 200000, distribusi plot mulai irregular. Kita dapat melihat bahwa kadang-kadang kita mempunyai harga yang diprediksi tinggi dan kadang-kadang rendah setelah 200000.
- Model yang diperoleh masih memiliki potensi untuk ditingkatkan melalui proses-proses tertentu. Namun, untuk saat ini, kami berasumsi bahwa model sudah mencapai hasil yang diharapkan. Selain itu, dalam proses pembuatan model, diperlukan pengetahuan yang mendalam mengenai industri mobil untuk dapat mengembangkan pengembangan model yang lebih baik lagi.

# Recommendations



Berikut ini adalah beberapa rekomendasi yang dapat dilakukan untuk meningkatkan performa model:

1. Menambahkan fitur yang mengkategorikan jenis mobil menjadi classic atau non-classic. Hal ini penting karena harga mobil bekas terhadap mileage dan tahun pembuatan dapat berbeda secara signifikan antara mobil-mobil classic dan mobil-mobil biasa. Dengan memasukkan fitur ini ke dalam model, kemungkinan besar akan meningkatkan akurasi prediksi harga mobil bekas.
2. Melakukan analisis terhadap nilai error tertinggi yang dihasilkan oleh model, dengan mengelompokkan error menjadi 3 kategori, yaitu overestimation (5%), underestimation (5%), dan mayoritas yang memiliki error mendekati nilai mean (90%). Setelah itu, dilakukan pemeriksaan hubungan antara error tersebut dengan setiap variabel independen. Dengan demikian, dapat dilakukan proses training ulang dan menghindari variabel yang menyebabkan error tinggi.
3. Jika tersedia tambahan data yang signifikan, dapat mencoba menggunakan model yang lebih kompleks seperti recursive neural networks (RNN). Namun, perlu diperhatikan bahwa jika jumlah data dan fitur masih sebatas dataset yang ada saat ini, kemungkinan peningkatan performa model secara signifikan dengan menggunakan model yang lebih kompleks tidak terlalu besar.

**Thanks!**