Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Name : Puneet Subhanji

Email : puneetsubhanji@gmail.com

- **Data understanding**
  Data Analysis
  Data pre processing and exploration
- **Feature Analysis**
  Univariate
  Bivariate and multivariate analysis
- **Feature engineering**
  Null value check
  Random sampling
  Feature creation , etc
- **Data visualisation**
- **Algorithm implementation**
  Logistic Regression
  Random forest classifier
  XGBClassifier
- **Research Analytics**
  Technical documents

Please paste the GitHub Repo link.

Github Link:- https://github.com/PUNEETSUBHANJI/Health_insurance_cross_sell_prediction

Please write a short summary of your Capstone project and its components. Describe theproblem statement, your approaches and your conclusions. (200-400 words)

PROBLEM

Our client is an insurance firm that has offered Health Insurance to its clients and now needs assistance in developing a model to forecast if the consumers from the previous year will be interested in the company's Vehicle Insurance.

APPROACH

- Initially, imported the data set in order to do descriptive analysis on the data set in to understand the information of data accessible.
- The data set supplied was cross-checked for missing and repeated values.
- Exploring all the variables of the data set (such as Id, Age, Vehicle, vehicle age, etc) concerning opting for vehicle insurance, to determine and comprehend the elements influencing the decision to get vehicle insurance.
- Data visualization with various plots was used to investigate the relationship between different factors.
- Encoding of categorical columns and fitting the different models, we used the below algorithms:-
  Logistic Regression
  Random forest classifier
  XGB classifier
- Then tuning into Hyperparameters and perfomance evaluation to identify best fit Model.

CONCLUSION

- More customers between the ages of 30 and 60 are likely to purchase insurance
- Vehicle insurance is not interesting to anyone under the age of 30. The lack of involvement could be a factor, they may not yet have expensive vehicles and have little knowledge about insurance.
- Consumers with 1-2-year-old vehicles are more interested as compared to others.
- Consumers with less than 1-year-old Vehicles have very little chance of purchasing Insurance
- Customers with driver license are more likely to get insurance.
- Vehicle damage customers are more likely to purchase insurance.
- The variable such as Age, Previously_insured, Annual_premium are more affecting the target variable.
- Comparing the ROC curve we can see that the Random Forest model performs better. Because curves closer to the top-left corner indicate better performance.