KAIZEN CONSULTING

Insurance **Strategy**

Final Project Rakamin Data Science Batch 25

Agenda

- Background and Understanding Needs
- Analysis Method
- Summary
- Impact
- Business Recommendation

KAIZEN CONSULTING

Background and Understanding

KAIZEN CONSULTING

Background

KAIZEN CONSULTING

80% future profits will come from **20% customers**.

Cross Selling

- Less efforts
- Improve satisfaction & retention
- Significantly increases products sold to analyzed customer up to 300%

Source:
- Leading on the Edge of Chaos - Emmet C. Murphy & Mark A. Murphy
- McKinsey

Client Overview

Company **X**, is an insurance company that currently wants to do cross selling product in order to increase the company's revenue. Currently they have **12% response rate**.

Goals

Time and Cost Efficiency

Objective

Machine learning model for filtering survey's target

Business Matric

Response rate

KAIZEN CONSULTING

Being Solution

Problems

Solution

KAIZEN CONSULTING

Technology Consulting

Data and Analytics

Machine Learning

Method
Analysis

KAIZEN CONSULTING

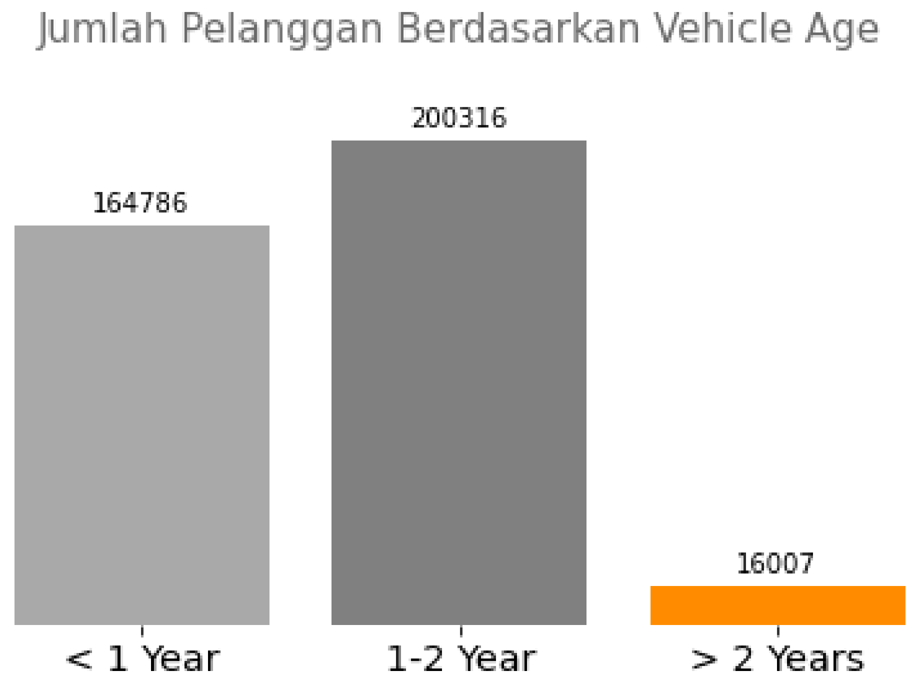
Data
Overview

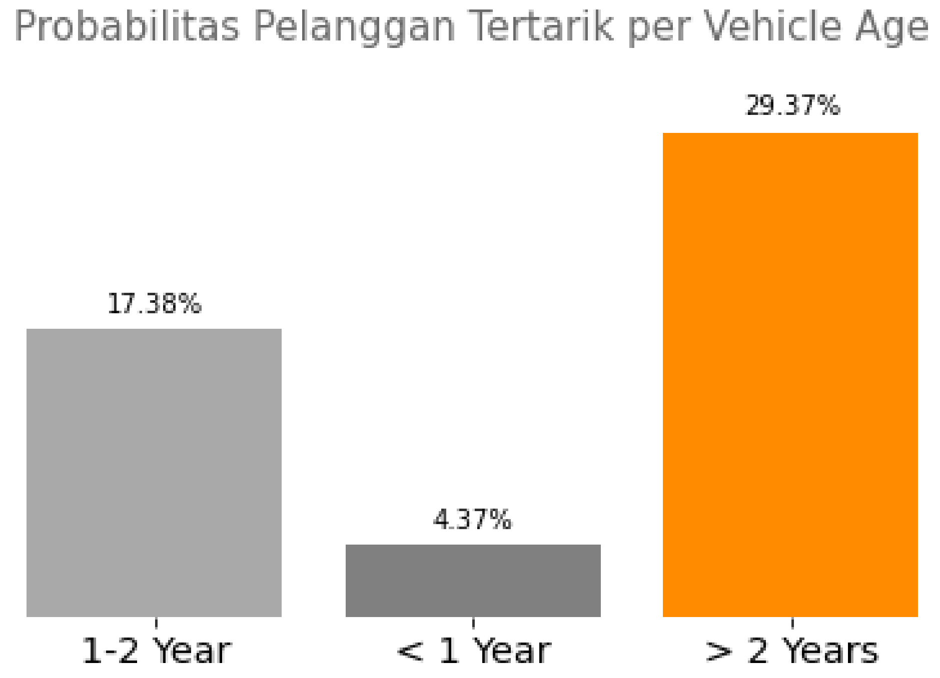
The dataset contains 381,109 rows, with no missing value and no duplicate.

	Details
id	Unique ID of the Customer
Gender	Gender of the Customer
Age	Age of the Customer
Driving License	Customer has/does not have driving license
Region Code	Unique code of the region of Customer
Previously Insured	Customer has/does not have vehicle insurance
Vehicle Age	Age of the vehicle
Vehicle Damage	Customer got their vehicle damaged/not in the past
Annual Premium	The amount customer needs to pay as premium in the year
Policy Sales Channel	Anonymized code for the channel of outreaching to the customer ie. Different Agents, Over Mail, Over Phone, In Person, etc.
Vintage	Number of days Customer has been associated with the company
Response	Customer is interested or not

KAIZEN CONSULTING

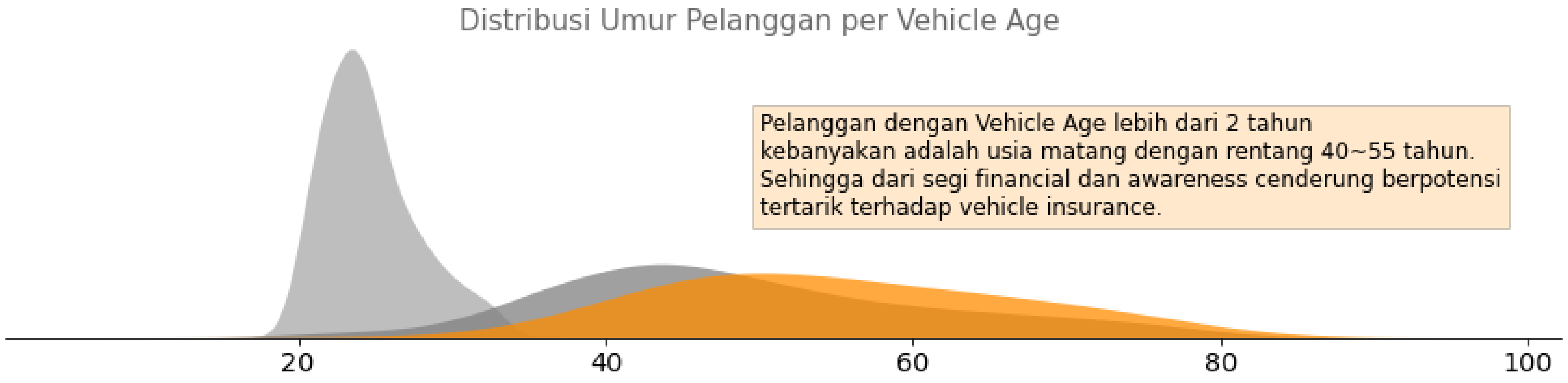
Data
Insights





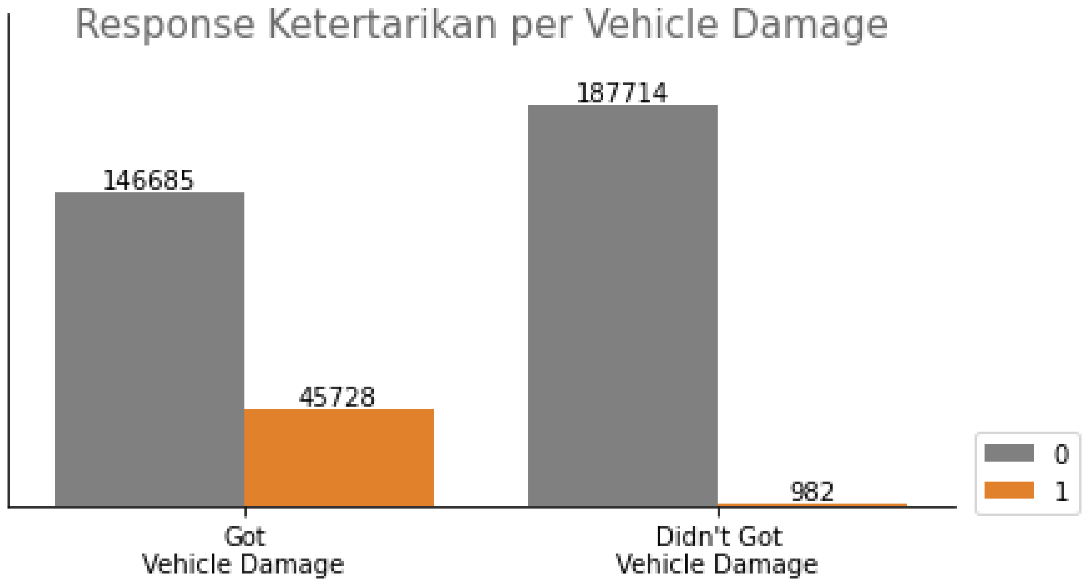
Quickwin based on "Vehicle Age" Data

To increase the response without the ML model, as quickwin can focus on offering new products to those who have a vehicle age of more than 2 years.



KAIZEN CONSULTING

Data Insights

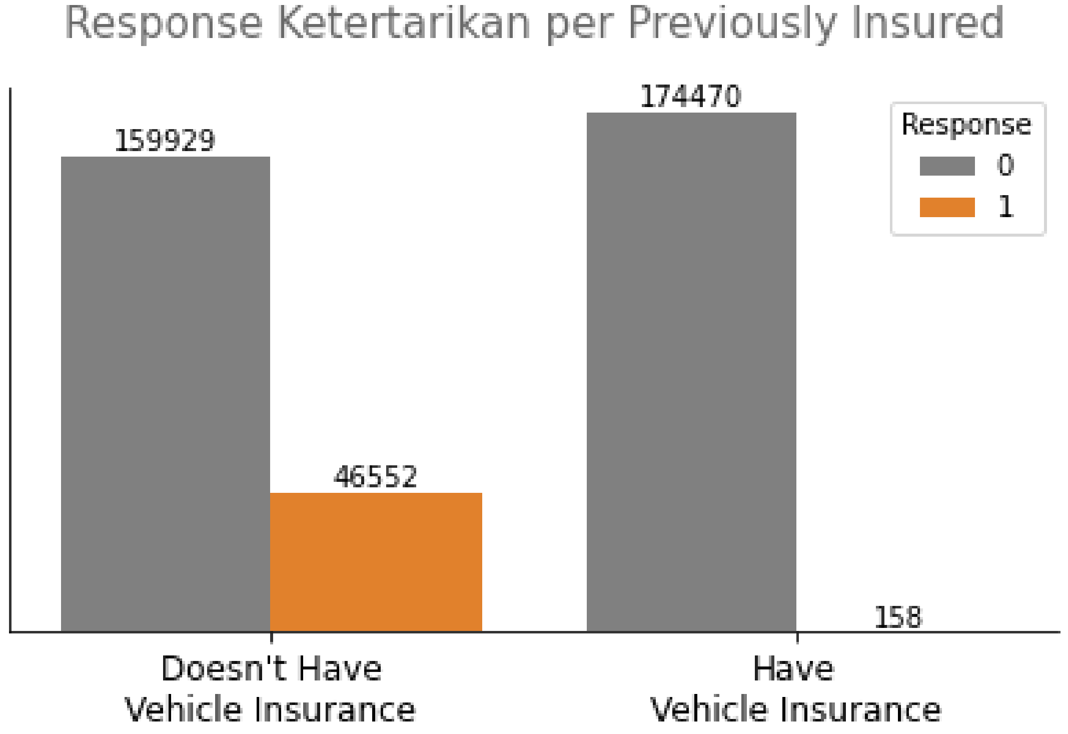


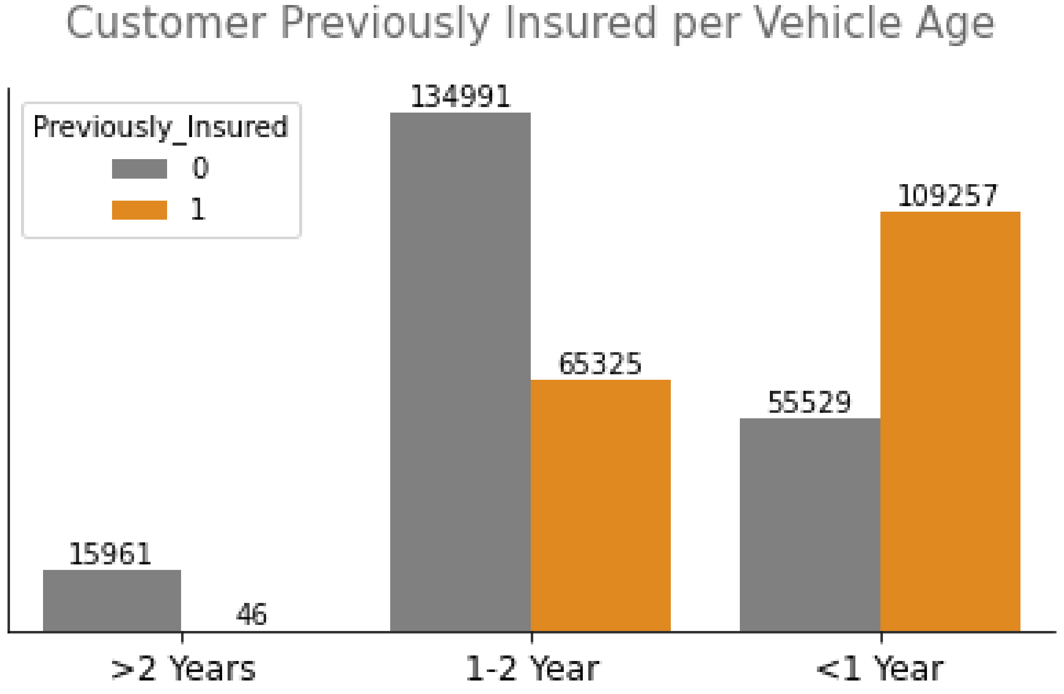
Quickwin based on "Vehicle Damage" Data

Focus on offering to customers who have experienced vehicle damage, as well as providing socialization on the importance of vehicle insurance.

Quickwin based on "Previously Insured" Data

Focus on offering to customers who do not have vehicle insurance or vehicle age > 1 year, because if you see customers who already have vehicle insurance or have vehicle age < 1 year, they will almost certainly not be interested (refuse).





KAIZEN CONSULTING

Data **Cleansing & Preprocessing**

Data cleansing

There's no missing value & duplicate data, but there's some outlier data and we handle it using standardscaler.

Imbalance treatment

Using class_weights = 'balanced' to optimize the scoring for the minority class.

Feature highlight

11 features: id, Previously Insured, Vehichle Damage, Age, Gender, Driving License, etc.

Train test split

We split the data into train and test for modelling and evaluation.

Model **Comparison**

	Precision			Recall			ROC_AUC		
	Train	**CV -Mean**	**CV - STD**	**Train**	**CV - Mean**	**CV - STD**	**Train**	**CV - Mean**	**CV - STD**
Logistic Regression	0.253	0.253	0.001	0.968	0.969	0.002	0.785	0.831	0.002
Decision Tree	0.862	0.292	0.004	0.999	0.312	0.003	0.989	0.604	0.002
XGBoost	0.298	0.287	0.001	0.942	0.903	0.004	0.816	0.854	0.001
Random Forest	0.889	0.316	0.002	0.989	0.238	0.004	0.986	0.810	0.001
AdaBoost	0.050	0.347	0.257	0.000	0.0003	0.0002	0.499	0.848	0.001

*Initial Comparison (before hyperparameter tuning)

Hyperparameter & Threshold Tuning

	Precision		Recall	
	Train	CV - Mean	Train	CV - Mean
Random Forest (before)	0.889	0.316	0.989	0.238

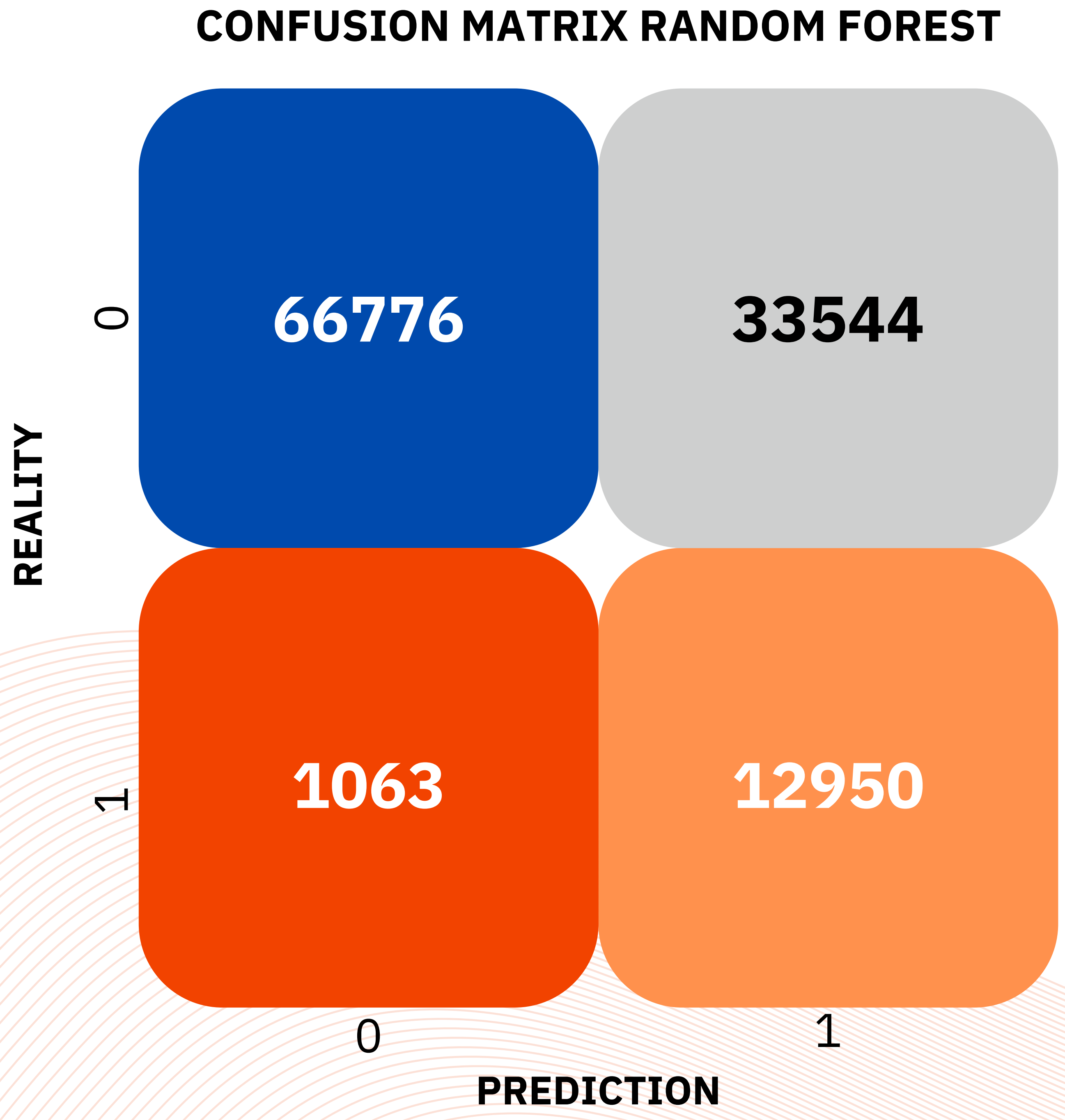
	Precision		Recall	
	Train	Test	Train	Test
Random Forest (after)	0.295	0.280	0.971	0.928

Parameter

- n_estimator= 200
- max_depth = 20
- max_features=0.3
- min_samples_leaf=10

- threshold = 0.415058

Simulation



From confusion matrix obtained

- TP (True Positive) : 12.950
- FP (False Positive) : 33.544
- TN (True Negative) : 66.766
- FN (False Negative) : 1.063

Metrics Score

- **Precision : 0.28** (response rate)
- Recall : 0.92

KAIZEN CONSULTING

Summary

The focus of the model made is to find the best increase in response rate/precision with the condition of loss of potential customer approaching zero/recall close to equal to 1. It can be seen that there is an increasement in **response rate/precision score** to **0.28 or 28%** (before **12%**) with a **recall score of 0.92**.

We also want to reduce the score of False Negative (FN), which means that there are customers who are predicted to be negative (not interested) by the model but actually they have an interest. This can lead to loss of potential profit from existing customers.

The recall score obtained from the Random Forest model is **0.92**, good enough to suppress the FN value. Because in this case there is a potential where **the loss from the potential customer has a much greater value than the loss experienced from the loss of approach** with a ratio of '**1 : infinity**'

KAIZEN CONSULTING

Impact

114,323**x** 46,494**x**

-59.3%

x represent of the number of times to do a survey/ approach and reflects the cost itself.

The cost saving can't be used as it is. There is about **1063** of need to analyze whether the number of potential of cost loss lower (in total) than the cost saving or not.

KAIZEN CONSULTING

Business **Recommendations**

Detailed data

Customer profile: vehicle type, occupation, family

Marketing Strategy

Family discount, bundling packages, car repair shop

Value

Cost of survey or approach, customer potential revenue (value)

Thank **You!**

KAIZEN
CONSULTING

Mentor:

Al Fath Terry

Annissa Sagita

Data Analyst

Galih Sawunggaling

Data Analyst

M. Wahid Adam

Data Scientist

Nur Hikmawati

Data Analyst

Pradipta Adyatma

Data Analyst

Rifka Amalia

Data Scientist

Riska Serli

Data Scientist