

School of Engineering and Applied Science
CSE523- Machine Learning

Project Report - 1

Vehicle Insurance Predictor

Submitted to: **Prof. Mehul Raval**

Date of submission: **9th Feb 2021**

Members:

Name	Roll Numbers	Branch
Samarth Shah	AU1841145	B. Tech, ICT
Priyank Sangani	AU1841136	B. Tech, ICT
Yash Patel	AU1841141	B. Tech, ICT
Shaili Gandhi	AU1841012	B. Tech, ICT

Task performed this week:

- Study of the 4 domains
- Problem framing
- Topic exploration
- Dataset search
- Background research for the selected topic

Outcomes of the tasks performed:

- Problem Definition: An Insurance company wants to know whether their customers are interested in vehicle insurance provided by them.
- Ideal Solution: Build a model to predict whether the customer is interested in vehicle insurance or not.

- Understanding of the problem: Here we consider various features like gender of the customer, age of the customer, whether the customer have license or not, regional code of the customer's region, whether the customer have old insurance or not, age of the vehicle, whether the vehicle was previously damaged or not, premium amount for insurance, no. of days the customer has been associated with the company, the channel of outreaching to the customer ie. Different Agents, Over Mail, Over Phone, In Person, etc.

- **Data-set Description:**

Data type	Variable	description
int	ID	Unique Id for customer
string	Gender	Gender of the customer
int	Age	Age of the customer
int	Driving_License	0 : Customer does not have DL, 1 : Customer already has DL
int	Region_Code	Unique code for the region of the customer
int	Previously_Insured	1 : Customer already has Vehicle Insurance 0 : Customer doesn't have Vehicle Insurance
int (range)	Vehicle_Age	Age of the Vehicle
int	Vehicle_Damage	1 : Customer got his/her vehicle damaged in the past. 0 : Customer didn't get his/her vehicle damaged in the past.
int	Annual_Premium	The amount customer needs to pay as premium in the year

int	Policy_Sales_Channel	Anonymized Code for the channel of outreaching to the customer ie. Different Agents, Over Mail, Over Phone, In Person, etc.
int	Vintage	Number of Days, Customer has been associated with the company
int	Response	1 : Customer is interested, 0 : Customer is not interested

Tasks to be performed in the upcoming week:

- Pre-process the dataset and try to get more of the numeric values from categorical features.
- We can plot some of the visualizations to get better interpretation of features mentioned.
- Also we will try to search for more features if possible and modify the dataset if needed.