Capstone Project Submission

Health Insurance Cross-Sell Prediction

Introduction:

Cross-Selling is a new marketing strategy based on data analysis, which found that different needs exist as well, who can become customers and meet their needs through sales of various related services or products.

We are utilizing the Decision Tree algorithm, Logistic Regression, and Random Forest for building different prediction models. Decision tree algorithms build nodes based on certain decisions and can be very useful for random data with no specific distribution. Logistic Regression is a parametric algorithm; hence, certain properties are to be verified for good results. The Random Forest engenders decision trees on randomly selected data samples, gets predictions from each tree, and selects the best solution by means of voting. Here we have features like age, driving license, and vehicle age, which will give a relationship with the response variable.

Problem Statement:

Our client is an insurance company that has provided Health insurance to its customers. Now they require our availability in building a model to prognosticate whether the policyholders from the past year will additionally be intrigued with the Vehicle insurance provided by the company. An insurance policy is an arrangement by which a company undertakes to provide an assurance of emolument for a designated loss, damage, illness, or death in reciprocation for the payment of a designated premium. A premium is a sum of profit that the customer needs to pay customarily to an insurance company for this assurance.

Abstract :

Cross-selling is a great way to make more money for any insurance agency without starting from scratch. We can build a business from the book you already have with your current customer relationships. It is not only benign for the company but withal for the clients. It is not only a matter of making more profit. But it is withal about integrating value and bringing solutions to the indemnification-cognate challenges your customers face. Just like medical insurance, there is vehicle insurance where every year the 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 compensation (called 'sum assured') to the customer. Our dataset is based on the Health Insurance Customers database. This Experiment can help to understand what can be affecting factors for cross-selling Insurance Plans for already existing customers. The model can be used for any insurance plan dataset to predict cross-selling.

Keywords: machine learning, Supervised machine learning, Cross-selling, Predictive Model building, Logistic Regression, Random Forest

Conclusion:

- Due to the Response variable's value of 1 being much lower than its value of 0, the provided dataset is an imbalanced dataset.

- Compared to their female counterparts, male consumers own a little bit more vehicles and have a higher likelihood to get insurance.

- Customers between the ages of 30 and 60 are the most likely to get insurance whereas Vehicle insurance is not interesting to anyone under the age of 30. The lack of involvement, a lack of knowledge about insurance, and possibly the lack of expensive vehicles are potential causes.

- Customers with driving licenses are more likely to purchase insurance

- Compared to consumers with vehicles less than one-year-old, those with vehicles between one and two years old are more interested in purchasing insurance.

- Due to their personal experience with the costs associated with vehicle repairs, customers with vehicle damage are more likely to purchase insurance.

- The variable such as Age, previously insured, and Annual premium is more affect the target variable.

GitHub Repo link - https://github.com/md-kamran98/HEALTH-INSURANCE-CROSS-SELL-CTION/blob/main/Md_Kamran's_classification_ML_Submission.ipynb

GitHub Link - https://github.com/md-kamran98