Capstone Project

Health Insurance Cross Sell Prediction

By

Ravi Kumar

Content

AI

- Introduction
- Problem Statement
- Data summary
- Exploratory Data Analysis(EDA)
- Feature Engineering
- Building Model
- Conclusion

Background

AI

- An health insurance company that has provided health insurance to its customers now they want to sell their customers from past years vehicle insurance
- There is vehicle insurance where every year a customer needs to pay a premium of a certain amount to the insurance provider company so that in case of an unfortunate accident by the vehicle, the insurance provider company will provide a compensation(called ‘sum assured’) to the customer.

Problem Statement

AI

Our client is an Insurance company that has provided Health Insurance to its customers now they need your help in building a model to predict whether the policyholders (customers) from the past year will also be interested in Vehicle Insurance provided by the company. An insurance policy is an arrangement by which a company undertakes to provide a guarantee of compensation for specified loss, damage, illness, or death in return for the payment of a specified premium. A premium is a sum of money that the customer needs to pay regularly to an insurance company for this guarantee. Building a model to predict whether a customer would be interested in Vehicle Insurance is extremely helpful for the company because it can then accordingly plan its communication strategy to reach out to those customers and optimise its business model and revenue.

Data Description

AI

- id: Unique ID for customer
- Age: Age of the customer
- Driving_License 0 : Customer has DL or not
- Region_Code: Unique code for the region of the customer
- Previously_Insured 1 : Customer already has Vehicle Insurance or not
- Vehicle_Age: Age of the Vehicle
- Vehicle_Damage 1 : Past damages present or not
- Annual_Premium: The amount customer needs to pay as premium
- Policy Sales Channel: Anonymized Code for the channel of outreaching to the customer ie. Different Agents, Over Mail, Over Phone, In Person
- Vintage: Number of Days, Customer has been associated with company
- Response :Customer is interested or not

Insights from the Problem and Data

AI

What is an insurance firm?

- **If a loss occurred a guarantee of compensation for specified loss, damage, illness, or death in return for the payment of a specified premium.**

What is the probability of buying an insurance?

- **In insurance industry, it refers to a situation in which people only buy insurance when they expect high risks. Buying insurance is not appropriate for all levels and types of risks. In many cases, people are better off taking actions to avoid risk, retain (accept) risk or reduce risk. Buying insurance makes the most sense when the potential loss is great and there is a significant probability of loss.**

Insights from the Problem and Data

AI

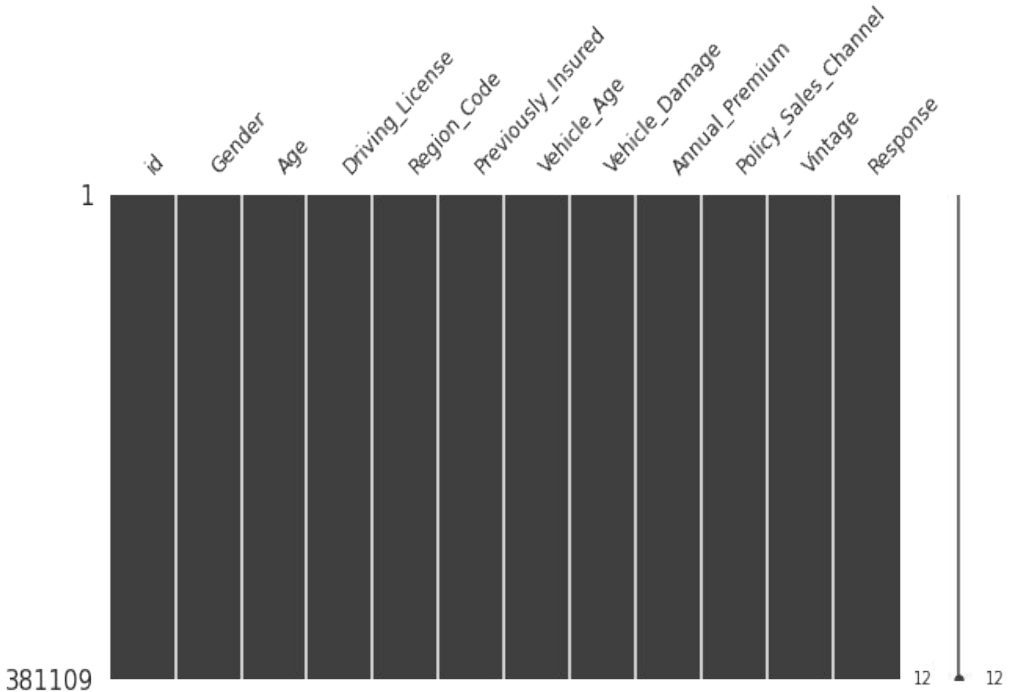
How many people are knowledgeable about insurance policy and how many of them claim insurance?

- **Let's say about four in 10 men describe themselves as being very knowledgeable about life insurance. As in the problem statement, about 2 or 3 get hospitalized out of 100, which means 2 to 3 percent claim the insurance. This way everyone shares the risk of everyone else.**

So we need to building a model to predict whether a customer would be interested in Vehicle Insurance is extremely helpful for the company because it can then accordingly plan its communication strategy to reach out to those customers and optimise its business model and revenue. Now, we need to predict whether the customer would be interested in Vehicle insurance or not.

Checking for missing values in the dataset

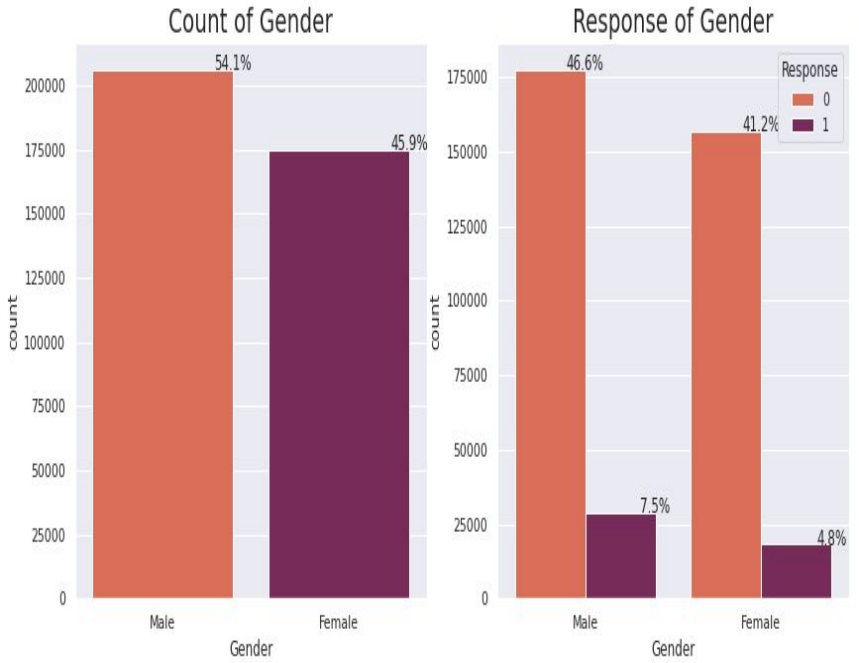
- **There are no missing values in the dataset**



Gender variable

As we can see from the graph,

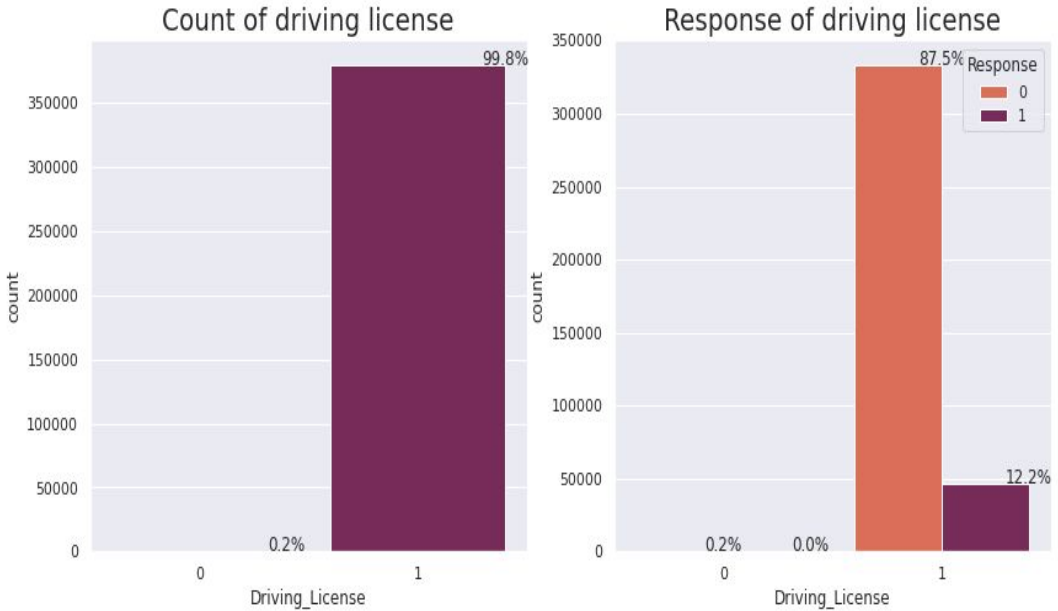
- **The gender variable in the dataset is spread nearly evenly. The male category is marginally larger than the female category, and the likelihood of purchasing insurance is also slightly higher.**
- **Only 12.3% people are interested in buying vehicle insurance and 87.7% are not interested to buy vehicle insurance**



Driving License

As we can see from the graph,

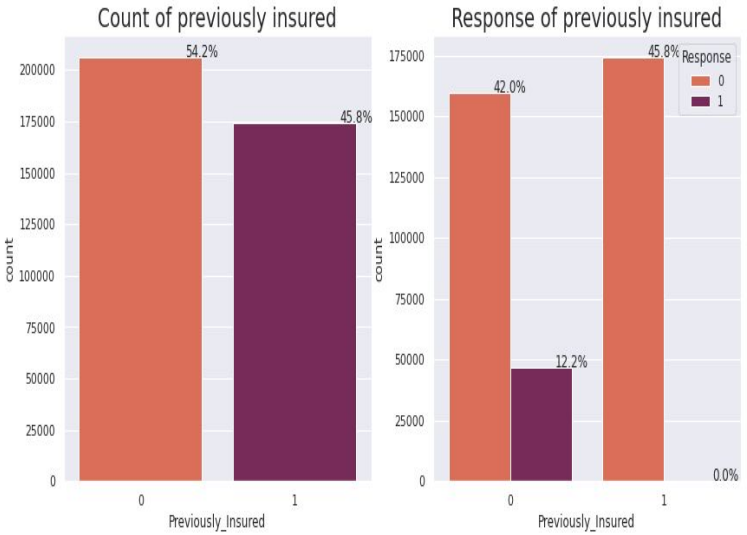
- **99.8% of customers have DL, whereas 0.2% do not have DL.**
- **Only a small percentage of people who have a DL (12.2%) are interested**



Previously Insured

As we can see from the graph,

- **45.8% people are insured previously, in that 12.2% people interested to buy the vehicle insurance again, Which means people are aware of insurance policy and ready to pay a premium amount, for better off taking actions to avoid certain risks or reduce risk.**
- **So buying insurance makes the most sense when the potential loss is great and there is a significant probability of loss.**

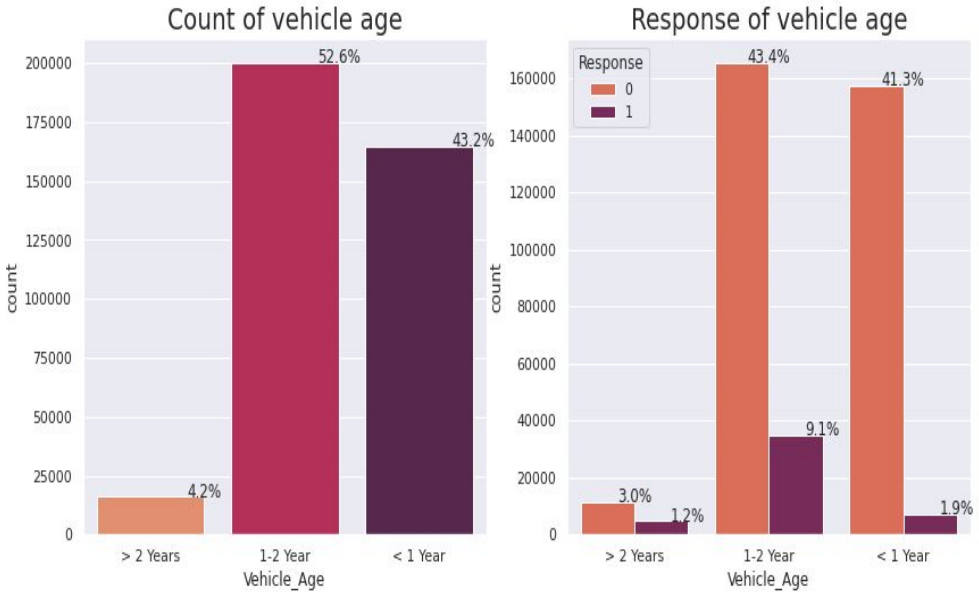


Vehicle Age

As we can see from the graph,

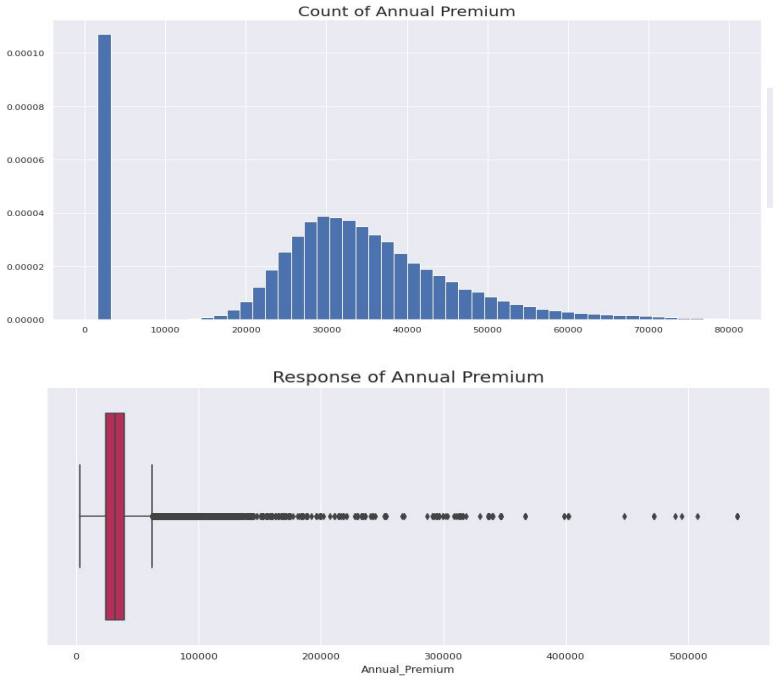
- **Around 4.2% of vehicles are more than two years old, 52.6% are between one and two years old, and 43.2% are under one year old.**

As vehicle age increases most of the people are aware of insurance and interested to buy the insurance for reducing the risk



Annual Premium

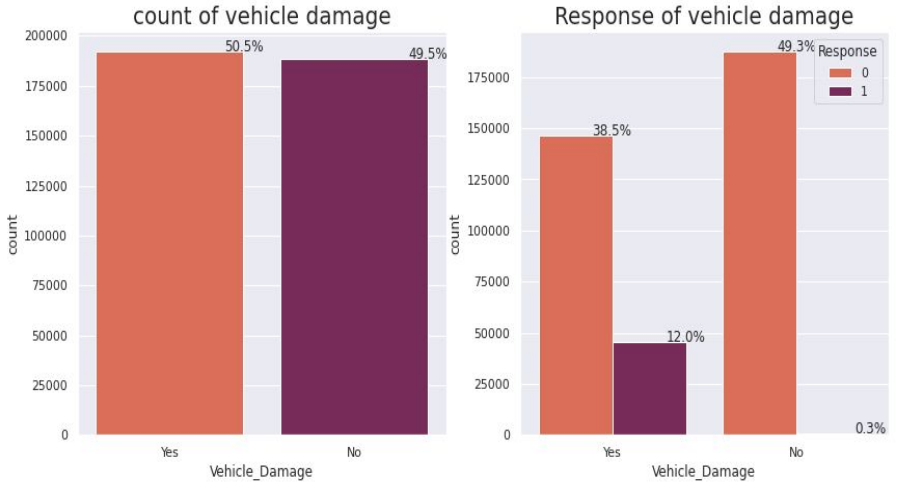
- From the distribution plot we can infer that the annual premium variable is right skewed
- From the boxplot we can observe lot of outliers in the variable



Vehicle Damage

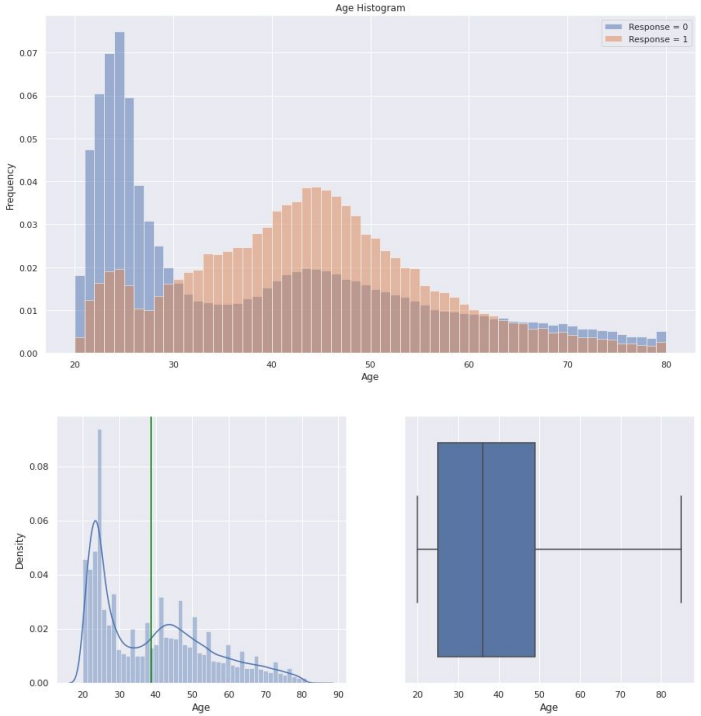
As we can see from the graph,

- **50.5% of the vehicles have past damage**
- **12.0% of people who have had a damaged vehicle in the past want to acquire vehicle insurance**



Age

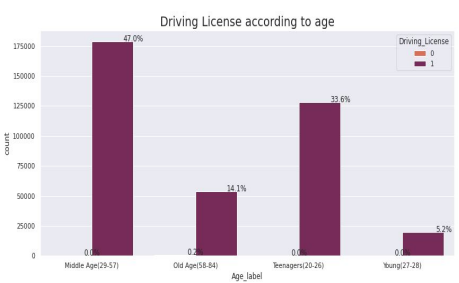
- The dataset has more individuals with an age of 24.
- 40 to 60-year-olds had a higher likelihood of purchasing vehicle insurance.
- From the above boxplot we can see that there no outlier in the dataset.

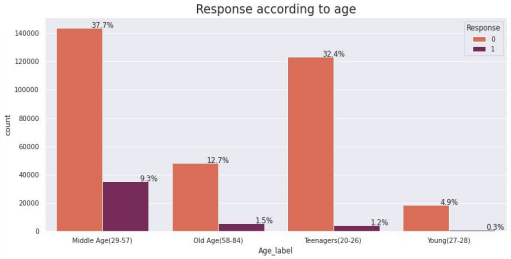


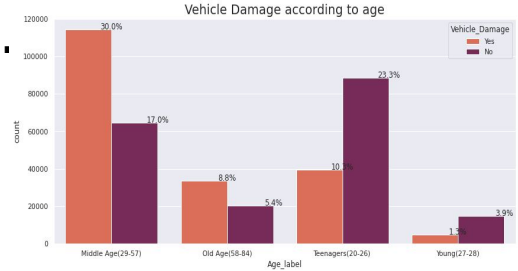
Age According to Response

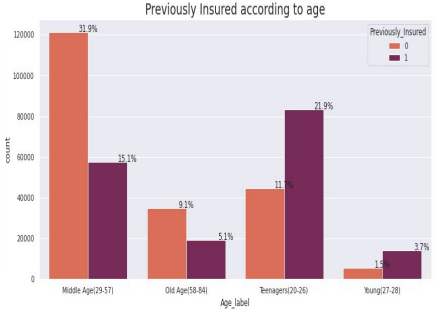
According to age wise:

- **9.3% of people in their middle age people are interested in purchasi insurance.**
- **Almost 47% of middle-aged individuals have a driver's licence.**
- **About 21.9% of people in their teens have health insurance.**
- **Around 21.9% of persons in their teens have insurance previously.**





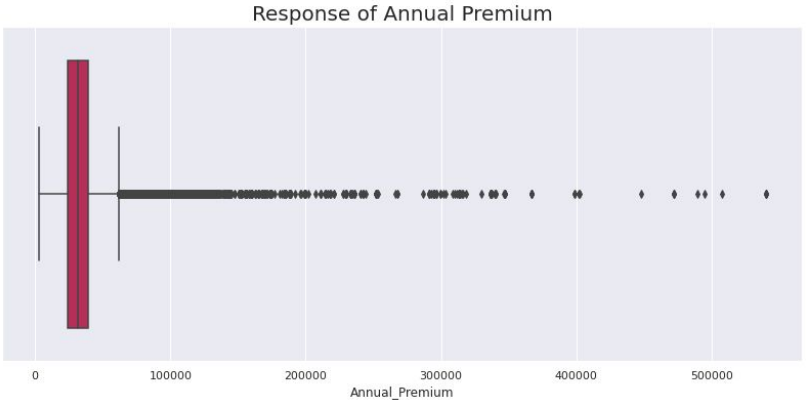




Outliers

Handling Outliers:

- **For identifying outliers, scatter plots and box plots are the most used visualisation techniques.Here we use boxplot.**
- **We have used the quantile approach to address outliers and eliminate Outlier.**

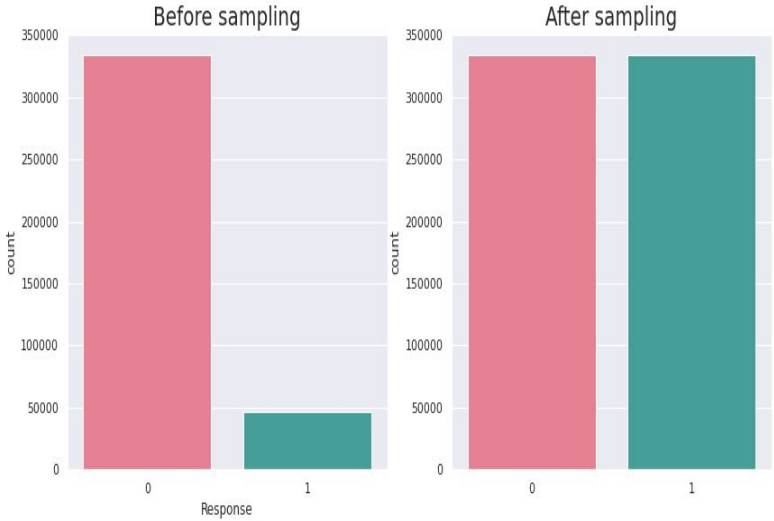


Feature encoding:

- We have encoded gender,vehicle damage and vehicle age which where of object type

Model Building

- As we can see from the figure, target variable response are not balanced.
- After oversampling data is balanced



Machine Learning Algorithms

AI

Let's try various machine learning models on our data set to see how they each perform.

- **Logistic Regression**
- **Decision Tree**
- **Random Forest**
- **KNN**
- **Gradient Boost**
- **XGBoost**
- **LightGBM**

Result of all models

- **Comparing the result of all the model Random Forest gives the best Result.**
- **We can deploy this models.**

model_name	Recall_Score	Precision_Score	f1_Score	Accuracy_Score	ROC_AUC Score
Logistic Regression	0.933993	0.717003	0.811238	0.782277	0.810925
Decision Tree	0.843728	0.843728	0.826262	0.822264	0.822224
Random Forest	0.893616	0.802839	0.845799	0.836782	0.836678
KNN	0.883408	0.780077	0.828533	0.816841	0.816719
Gradient Boosting	0.918769	0.760795	0.832353	0.814608	0.814417
Extreme Gradient Boosting	0.916610	0.769674	0.836740	0.820828	0.820652
LGBM	0.915028	0.765703	0.833732	0.817185	0.817005

Summary and Conclusion

AI

- The gender variable in the dataset is evenly distributed, with 50.5% of vehicles having past damage and 12.0% of people who have had a damaged vehicle wanting to acquire vehicle insurance. 99.8% of customers have DL, while 0.2% do not have DL.
- Vehicle age increases, making it more important to buy insurance to reduce risk.
- Middle-aged individuals are more likely to purchase insurance, with 47% having a driver's licence and 21.9% having vehicle insurance.
- Further, we applied Machine Learning Algorithms to determine whether a customer would be interested in Vehicle Insurance. Random Forest is the model that performs the best.

Thank You