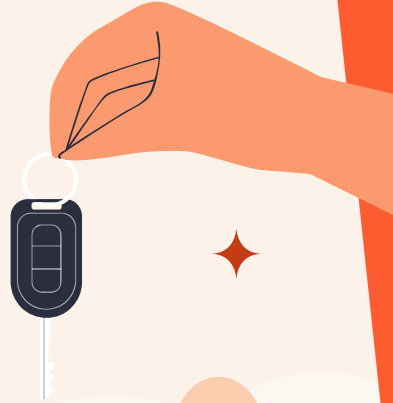
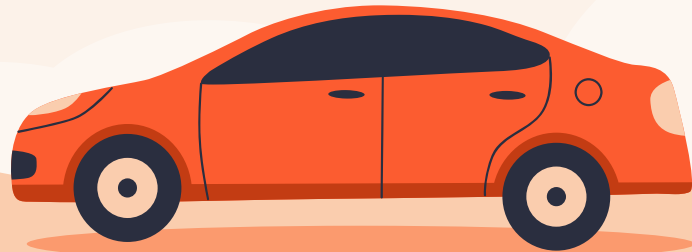


# Saudi Arabia Used Cars

**Capstone Module 3**

Ega Adiwena - JCDSOL-16



# What will we discuss?

**01**

**Background**

**02**

**Data  
Understanding**

**03**

**Data  
Preprocessing**

**04**

**Modeling**

**05**

**Conclusion**

**06**

**Recommendation**



An illustration featuring two hands. On the left, a light-skinned hand holds a dark blue car key. On the right, a dark-skinned hand is reaching out. In the center, a white square box with an orange border contains the number '01' in orange. The background is a light beige color with stylized white clouds, small orange stars, and a landscape at the bottom with rolling hills and two orange trees.

01

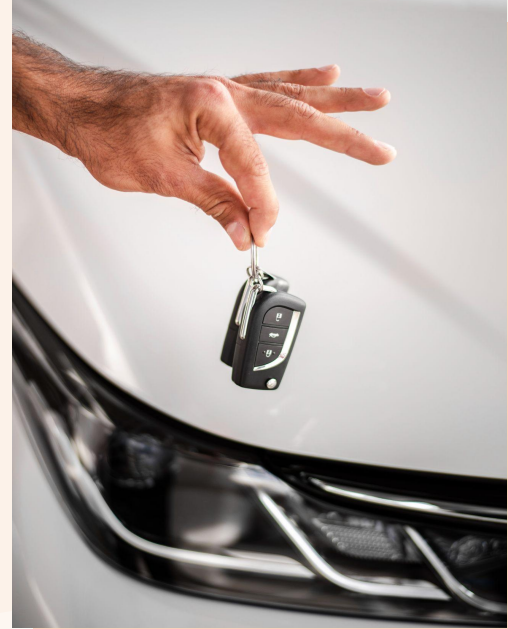
**Background**

# Before we begin..

Pasar mobil bekas di Arab Saudi sedang mengalami pertumbuhan yang signifikan. Diperkirakan bernilai **USD 6,41 miliar pada tahun 2024**, pasar ini diproyeksikan mencapai **USD 9,03 miliar pada tahun 2029**, dengan **CAGR lebih dari 7,10%** selama periode perkiraan.

**Digitalisasi** yang pesat di Arab Saudi, dengan penetrasi internet mencapai **98,6% populasi**, telah mengubah lanskap pasar mobil bekas.. Transaksi peer-to-peer dan lelang online semakin populer, menyederhanakan proses jual beli.

<https://www.mordorintelligence.com/industry-reports/saudi-arabia-use-d-car-market>



# So here we are now..



## Problem Statement

Bagaimana cara melakukan valuasi yang akurat terhadap mobil bekas agar mendapatkan **harga yang wajar** dan tidak merugikan bagi pembeli maupun penjual?



## Objective

Diciptakan sebuah perangkat yang dapat **memberikan prediksi harga yang tepat**. Alat ini diharapkan dapat menjadi referensi bagi pembeli maupun penjual dalam proses transaksi.



# Meet the stakeholders



**Data Engineer Team**



**Data Analyst and ML Team**



**Software Engineer Team**



**Sales and Marketing Team**

# Evaluation Matrix

## MAE

**Mean Absolute Error**  
Melihat kesalahan absolut  
dari model



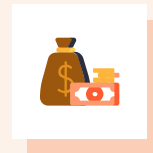
## R-Squared

**R-Squared**  
Melihat generalisasi  
model



## RMSE

**Root Mean Squared Error**  
Melihat kesalahan prediksi  
model



An illustration featuring two hands, one light orange and one brown, exchanging a dark blue car key. In the center, a white square box with an orange border contains the number '02' in orange. The background is a light beige color with stylized white clouds, small orange four-pointed stars, and a landscape at the bottom with orange hills and two stylized trees. The entire scene is framed by vertical orange bars on the left and right sides.

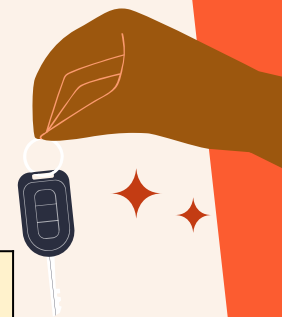
02

# Data Understanding



# Data Information

Features	Data Type	Description
Engine_Size	Float	The engine size of used car
Gear_Type	Text	Gear type size of used car (automatic / manual)
Make	Text	The company name
Mileage	Int	Mileage of used car
Negotiable	Bool	True if the price is 0, that means it is negotiable
Option	Text	Options of used car (full options / semi-full / standard)
Origin	Text	Origin of used car (Gulf / Saudi / other)
Price	Int	Used car price (in SAR)
Region	Text	The region in which the used car was offered for sale
Type	Text	Type of used car
Year	Int	Manufacturing year



An illustration featuring two hands, one light orange and one brown, exchanging a dark blue car key. In the center, a white square box with an orange border contains the orange number '03'. The background is a light orange gradient with stylized white clouds, small orange stars, and a landscape with rolling hills and two orange trees at the bottom.

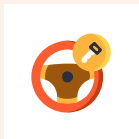
03

# Data Preprocessing

# The data preprocessing steps involve..

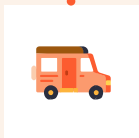
## Step 1

Column adjustment



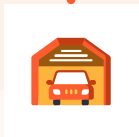
## Step 2

Missing value and duplicate checking



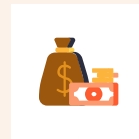
## Step 3

Drop column



## Step 6

Data correlation checking



## Step 5

Data distribution checking

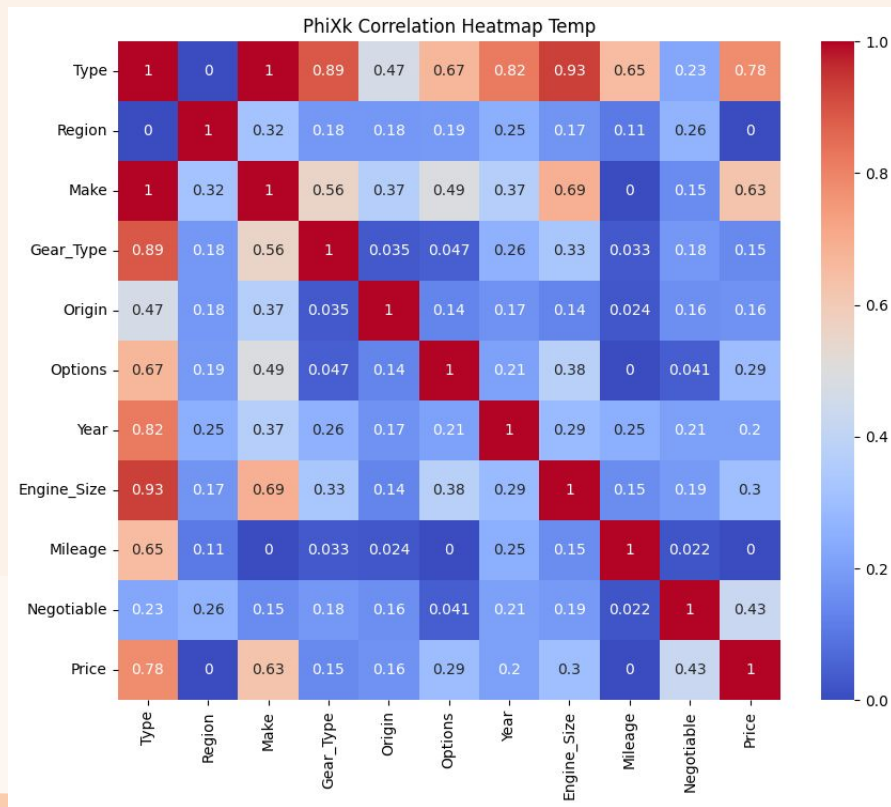


## Step 4

Outliers checking



# Let's see how the data connects..





04

# Data Modeling

# Let's dive into the modeling process..

**Modelling**

**Encoding**

**Train and  
Splitting**



## Train and Splitting

Split data into train with testing proportion of 70 : 30



## Encoding

One Hot (Gear\_type, Origin, Options)  
Binary (Type, Region, Make)



## Modeling

Linear Regression, SVR, KNN, Decision Tree, Random Forest, Gradient Boost, XG Boost, LGBM Regressor



# Our Modeling Result

No	Model	Mean_MSE	Mean_RMSE	Mean_MAE	Mean_R-Squared
1	Linear Regression	2,787,581,957	52,473	32,161	0.462
2	SVR Regressor	5,440,077,304	73,481	45,168	-0.058
3	KNR Regressor	1,762,044,610	41,562	22,701	0.661
4	Decision Tree Regressor	2,150,918,105	46,331	17,643	0.566
5	Random Forest Regressor	1,132,243,451	33,425	13,980	0.778

# Our Other Modeling Result

Using model boosting

No	Type	RMSE	MAE	R-Squared
1	Gradient	35,786	19,022	0.709
2	XGB	30,001	14,004	0.795
3	LGBM	29,339	13,771	0.804



# Let's tune it up!

Using hyperparameter tuning

Model LGBM	RMSE	MAE	R-Squared
GridSearchCV	30,644	14,640	0.786
RandomizedCV	31,155	15,209	0.779

Setelah dibandingkan dengan “based” model LGBM, ternyata hasil dari **“based” model LGBM** lebih bagus dibandingkan dengan hasil setelah Hyperparameter Tuning

An illustration of a light-skinned hand holding a dark blue car key with a silver blade. The hand is positioned on the left side of the frame.

05

An illustration of a dark-skinned hand reaching out from the right side of the frame, palm facing up.

**Conclusion**

# To conclude..



Performa terbaik pada test set dicapai oleh **Base Model LightGBM**, dengan nilai RMSE sebesar 29,339, MAE sebesar 13,771, dan  $R^2$  sebesar 0.804.



Model yang kami kembangkan memiliki tingkat akurasi yang cukup baik, dengan **rata-rata kesalahan prediksi sekitar 29,339 satuan (RSME)**. Model ini **mampu menjelaskan 80,4% variasi data (R-Squared)**.



Meskipun performa test set Base Model lebih baik dibandingkan GridSearchCV dan RandomizedSearchCV, **terdapat perbedaan cukup besar antara performa training ( $R^2$ : 0.925) dan test set ( $R^2$ : 0.804).**

Hal ini mengindikasikan model mungkin sedikit **overfitting** terhadap data training.



An illustration of a light-skinned hand holding a dark blue car key with a silver blade. The background is a warm, orange-toned landscape with stylized clouds, trees, and a central white square containing the number 06.

06

An illustration of a dark-skinned hand reaching out towards the left, palm facing up. The background is a warm, orange-toned landscape with stylized clouds, trees, and a central white square containing the number 06.

# Recommendations

# What we can improve..



**Penambahan  
Fitur Relevan**



**Analisis  
Residual Error**



**Cross-Validation  
yang Lebih  
Mendalam**



**Eksplorasi  
Model Lain**



**Optimisasi  
Hyperparameter**

# Thanks!

**Do you have any questions?**

egadiwena@gmail.com

+62 8111 0811 40

**CREDITS:** This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

