The background features a stylized red car on a dark teal road. To the left of the road are three yellow directional signs pointing right. The top sign is a large arrow, and the two below it are smaller rectangular flags.

Used Car Auction Prices Machine Learning Model

Car-a-
thon

Our Team



Bagus Ganjar Lugina



Samuella Magdalena E



Raihan Kurniasugianto



Bernadetha Stella



M Rifqi Sarosa

Table of Contents



1

Problem Statement

Problem definition and business metrics



2

Insights

Business insights from dataset



3

Data Pre-processing

Data cleansing, handling data and feature engineering



4

Modelling

Machine Learning model



5

Recommendation

Model impact and result to business

Introduction

Kami merupakan **tim data scientist** pada **PT Car-a-thon** (perusahaan jual beli mobil bekas berbasis yang di Amerika) dan bertanggung jawab untuk memberikan rekomendasi atas masalah yang dihadapi perusahaan untuk meningkatkan performa perusahaan.

Basis data menggunakan data historikal penjualan mobil bekas di Amerika ±2 tahun sebelumnya (Jan 2014 sampai Juli 2015).



O1

Problem Statement



The Problem



1. Waktu yang dibutuhkan untuk menilai harga mobil bekas kurang efektif dan efisien

Penentuan harga mobil yang tidak efektif dan efisien menyebabkan customer memilih tempat lain untuk menjual mobil.



2. Pelanggan/pembeli mobil bekas belum mendapatkan jaminan *value-for-money*

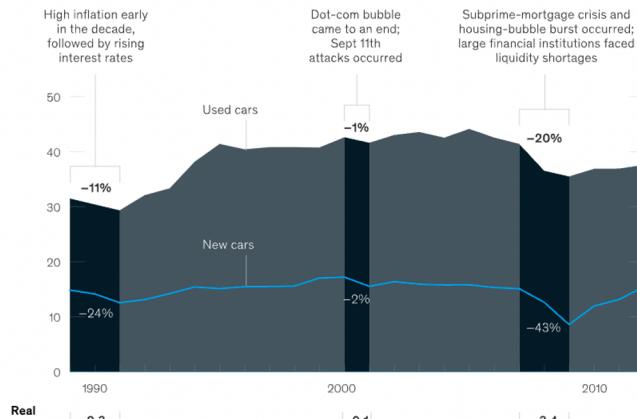
Penentuan harga mobil yang belum terstandarisasi dapat menimbulkan keraguan pada customer untuk menjual mobil pada perusahaan kita.

Waktu yang dibutuhkan untuk menilai harga mobil bekas kurang efisien

Consumer Habit

The used-car market is less susceptible to market shocks.

Used- and new-vehicle sales peak-to-trough decline in sales volume, million



Source: Federal Reserve Board; National Bureau of Economic Research; McKinsey analysis

Pasar mobil bekas di US besarnya 2x dari pasar mobil baru.

"The **used car appraisal** process may slightly different each dealership. But the entire process normally **takes from 15 to 30 minutes** depending how busy car manager or appraiser at that time."

Dari penjelasan di atas dapat ditarik kesimpulan bahwa rata-rata appraisal dalam used car yaitu **20 menit**.

**20
MIN**

- <https://www.autocheatsheet.com/trade-in/how-appraisal-process-works.html#:~:text=The%20used%20car%20appraisal%20process,appraiser%20is%20at%20the%20time.>
- <https://www.carhistory.com.au/news-and-media/infographic-used-car-survey-2016>
- <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/used-cars-new-platforms-accelerating-sales-in-a-digitally-disrupted-market>

Pembeli mobil bekas belum mendapatkan jaminan value-for-money

Consumer Habit



Quality was the most important factor for brands in the auto industry, with 30% of the vote. Affordability (21%) was the next most important attribute. (*Brandwatch*)

Important Factors for Consumers



30% - Quality



21% - Affordability



10% - Customer Service

Konsumen masih belum memiliki kepercayaan penuh atas penilaian harga mobil bekas.

Goal, Objectives ,and Business Metrics



GOAL

Meningkatkan **ketepatan nilai dan kecepatan waktu** dalam penentuan harga mobil bekas sehingga dapat meningkatkan performa penjualan mobil



OBJECTIVES

Membuat model *machine learning* untuk **memprediksi harga** mobil bekas



BUSINESS METRICS

- Appraisal time
- Sales Revenue

02

Business Insights



Insight Segmentation

Dataset Used Car Auction



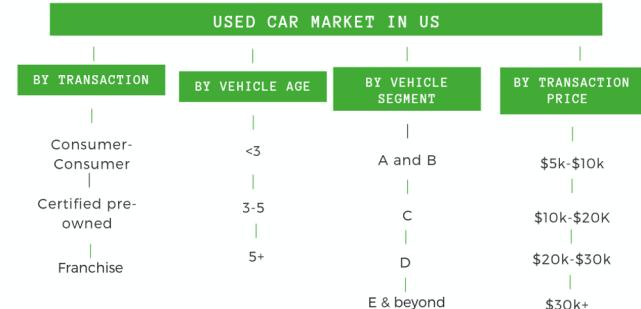
Dari **count plot** berikut menunjukkan harga mobil dengan kategori **medium** memiliki jumlah terbanyak, sedangkan kategori **high** memiliki kategori yang paling sedikit.

Impact:

Melihat range harga mobil paling banyak berada di segmen medium, dapat membantu perusahaan memperbanyak stok mobil dengan range medium.

External Data

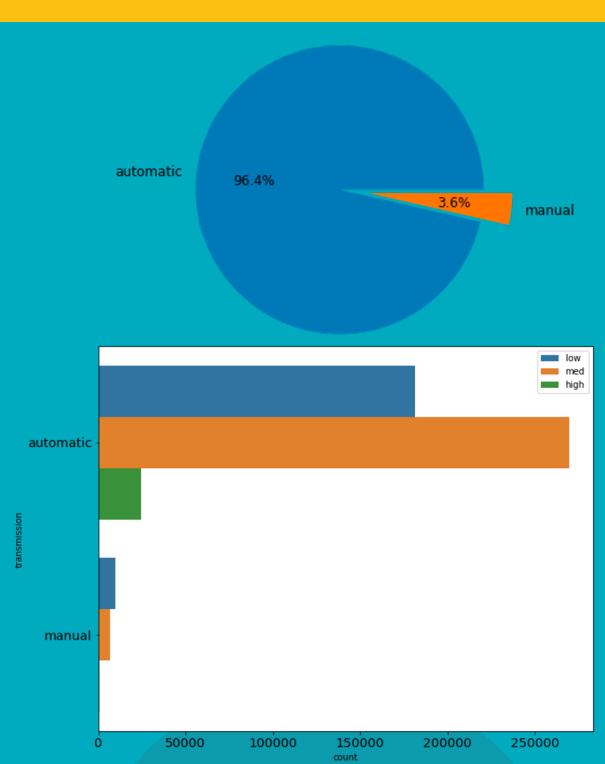
MARKET SEGMENTATION



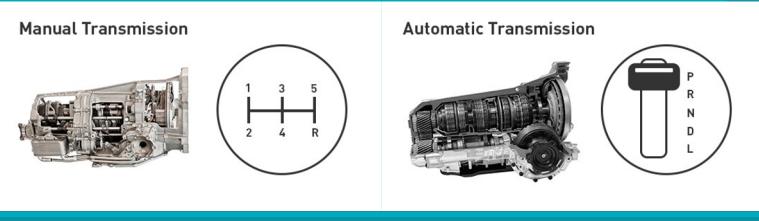
Insights

Car Transmission

Transmission Distribution



Dari kedua plot berikut, dapat disimpulkan bahwa jenis transmisi didominasi oleh transmisi **automatic**, dimana mobil untuk transmisi automatic didominasi dengan mobil pada **segmen harga medium**.



Manual:

Less than 3% of the cars sold in US have a manual transmission.

Automatic:

Popular in the US - 95% of the cars sold have automatic transmissions.



Impact:

Melihat jumlah datanya, model dapat bekerja baik untuk penentuan harga mobil dengan transmisi otomatis, namun perlu pengecekan ulang untuk mobil dengan transmisi manual.

Insights

Car Preferences

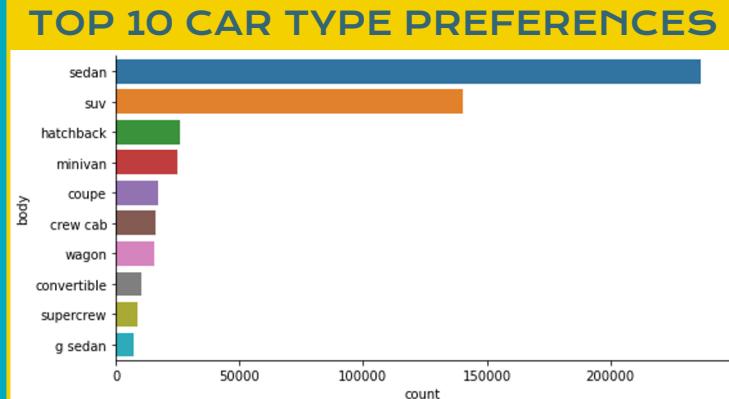
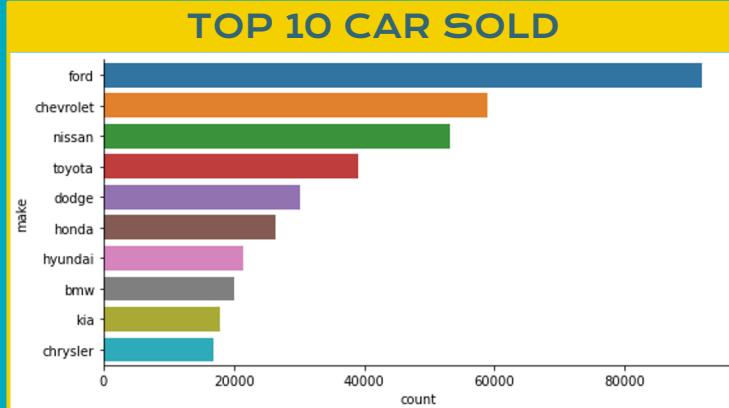


Ford dan **Chevrolet** merupakan brand mobil bekas yang paling banyak terjual.

Preferensi tertinggi tipe mobil berdasarkan mobil yang terjual adalah tipe **sedan** dan **SUV**.

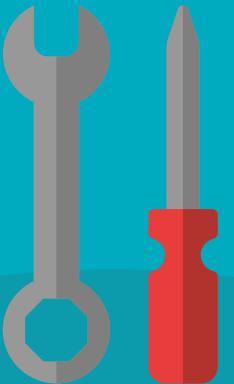
Impact :

Membantu perusahaan dalam menawarkan atau melakukan pembelian merk dan tipe mobil dengan penjualan terbanyak di pasaran.



03

Data Pre- Processing



Pre-Processing

```
↳ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 558811 entries, 0 to 558810
Data columns (total 16 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   year        558811 non-null   int64  
 1   make         548510 non-null   object  
 2   model        548412 non-null   object  
 3   trim         548160 non-null   object  
 4   body          545616 non-null   object  
 5   transmission 493458 non-null   object  
 6   vin           558811 non-null   object  
 7   state         558811 non-null   object  
 8   condition     547017 non-null   float64 
 9   odometer      558717 non-null   float64 
 10  color         558062 non-null   object  
 11  interior      558062 non-null   object  
 12  seller         558811 non-null   object  
 13  mmr            558811 non-null   int64  
 14  sellingprice  558811 non-null   int64  
 15  saledate       558811 non-null   object  
dtypes: float64(2), int64(3), object(11)
memory usage: 68.2+ MB
```

Sebelum Pre-Processing

- 558811 baris, 16 kolom
- Tipe data sudah sesuai
- Tidak ada duplikat
- Target: 'sellingprice'



Data Cleansing

- Drop kolom dengan **unique values tinggi** ('state', 'vin', 'model', 'interior', 'seller', 'trim')
- Drop kolom yang memiliki **korelasi dengan target terlalu tinggi** ('mmr')
- Drop baris target ('sellingprice') yang memiliki **nilai anomali**



Handling Missing Value

- **Drop baris** dengan missing value <5% ('body', 'model', 'trim', 'condition', 'make', 'color', dan 'odometer')
- **Impute data modus** untuk baris dengan missing value >10% ('transmission')

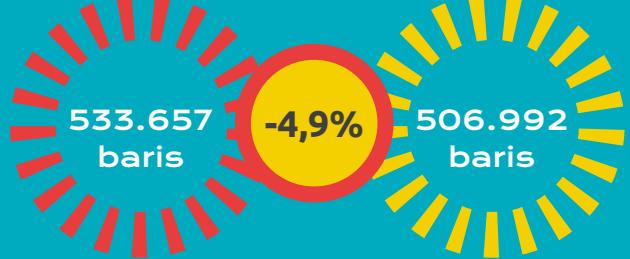
Pre-Processing



Handling Outlier

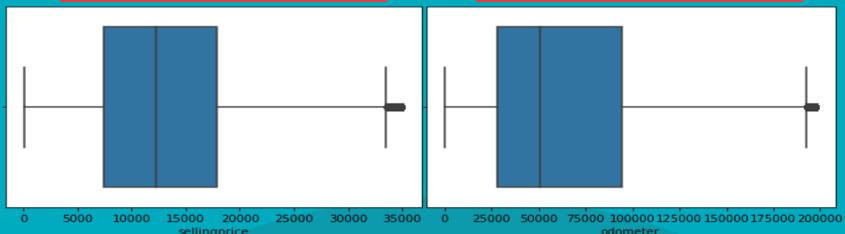
Dilakukan pada 2 fitur: '**sellingprice**' dan '**odometer**' dengan metode IQR

Sebelum
memfilter outlier



'**sellingprice**'

'**odometer**'



Feature Transformation



Normalization **year** dan **odometer**

→ distribusi yang sangat skew



Standardization **sellingprice** dan **condition**

→ distribusi yang hampir mendekati normal



Feature Encoding

Label Encoding

'Transmission', 'top_10_model',
'top_10_trim', 'top_10_body',
'top_10_color', 'top_10_make'



Feature Selection

9 Fitur

odometer, year, body_label, make_label,
condition, trim_label, color_label,
model_label, transmission_label

04

Model Evaluation



Model Comparison

Algorithm	RMSE	MAE	R ²	Train	Test
Linear Regression	5193.46	4006.74	0.537	0.540	0.537
Decision Tree	3581.19	2447.57	0.780	0.795	0.780
Random Forest	3221.76	2215.91	0.810	0.903	0.810
XgBoost	3359.42	2296.73	0.806	0.818	0.805
LightGBM	3294.39	2244.70	0.813	0.822	0.813

1. 5 model algoritma sudah dilakukan **hyperparameter tuning**,
2. Split data **train (70%)** dan **test (30%)** dari **506.992 rows**,
3. Evaluasi metrics yang digunakan adalah **R²**,

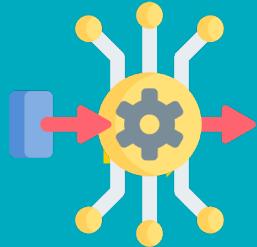
Selected Model



R² Score:
0.803



Train: 0.804
Test: 0.803



R² Score:
0.812



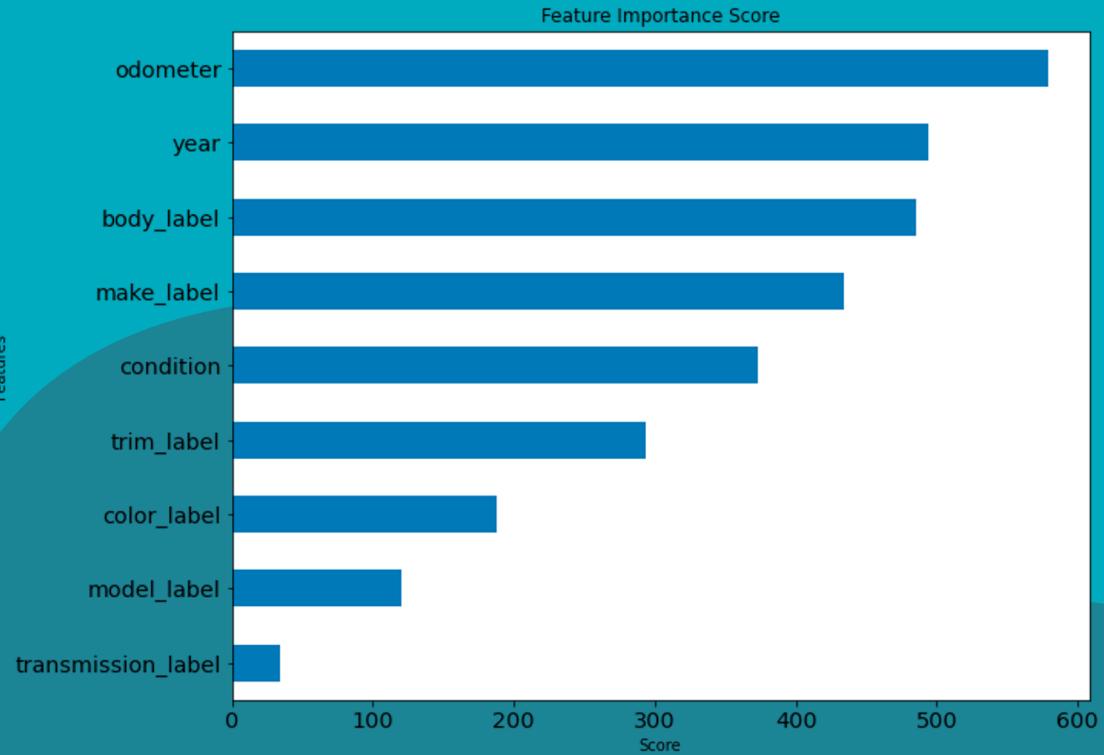
Train: 0.822
Test: 0.813

Before
Hyperparameter Tuning

After
Hyperparameter Tuning



Selected Feature



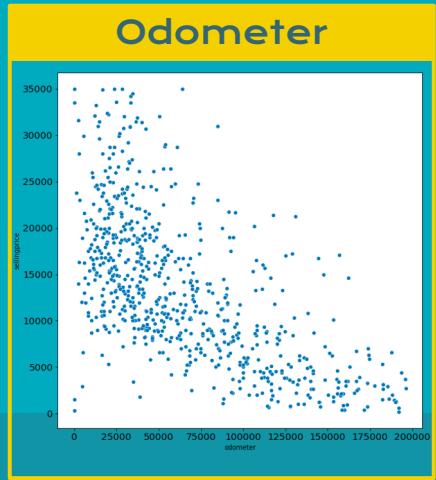
Feature Importance



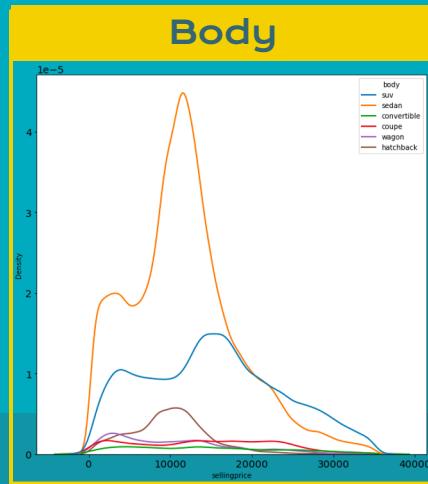
Pengurangan fitur **transmission_label**, **model**, dan **color_label** berdampak pada penurunan akurasi model untuk mencapai model dengan hasil yang lebih optimum sehingga fitur yang digunakan pada model tetap/tidak dikurangi.

(9 fitur: **odometer**, **year**, **body_label**, **make_label**, **condition**, **trim_label**, **color_label**, **model_label**).

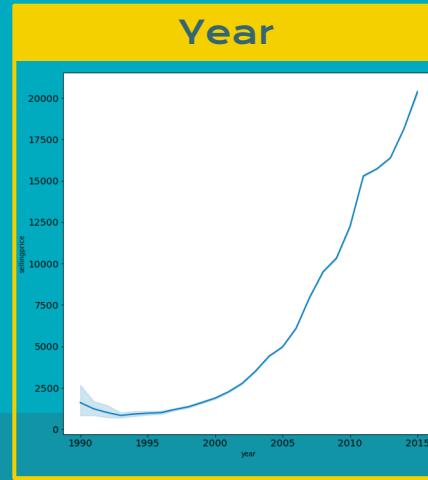
Interpretasi Feature Selection



Semakin tinggi
odometer maka
semakin rendah harga
mobil yang dijual.



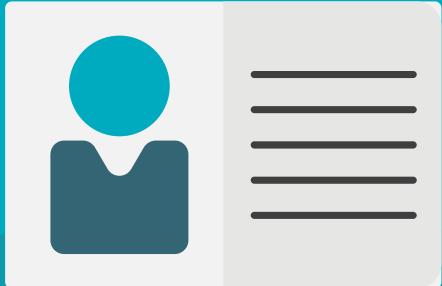
Persebaran harga
mobil berdasarkan
body cukup merata
berdasarkan
harganya.



Semakin baru
tahunnya akan
semakin tinggi harga
mobil yang dijual.

05

Recommendation and Impact



Recommendation

Menyempurnakan aplikasi penentuan harga mobil dengan **menambahkan model machine learning** untuk **prediksi harga mobil** berdasarkan feature-feature yang diinput. Aplikasi tersebut akan digunakan oleh staff/appraiser untuk membantu **mempercepat penentuan harga** mobil.

Before ML Model

CAR - A - THON Car Price Prediction

Car Appraisal Form

Date: _____

Year: _____

Make: _____

Model: _____

Trim: _____

Body: _____

Transmission:
 Automatic Manual

Vin: _____

Condition: 1 2 3 4 5

Odometer: _____

Color: _____

Interior: _____

Seller: _____

Car Price Analysis:

MMR: _____

Calculation accumulation based on condition:

Selling price:

Appraiser,

After ML Model

CAR - A - THON Car Price Prediction

Input Car Specification

Year: 2012 Body: Sedan

Brand: BMW Color: Black

Model: Altima Condition: 4.3

Trim: base Transmission: Manual

Odometer: 16399

Selling Price \$ 1233333333



Existing Condition

1.176

Pengajuan Harian

- Dengan asumsi terdapat 25 hari kerja per bulan, maka terdapat:
[558.811 pengajuan/(19 bulan x 25 hari)] = **1.176 pengajuan per hari**

21

Pengajuan/hari/staff

- 1 Orang Staff
- Asumsi: 1 hari = 7 jam kerja dan proses penilaian 20 menit/pengajuan

56

Staff

Dibutuhkan 56 orang staff untuk melakukan perhitungan harga mobil setiap harinya

Impact Appraisal Time

Based on data

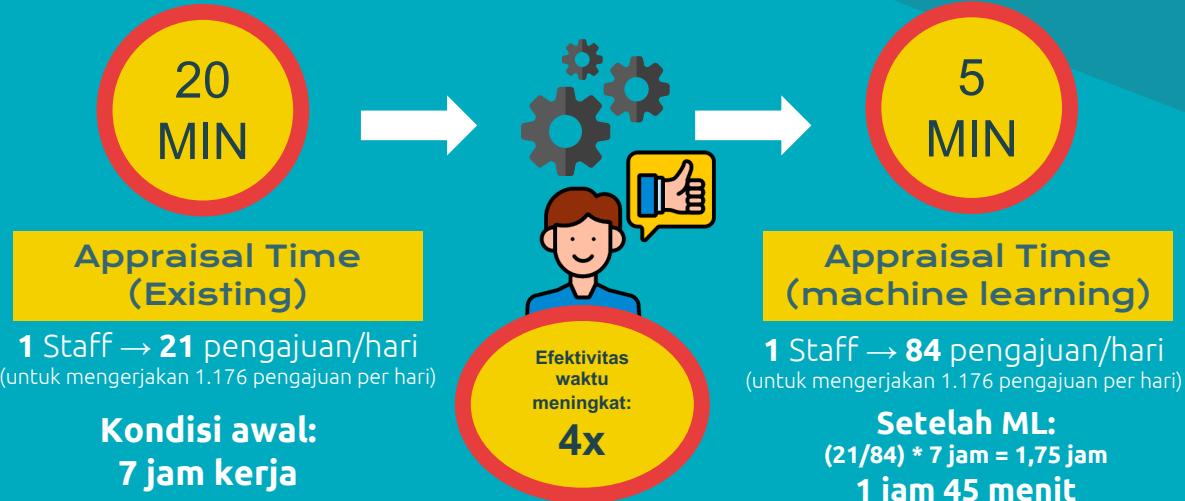
"The used car appraisal process may slightly different each dealership. But the entire process normally takes from 15 to 30 minutes depending how busy car manager or appraiser at that time."

Sumber :
<https://www.autocheatsheet.com/trade-in/how-appraisal-process-works.html#:~:text=The%20used%20car%20appraisal%20process,appraiser%20is%20at%20the%20time.>

Opsi 1: Lay-off

Efisiensi maksimal pengeluaran perusahaan untuk gaji karyawan:
100% - [(14/56)x 100%] = 75%

Perusahaan melakukan cost efficiency sebesar 75%.

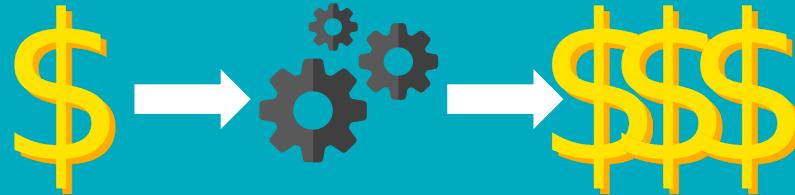


Impact Sales Revenue

Opsi 2: Sales Improvement

Peningkatan profit apabila
peningkatan sales mencapai angka
optimum: $4.704/1.176 = 4x$

**Perusahaan dapat
meningkatkan sales sampai
dengan **4x**.**



Jumlah staff: 56 orang

Total pengajuan yang dapat dikerjakan :
1.176 pengajuan per hari

Jumlah staff: 56 orang

Total pengajuan max yang dapat dikerjakan :
4.704 pengajuan per hari

Rekomendasi untuk optimalisasi sales:

- Divisi/tim **Sales** di perusahaan untuk bisa **meningkatkan stok** penjualan mobil dengan **merek tertentu (Ford, Chevrolet) dan tipe tertentu (sedan dan SUV)** agar penjualan mobil dapat meningkat.
- Memprioritaskan penjualan mobil keluaran **tahun 2000 ke atas dan membatasi masuknya stok mobil keluaran tahun 2000 ke bawah.**
- Divisi/tim **Marketing** di perusahaan melakukan promo/campaign terhadap mobil yang menjadi **top 10 penjualan**.

Thank You

Appendix: Profit Estimation

	Rata-rata (\$)	Profit Sebelum ML Model (\$)	Profit Setelah ML Model (\$)
mmr	13.293,35	Asumsi: 1.176 pengajuan per hari $(1.176 \times 172,74)$	Asumsi: 4.704 pengajuan per hari $(4.704 / 1.176) \times$ 203.142
selling price	13.120,61		
SELISIH	172,74	203.142	812.569

Selisih profit: \$ 609.427
Kenaikan profit: 3%