

Data Science Training - Assignment

Health Insurance Cross Sell Prediction

Health Insurance Cross Sell Prediction.

Project Description:

The 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 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.

For example, you may pay a premium of Rs. 5000 each year for a health insurance cover of Rs. 200,000/- so that if, God forbid, you fall ill and need to be hospitalised in that year, the insurance provider company will bear the cost of hospitalisation etc. for up to Rs. 200,000. Now if you are wondering how company can bear such high hospitalisation cost when it charges a premium of only Rs. 5000/-, that is where the concept of probabilities comes in picture. For example, like you, there may be 100 customers who would be paying a premium of Rs. 5000 every year, but only a few of them (say 2-3) would get hospitalised that year and not everyone. This way everyone shares the risk of everyone else.

Just like medical insurance, there is vehicle insurance where every year customer needs to pay a premium of certain amount to insurance provider company so that in case of unfortunate accident by the vehicle, the insurance provider company will provide a compensation (called 'sum assured') to the customer.

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, in order to predict, whether the customer would be interested in Vehicle insurance, you have information about demographics (gender, age, region code type), Vehicles (Vehicle Age, Damage), Policy (Premium, sourcing channel) etc.

Data Description:

Feature	Description
ID	Unique ID for the customer
Gender	Gender of the customer
Age	Age of the customer
Driving_License	Customer does not have DL, 1 : Customer already has DL
Region_Code	Unique code for the region of the customer
Previously_Insured	Customer already has Vehicle Insurance, 0 : Customer doesn't have Vehicle Insurance
Vehicle_Age	Age of the Vehicle
Vehicle_Damage	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
PolicySalesChannel	Anonymized Code for the channel of outreaching to the customer ie. Different Agents, Over Mail, Over Phone, In

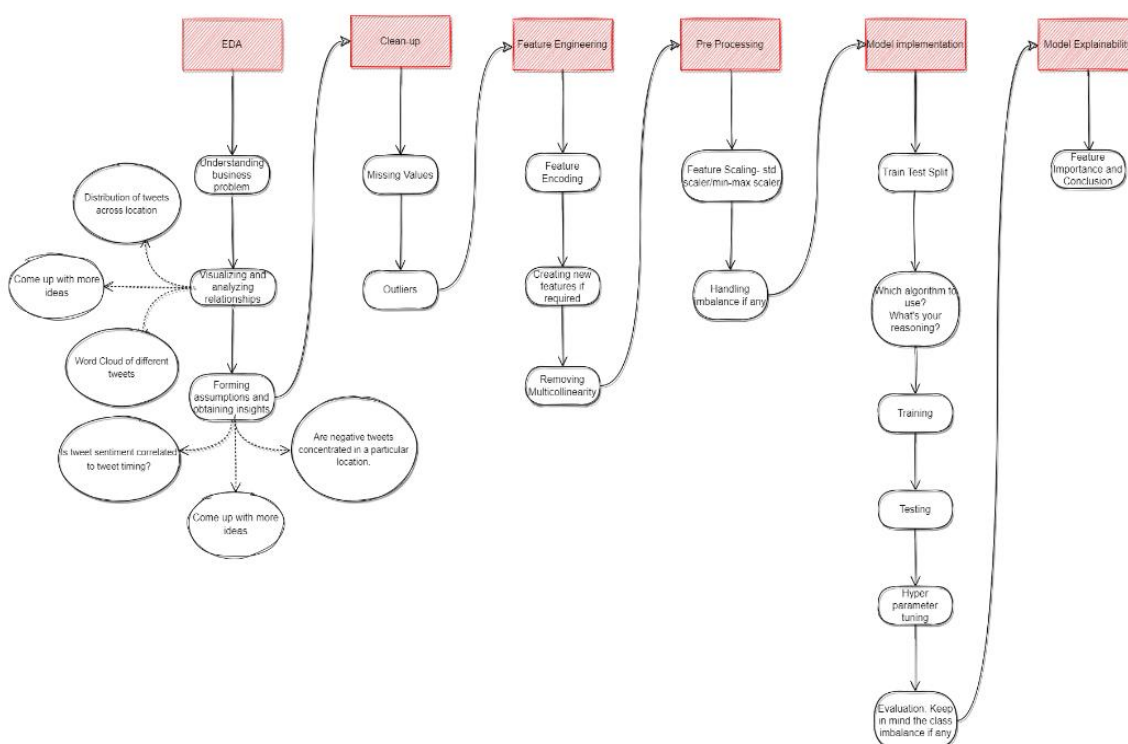
Person, etc

Vintage	Number of Days, Customer has been associated with the company
Response	Customer is interested, 0 : Customer is not interested

Main libraries to be used:

- Pandas for data manipulation, aggregation
- Matplotlib and seaborn for visualization and behaviour with respect to the target variable
- NumPy for computationally efficient operations
- SciKit learn for model training, model optimization and metrics calculations.

Project Architecture:



Project Evaluation Criteria:

- Efficient approach to problem statement.
- Data Exploration techniques and logic
- Logic of handling missing values and outliers
- Visualization logic
- Forming insights and explaining your understandings
- Understanding of how your project is to stakeholders.

JM//N
GROUP

