

Midterm Project Proposal

Jingwen Xu U77877194

Personal Statement:

We have made the virtual resume for our future career goal in MA675. In the career plan, I prepare to be a statistics consultant when I graduate from MSSP and try to become an actuary three years later. Based on this plan, I found a large dataset about insurance which combined the characteristics of both of my two career objects – statistical and insurance related. The detailed introduction of the dataset is in the link of data source.

From this dataset, I can practice my skills that I will operate for hundreds of times as a statistics consultant such as data cleaning, exploratory data analysis and data modeling. Besides, it can help me a lot to have more precise evaluate on insurance risk as an actuary. Actuaries are professionals who analyze risks and quantify their financial impact. I think that I should have a deep grasp of the information from the insurance data at first so that I can have an effective risk analysis.

Question:

- (1) Find the mathematical relationship between all kinds of predictor (Previously_Insured, Vehicle_Damage, Annual_Premium and so on) and the indicator whether the customers will be interested in the vehicle insurance.
- (2) Make a prediction of the attitude of customer towards the vehicle insurance given a hypothetical new customer.

The data source:

<https://www.kaggle.com/anmolkumar/health-insurance-cross-sell-prediction?select=train.csv>

Proposed Timeline of work:

- (1) EDA: By 2020/11/15
- (2) Data processing: By 2020/11/20
- (3) Modeling and Validation: By 2020/11/25
- (4) Write up: By 2020/12/01