horizontal line

Project Summary

|  |  |
| --- | --- |
| Batch details | PGPDSE-FT Chennai September 21 |
| Team members | Mihir.V. Joshi  Alan David.S  Kishore.P. G  Thomas Loyola  Dayanidhi |
| Domain of Project | Marketing Analytics |
| Proposed project title | Health Insurance Cross Sell Prediction |
| Group Number | GRP 5 |
| Team Leader | Mihir.V. Joshi |
| Mentor Name | Vikash Chandra |

Date: 07/01/2022

Signature of the Mentor Signature of the Team Leader

Mihir.V Joshi

# 

Table of Contents

|  |  |  |
| --- | --- | --- |
| Sl NO | Topic | Page No |
| 1 | Overview | 3 |
| 2 | Business problem goals | 3,4 |
| 3 | Topic survey in depth | 4 |
| 4 | Critical assessment of topic survey | 5 |
| 5 | Methodology to be followed | 5,6,7 |
| 6 | References | 7 |

Project Details

**OVERVIEW**

Cross-selling involves selling products to existing customers. It is one of the highly effective techniques in the marketing industry. The main objective behind this method is to increase sales revenue and profit from the already acquired customer base of a company.

**Business Problem Understanding**

Insurance company wants to understand whether a customer would be interested in Vehicle Insurance as it would be extremely helpful for the company because the company can then accordingly plan its communication strategy to reach out to those customers and optimize its business model and revenue.

**Business Objective**

To find the right customers who would be interested in Availing vehicle insurance.

**Approach**

The structured thinking approach will help us here. Let me state some hypotheses from our problem statement.

1. Male customers are more tend to buy vehicle insurance than females.
2. The middle-aged customers would be more interested in the insurance offer.
3. Customers having a driving license are more prone to convert.
4. Those with new vehicles would be more interested in getting insurance.
5. The customers who already have vehicle insurance won’t be interested in getting another.
6. If the Customer got his/her vehicle damaged in the past, they would be more interested in buying insurance.

.

**Conclusion**

Cross-sell prediction is a very common machine learning problem that is relevant in the industry. This can be solved by building a machine learning model which would help company in approaching the right customers.

**Problem understanding**

With the help of the below mentioned Features, the objective is to predict whether a customer will avail for the health insurance or not.

Parameters: Gender, age, Driving License, Region Code, Previously Insured, Vehicle Age, Annual Premium, Policy Sales channel, Vintage

**Current Solution to the problem:**

The Company needs to conduct a survey for vehicle insurance for the below details.

Gender, age, Driving License, Region Code, Previously Insured, Vehicle Age, Annual Premium, Policy Sales channel, Vintage.

**Proposed Solution to the problem:**

Build an efficient machine learning model which would predict the right customers who would be interested in availing vehicle insurance.

**Reference to the problem:**

[**https://www.duckcreek.com/blog/predictive-analyitics-reshaping-insurance-industry/**](https://www.duckcreek.com/blog/predictive-analyitics-reshaping-insurance-industry/)

We have referred to the above link to understand how predictive analytics in insurance can be done.

**CRITICAL ASSESSMENT OF TOPIC SURVEY:**

1.Find the key area, gaps identified in the topic survey where the project can add value to the customers and business.

Key areas are to find whether a customer has vehicle insurance or not, Age of the vehicle, Past Vehicle Damage, Premium Amount Paid for a year.

2. What key gaps are you trying to solve?

Since the company is starting its first vehicle insurance policy scheme. The company has to reach out to the right customers which would be challenging.

# METHODOLOGY TO BE FOLLOWED

**Business Understanding:**

Insurance company wants to understand whether a customer would be interested in Vehicle Insurance as it would be extremely helpful for the company because the insurance company can then accordingly plan its communication strategy to reach out to those customers and optimize its business model and revenue.

**Data Understanding:**

The data consist of 381k rows and 12 columns. Below is the column description.

|  |  |
| --- | --- |
| Feature Name | Feature Description |
| id | Unique ID for the customer |
| Gender | Gender of the customer |
| Age | Age of the customer |
| Driving\_License | 0: Customer does not have DL, 1: Customer already has DL |
| Region\_Code | Unique code for the region of the customer |
| Previously\_Insured | 1: Customer already has Vehicle Insurance, 0: Customer doesn't have Vehicle Insurance |
| Vehicle\_Age | Age of the Vehicle |
| Vehicle\_Damage | 1: Customer got his/her vehicle damaged in the past. 0: Customer didn't get his/her vehicle damaged 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 i.e. Different Agents, Over Mail, Over Phone, In Person, etc. |
| Vintage | Number of Days, Customer has been associated with the company |
| Response | 1: Customer is interested, 0: Customer is not interested |

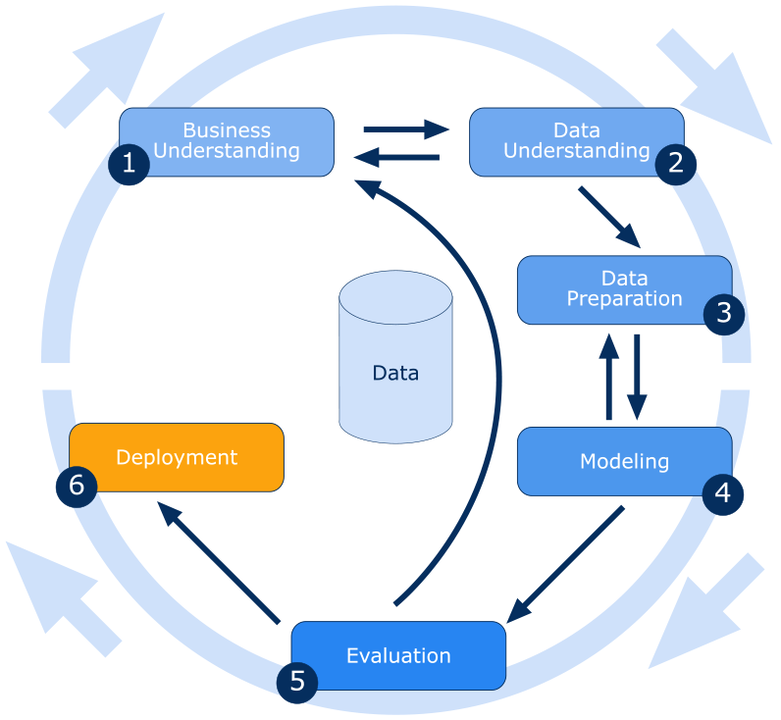
**Data Preparation:**

1.Missing value treatment

2.Outlier Treatment

3.Encoding Categorical Data

4.Feature Scaling



**Modeling:**

1.Build the base model for various algorithms

2.Identify the significant Features.

3.Identifying the best hyperparameters for the corresponding models.

4.Build the model using the obtained significant features and the hyperparameter values.

**Evaluation:**

Since we are solving a binary classification problem below are the various metrics, we could consider for evaluating the performance of the model

1.Accuracy

2.AUC score

3.Confusion Matrix, Classification Report

**Reference:**

[**https://www.mckinsey.com/industries/financial-services/our-insights/insurance-blog**](https://www.mckinsey.com/industries/financial-services/our-insights/insurance-blog)

**Notes For Project Team**

*Sample Reference for Datasets (to be filled by team and mentor)*

|  |  |
| --- | --- |
| Original owner of data | Anmol Kumar |
| Data set information | 381k rows  12 columns |
| Any past relevant articles using the dataset | <https://www.analyticsvidhya.com/blog/2021/09/cross-sell-prediction-using-machine-learning-in-python/> |
| Reference | <https://www.mckinsey.com/industries/financial-services/our-insights/insurance-blog> |
| Link to web page | <https://www.kaggle.com/anmolkumar/health-insurance-cross-sell-prediction/activity> |

### 

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*