



Used Car Price Model Machine Learning Prediction — DatS Your Future



Our Team



Joshua Abisha
Group Leader



Muhammad Faisal
Group Member



Daniel Jeffry
Group Member



Muhammad Arif
Group Member



Tri Utomo
Group Member



Queenta Rahatesa
Group Member

Table of Content

- 01.** Problem Statement
- 02.** Insights
- 03.** Data Pre-processing
- 04.** Modelling
- 05.** Recommendation & Impact

Introduction

Kami merupakan tim data scientist dalam suatu dealer jual beli mobil bekas secara online maupun offline yang menganalisa penjualan mobil bekas di Amerika dan bertanggung jawab untuk memberikan rekomendasi harga mobil bekas berdasarkan data histori pembelian 2 tahun sebelumnya yang ada di Amerika yang diasumsikan ideal karena mobil berhasil dibeli dan dijual kembali ke konsumen dengan harga ideal.



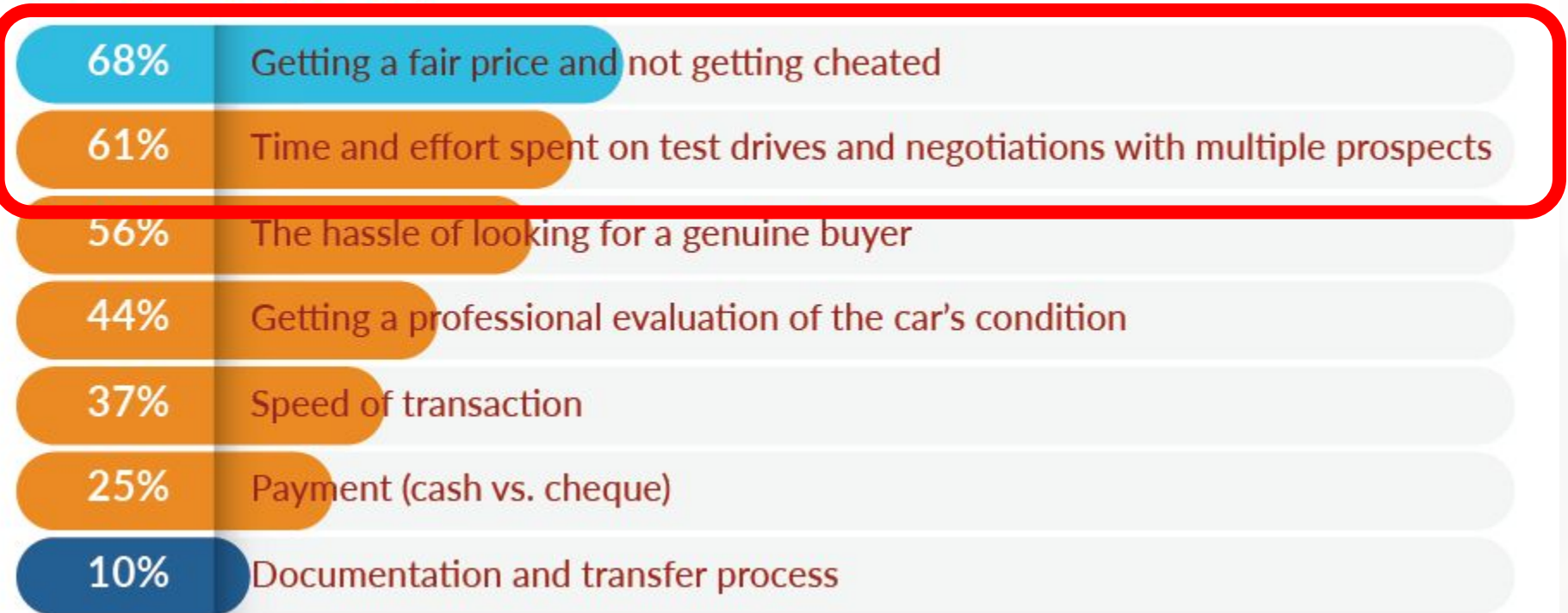
Problem Statement

Understanding The Problem

Price and convenience key



Key concerns (Multiple-responses)



The Problem

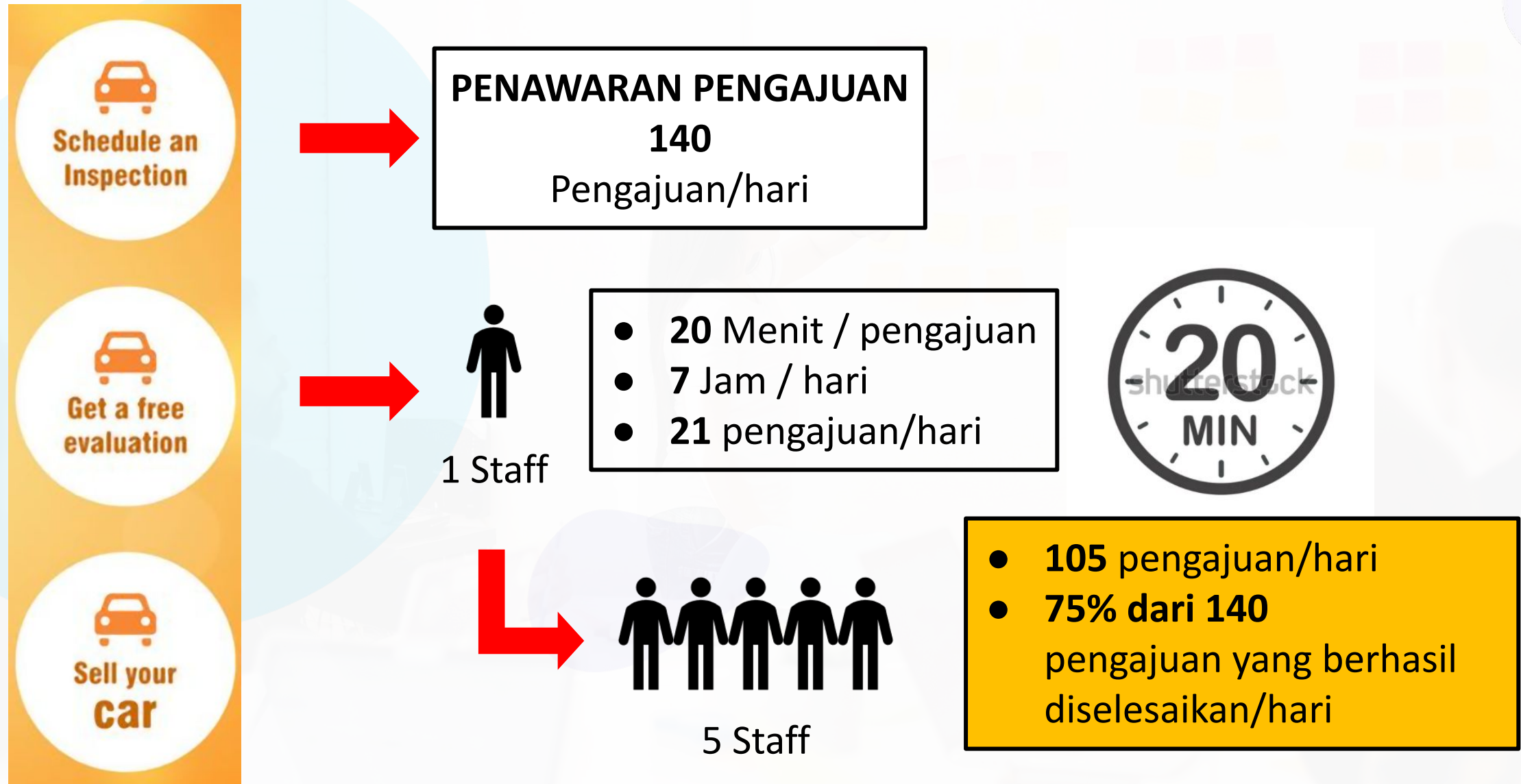
1. Conventional Business Process Consumes Too Much Time

Proses screening dan negosiasi yang cukup lama akhirnya customer menjadi ragu dan cenderung tidak menjual mobil nya ke perusahaan kita.

2. Lack of Trust from Customers due to Staff's Subjectivity

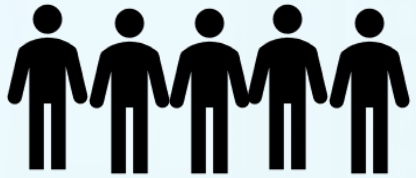
Perusahaan menentukan harga beli mobil bekas tidak terstandar (tidak sesuai kondisi mobil sebenarnya) sehingga mendapat harga mobil bekas yang berbeda dari perspektif yang bias oleh setiap staff.

1. Conventional Business Process Consumes Too Much Time

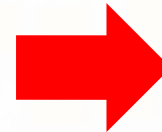
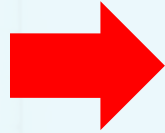


2. Lack of Trust from Customers due to Staff's Subjectivity

DECISION MAKING BIASES



5 Staff



Staff A

\$23.000

(overpricing??)

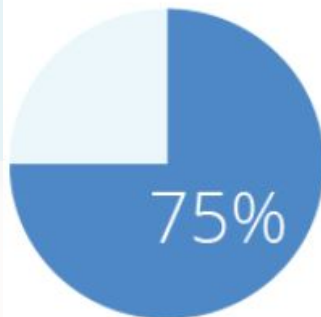


Staff B

\$18.000

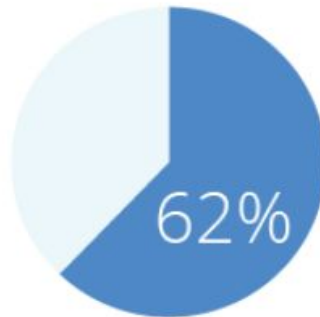
(underpricing??)

DO YOU BELIEVE YOU GET FAIR VALUE
ON TRADE-INS AT THE DEALERSHIP?



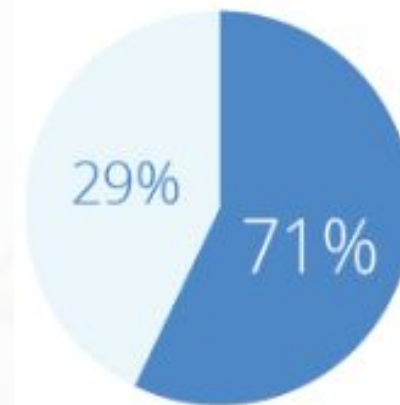
Consumers

■ - NO
□ - YES



Franchise Dealer
Perception

DO YOU LIKE TO NEGOTIATE OR BE
GIVEN THE "BEST PRICE" UP FRONT?



■ - Best price
□ - Negotiate

Understanding The Problem

Goal:

Bisa membantu perusahaan menentukan harga beli mobil bekas yang dapat terstandarkan berdasarkan model machine learning.

Objective:

Membuat tools / model machine learning untuk membantu memprediksi harga mobil berdasarkan kondisi sebenarnya.

Business Metrics:

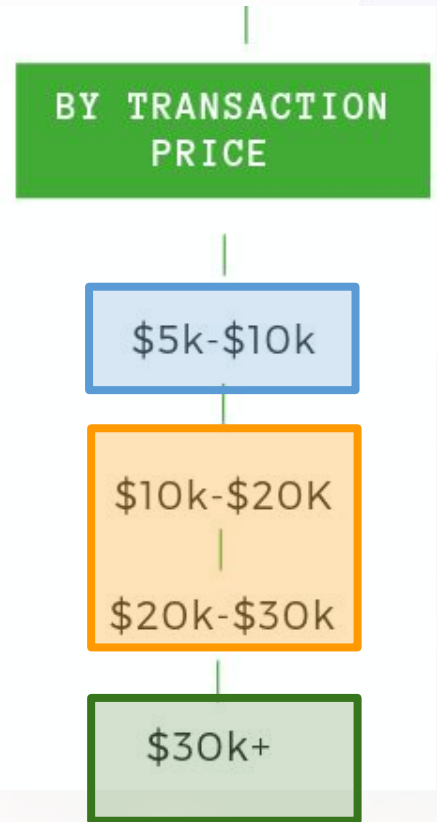
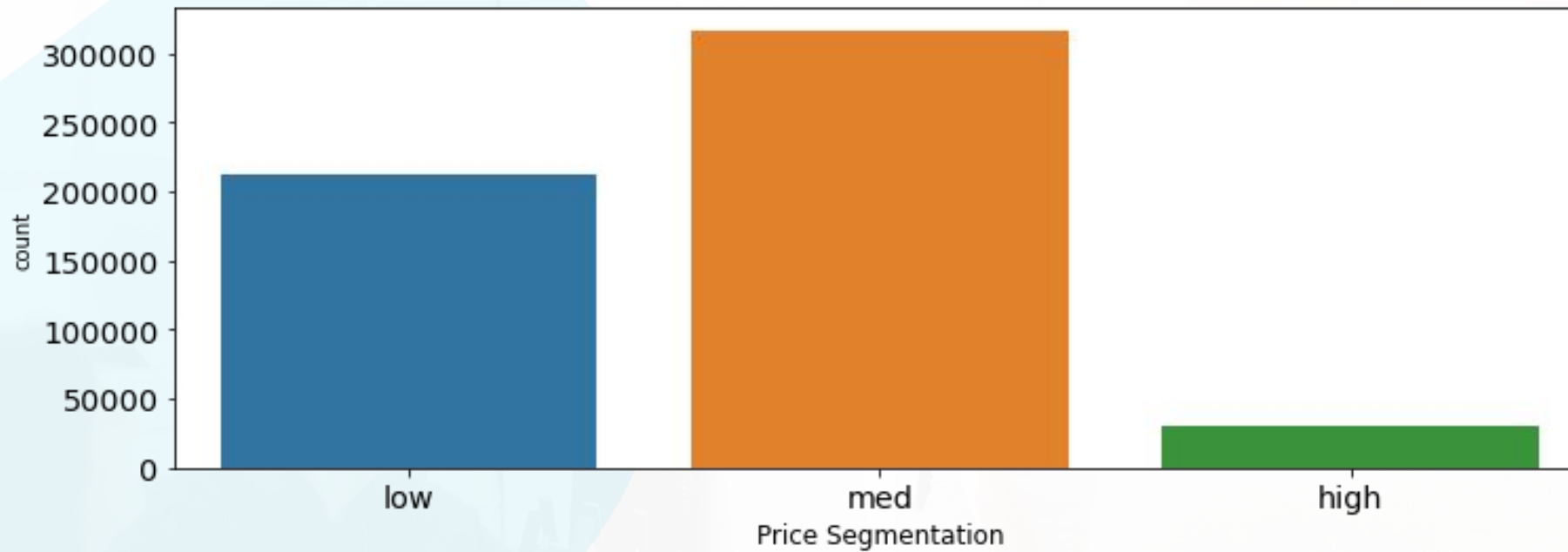
Sales Pembelian, Efisiensi Waktu



Insights

Insights

DATA VISUALISASI (Price Segmentation)



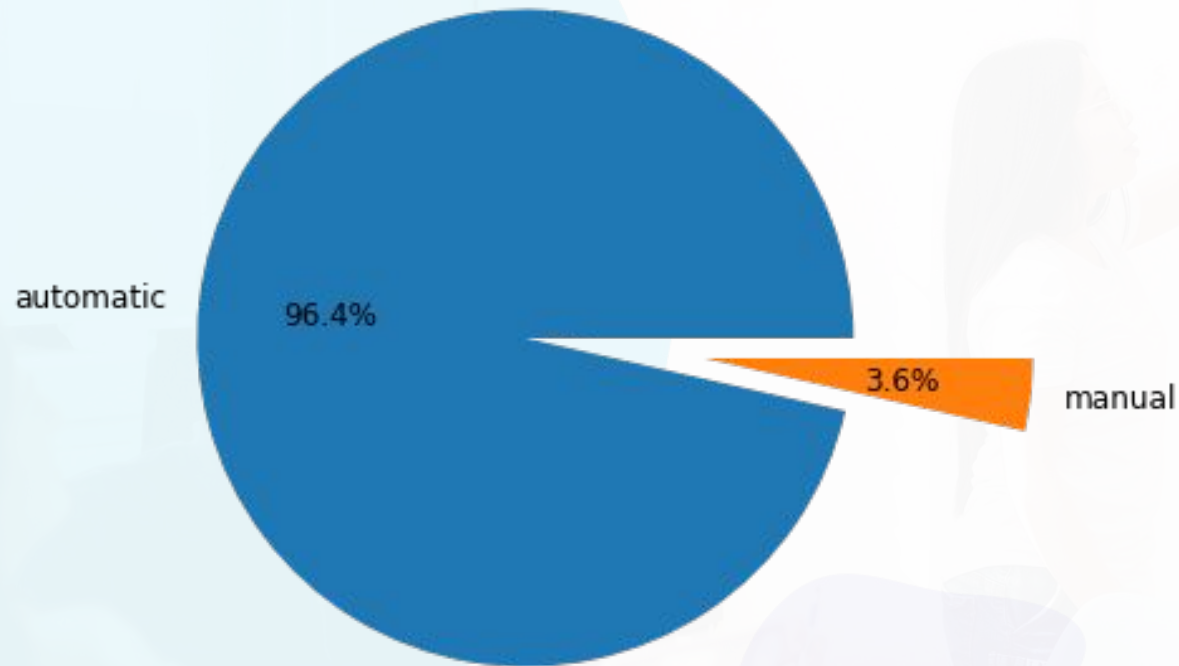
> 300.000 cars in medium segmentation

Source: https://res.cloudinary.com/dkrav2qn2/images/w_1024,h_768/f_webp,q_auto/v1621681071/Used-car-market-in-US-Market-segmentation/Used-car-market-in-US-Market-segmentation-png?_i=AA

Insights

DATA VISUALISASI (Car Transmission)

Car Transmission



96,4% Automatic

END OF THE ROAD FOR MANUAL TRANSMISSIONS?

2007

29% OF 2,391 VEHICLES SOLD IN THE U.S. ARE OFFERED WITH A MANUAL TRANSMISSION

29%



= 693 MODELS

= 50 CARS

2012

19% OF 2,360 VEHICLES SOLD IN THE U.S. ARE OFFERED WITH A MANUAL TRANSMISSION

19%

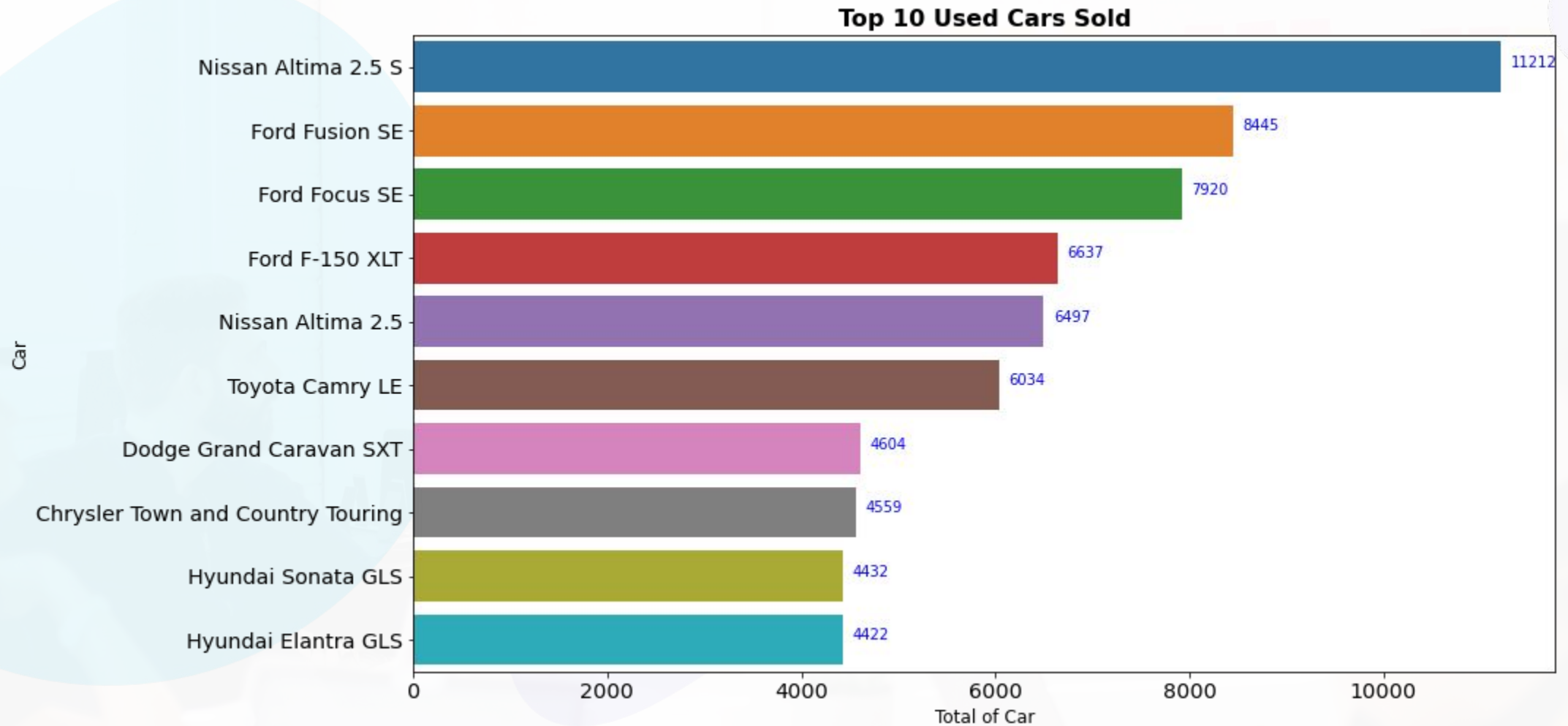


= 448 MODELS

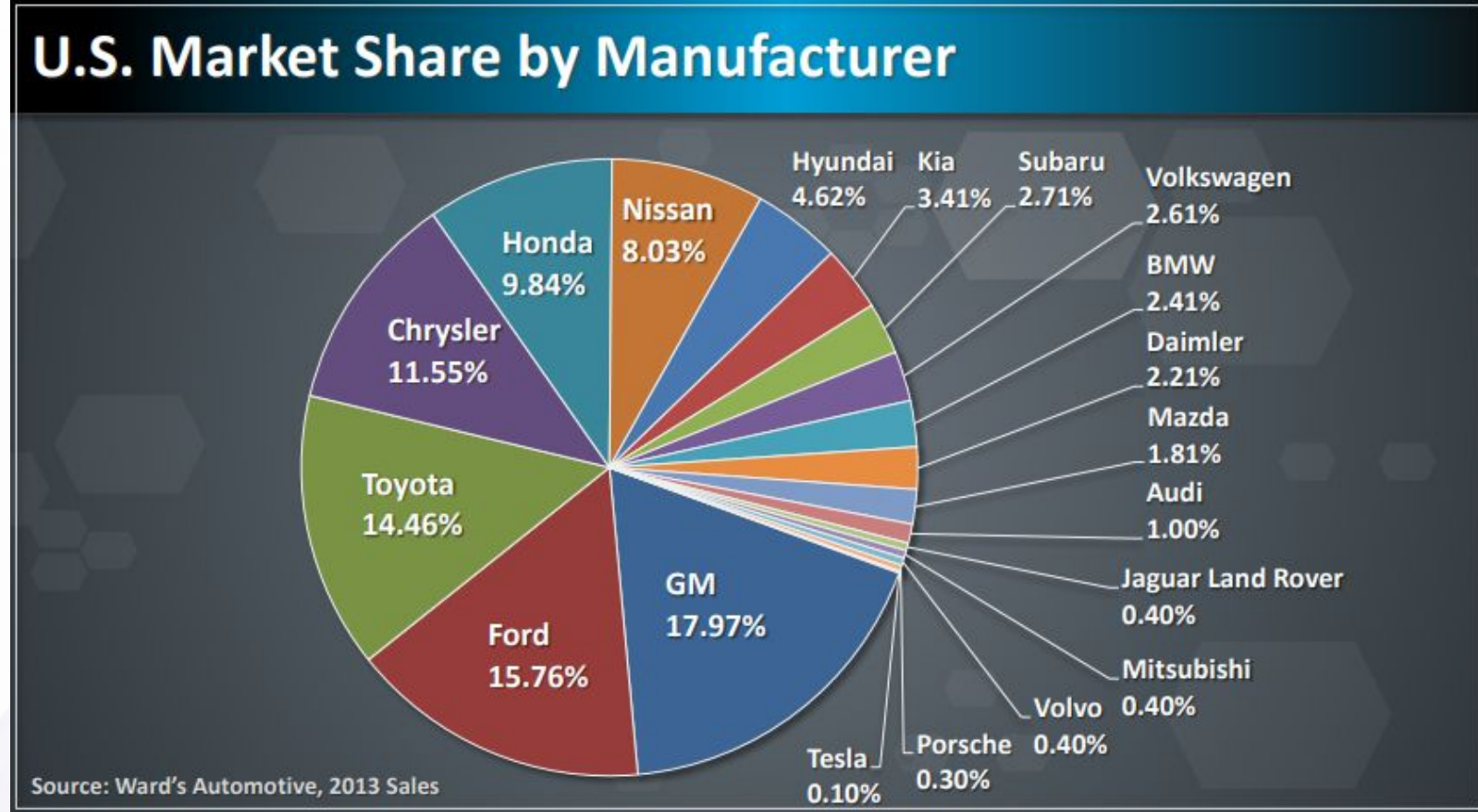
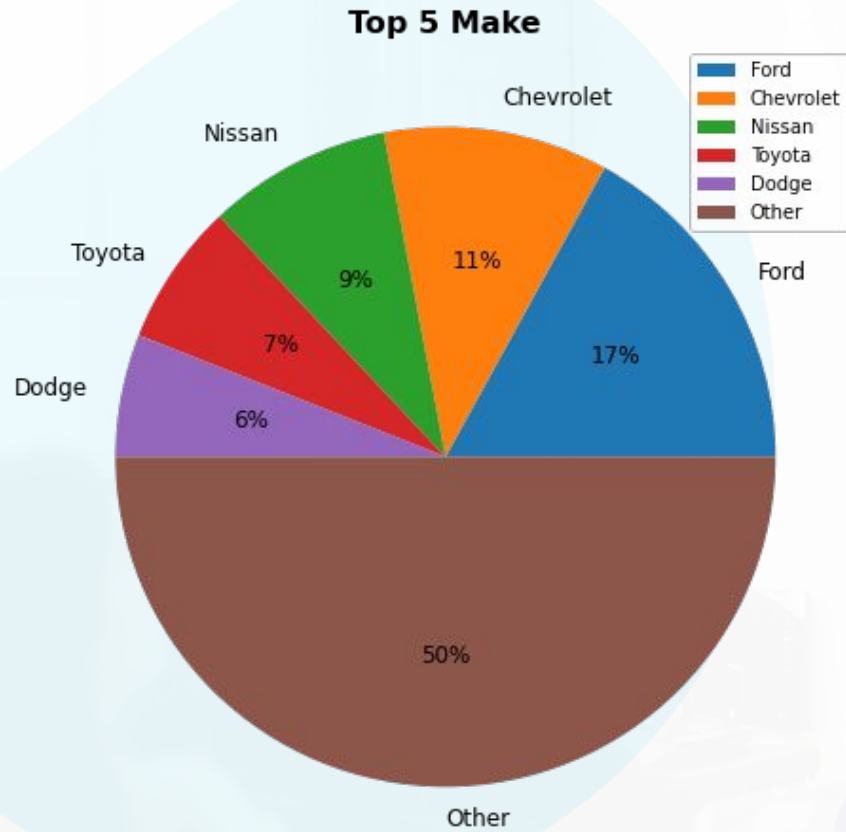
= 50 CARS

Source: <https://www.fix.com/blog/manual-vs-automatic-transmissions/>

Insights

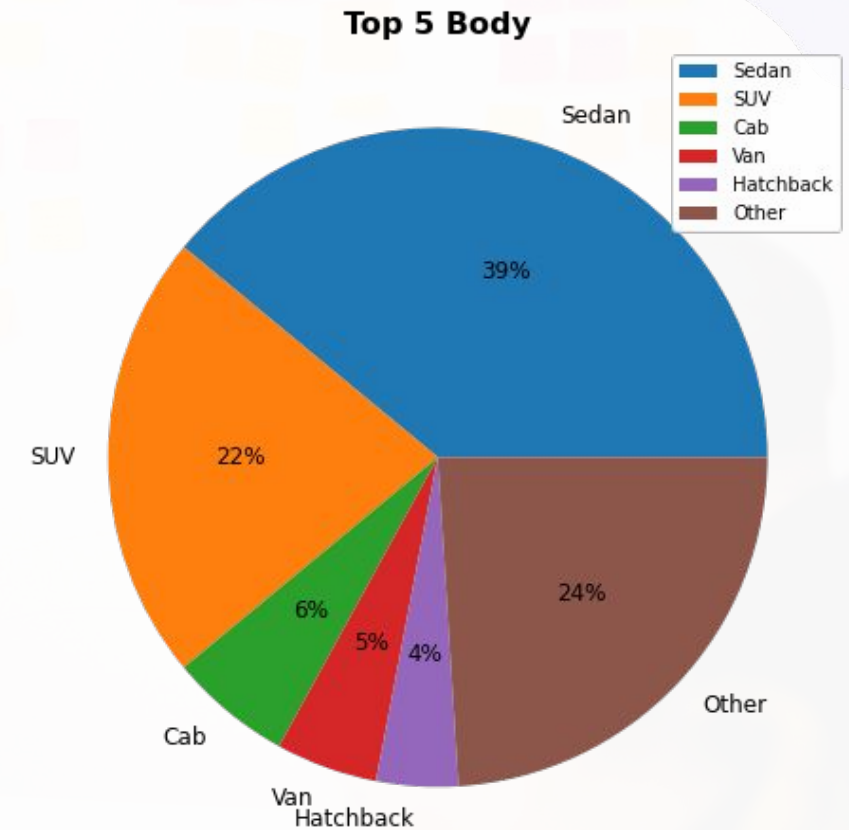
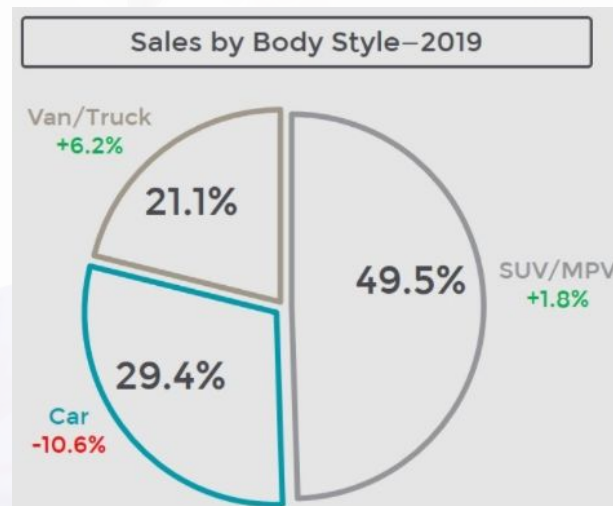
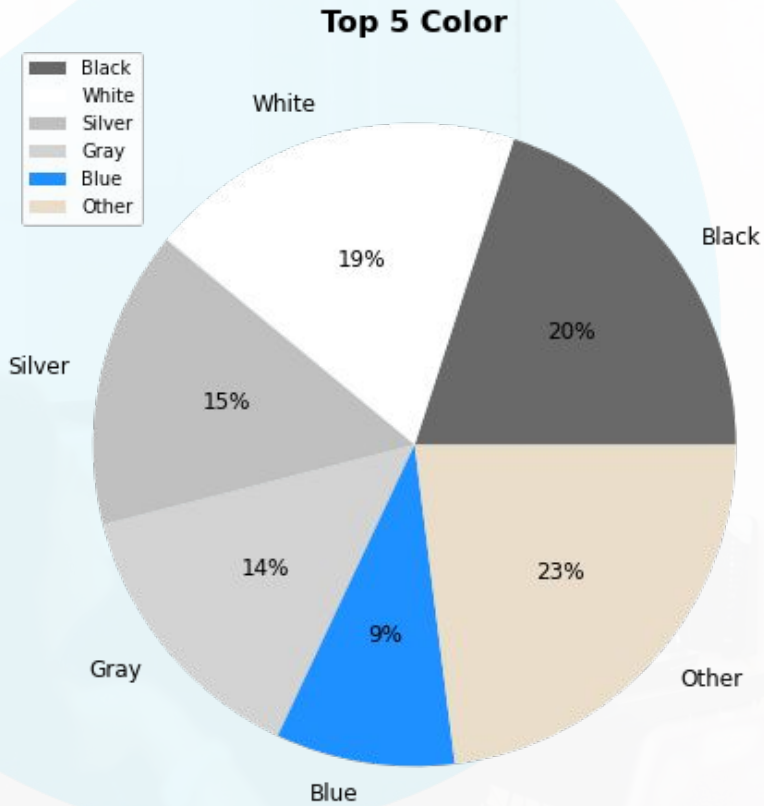


Top 10 mobil paling laku terjual



Top 5 pabrik terlaris adalah Ford, Chevrolet, Nissan, Toyota dan Dodge

Insights



Top 5 color dan body sesuai dengan market di US



Data Pre-processing

Pre-processing

No	Column Name
1	Year
2	Make
3	Model
4	Trim
5	Body
6	Transmission
7	Vin
8	State
9	Condition
10	Odometer
11	Color
12	Interior
13	Seller
14	Selling price
15	Sale date

- ★ 55881 Rows
- ★ 15 Column

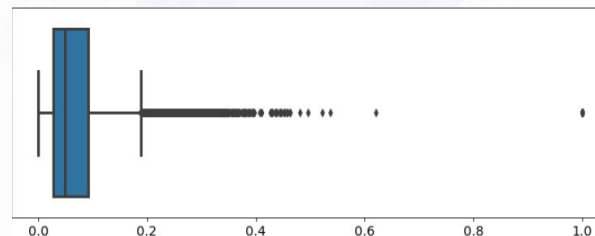
Data Cleansing

- ❖ Menghapus keseluruhan data kosong karena jumlahnya tidak mencapai 5% dari total keseluruhan data
- ❖ Menghapus kolom model yang memiliki nilai unique terlalu banyak serta kolom selling price dan odometer yang memiliki nilai tidak masuk akal

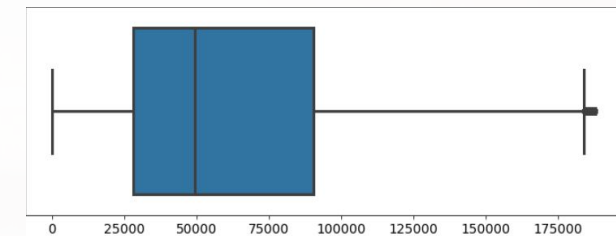
Outlier

- ❖ Handling outlier dengan IQR

• Odometer

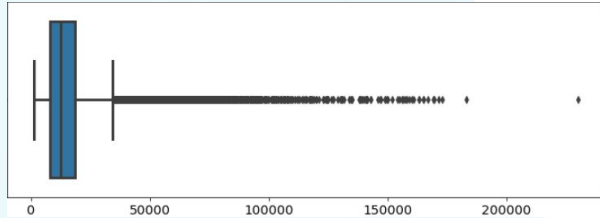


Handle
Outliers

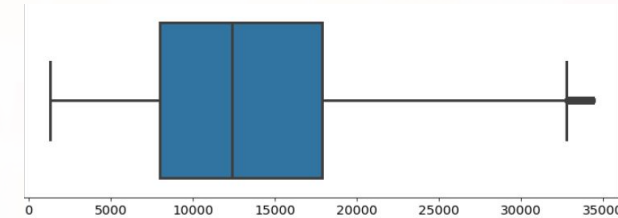


Pre-processing

- Selling price



Handle
Outliers



Scaling

- ❖ Membuat kolom baru dengan melakukan scaling pada kolom year, condition, odometer dan selling price

Feature Encoding

- ❖ Rekategorisasi pada kolom Make berdasarkan Brand mobil
- ❖ Rekategorisasi pada kolom body berdasarkan jenis mobil
- ❖ Melakukan label encoder pada kolom transmission, make, body dan color

Feature Selection

- ❖ vin, state, interior, trim, seller



DROP

Modelling

Modelling Experiments

Algorithm	MAE	RMSE	MSE	R2 Score
Linear Regression	4091.92	5248.97	27551691.67	0.49
Decision Tree	2483.07	3497.80	12234575.06	0.77
Random Forest	2306.28	3309.37	10951941.12	0.80
XGBoost	2250.57	3210.44	10306946.34	0.81

1. 4 Model sudah di Hyperparameter Tuning
2. Split Data; Train 70%, Test 30% from 432916 rows
3. Features Importance: Odometer, Make_label (Pabrikan), Body_label (Type of car), Condition, Year, Color_label

Selected Model

XGBoost

R2 Score: 0.73

Train	0.730
Test	0.728

Before
Hyperparameter Tuning



R2 Score: 0.81

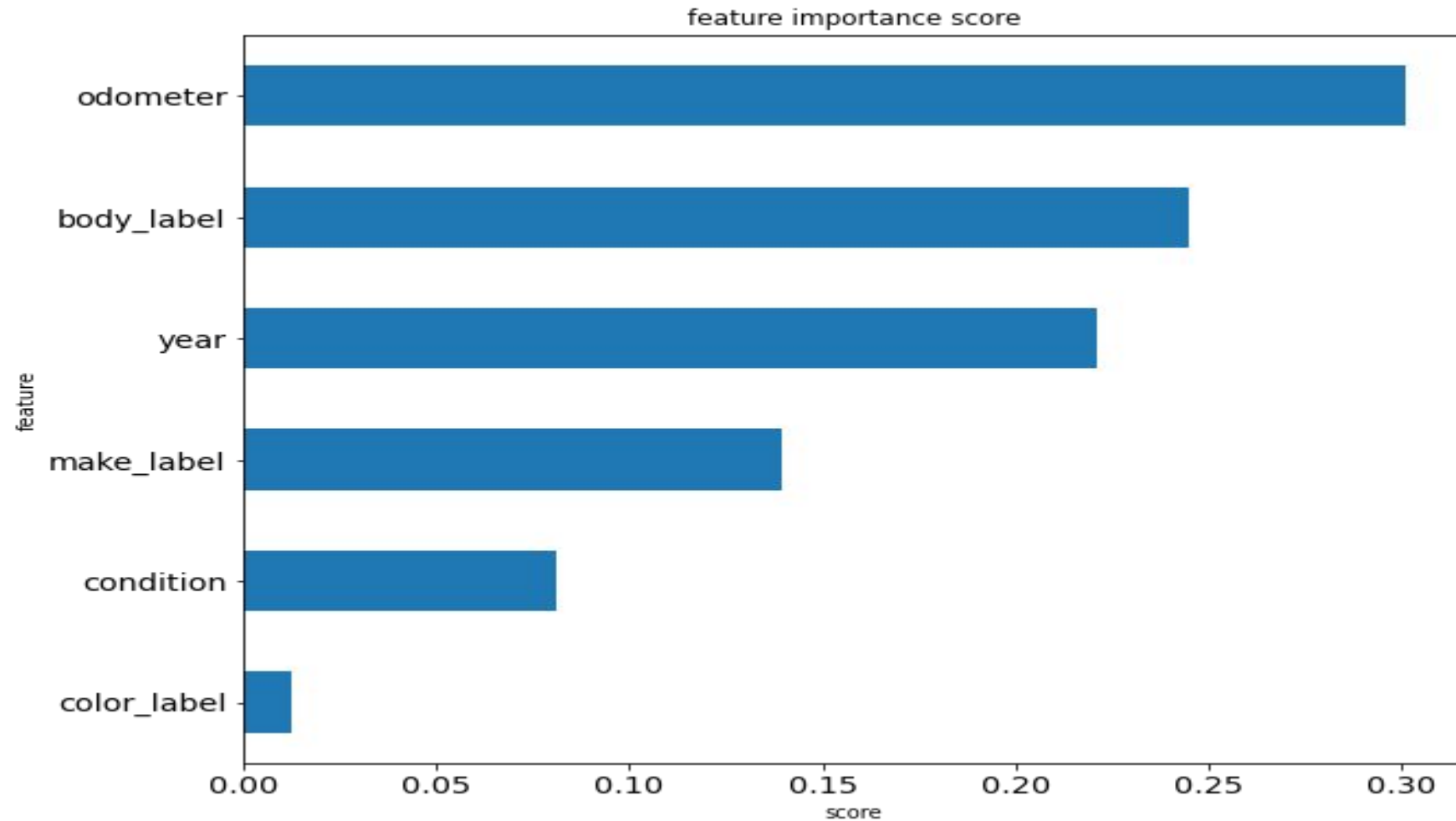
Train	0.829
Test	0.807

After
Hyperparameter Tuning

Hyperparameter:

- Max_depth
- Eta
- Min_child_weight
- Tree_method

Selected Model



Recommendation & Impact

Recommendation

TERKAIT DENGAN MODEL

Membuat **aplikasi web** berdasarkan model machine learning yang telah tim data buat untuk memprediksi harga mobil bekas berdasarkan kondisi dan fitur yang diberikan.



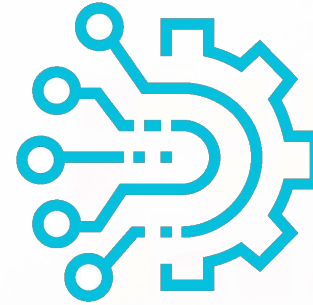
Jual Mobil

BRAND	
MODEL	
TYPE	
COLOR	
YEAR	
ODOMETER	
TRANSMISSION	
PREDICTED PRICE	

**Jual mobil kamu
dengan instan,
aman dan
nyaman!**

- GRATIS CEK CEPAT
PERKIRAAN HARGA JUAL
MOBIL!
- GRATIS INSPEKSI MOBIL!
- PENAWARAN DAN
PEMBAYARAN INSTAN!

Impact



Machine
Learning



WE BUY ANY CAR
IN 30 MINUTES



Instant Valuation



Free Inspection

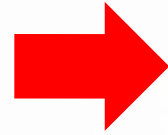


Instant Payment



Free Transfer

Impact



Staff A

\$23.000
(overpricing??)



Staff B

\$18.000
(underpricing??)

Jual Mobil

KIA

SORENTO

YEAR

ODOMETER

1663

TRANSMISSION

AUTOMATIC



\$21.500

Jual mobil ka
dengan i
am

PERKILAN HARGA JUAL MOBIL!
• GRATIS INSPEKSI MOBIL!
• PENAWARAN DAN PEMBAYARAN INSTAN!

Penentuan harga berdasarkan model machine learning meningkatkan ketepatan dan transparansi penentuan harga mobil untuk menghindari underpricing dan overpricing sehingga customer mendapatkan harga yang masuk akal dan tidak merasa dicurangi

Impact

	Konvensional	Machine Learning
Efisiensi Waktu 	Pengecekan harga mobil membutuhkan waktu sekitar 20 menit dan hanya dapat dilakukan pada jam kerja (7 jam).	Hanya membutuhkan waktu sekitar 2 menit dan dapat dilakukan 24 jam melalui website.
Sales 	Hanya mampu menyelesaikan 105 dari total 140 pengajuan dalam waktu 7 jam kerja.	Mampu menyelesaikan 140 pengajuan dalam 4 jam 40 menit atau mampu menyelesaikan 210 pengajuan dalam waktu jam kerja

Impact

Make	Model	Body	Color	Selling Price [Target]	ML Price	ML Relative Error
nissan	altima	sedan	white	10900	11745	7,75%
nissan	altima	sedan	white	13500	14238	5,47%
ford	fusion	sedan	black	8700	9363	7,62%
ford	fusion	sedan	brown	15200	15915	4,70%
ford	fusion	sedan	silver	7200	6527	9,35%
%AVERAGE						6,77%

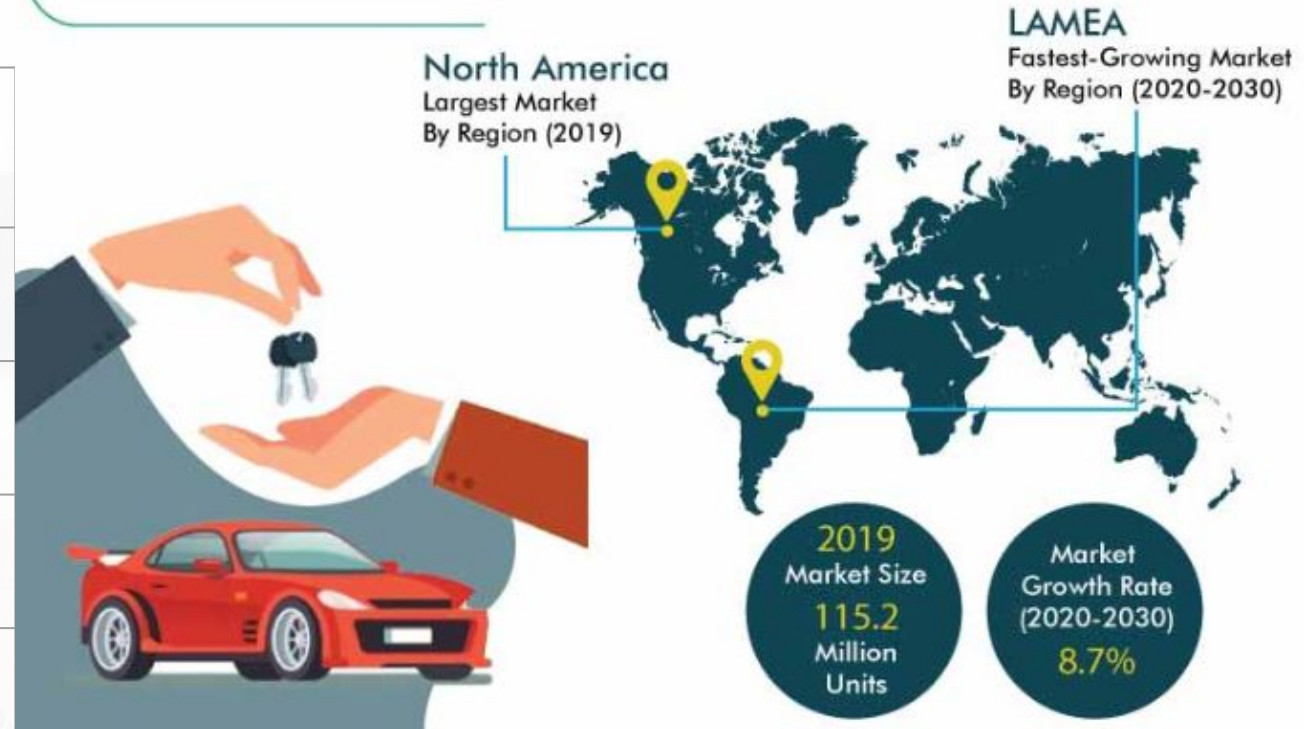
Range MAPE	Arti Nilai
< 10%	Kemampuan model peramalan sangat baik
10 - 20%	Kemampuan model peramalan baik
20 - 50 %	Kemampuan model peramalan layak
>50%	Kemampuan model peramalan buruk

Recommendation

TERKAIT DENGAN BUSINESS INSIGHTS

Price Segmentation	Medium, Low
Transmission	Automatic
Make	Ford, Chevrolet, Nissan
Model	Sedan, SUV, Cab
Color	Black, White, Silver

Global Used Car Market



Orang Amerika 'Takut' Habiskan Uang untuk Beli Mobil Baru, Lebih Suka Mobil Tua

Bisa memproduksi mobil sendiri, bukan berarti orang-orang di Amerika Serikat doyan gonta-ganti mobil. Studi menunjukkan rata-rata usia pakai mobil orang Amerika Serikat mencapai 11,9 tahun.

Selasa, 04 Agustus 2020 22:15
Editor : Dini Arining Tyas





THANK YOU