STA 2453: Project 2 - Bankruptcy Report

March 23, 2023, University of Toronto

Vignesh Edithal

Problem 1

Model development scheme and training

Data pre-processing

- City and Province are concatenated to get Address column which is then frequency encoded such that top 9 frequent addresses are mapped from 1 to 9 and others 10.
- AmountOfInsurance and Earnings are transformed by first dividing by 1000 then taking log base 10.
- BusinessType is encoded based on categorical codes.
- The 250k rows of policies data with the above transformation are clustered into 12 categories (corresponding to best BIC score) using GaussianMixture API from sklearn.
- Policies data and Claims data are merged using outer join on BusinessID.
- Rows are de-duplicated and NumClaims and ExpectedSeverity columns are added by aggregating over duplicate BusinessID rows.

NumClaims Model

- All 250k rows are considered for training data, NumClaims for businesses which did not claim is set to 0.
- XGBRegressor API from sklearn was fit on training data with 5 fold cross validation and grid search for hyperparameter tuning.
- Best hyperparameters found, n_estimators: 200, learning_rate: 0.1, max_depth: 20, subsample: 0.6, colsample_bytree: 0.95

ExpectedSeverity Model

- Only 8472 rows are considered for training data, these are the business which have claimed at least once.
- XGBRegressor API from sklearn was fit on training data with 5 fold cross validation and grid search for hyperparameter tuning.
- Best hyperparameters found, n_estimators: 200, learning_rate: 0.1, max_depth: 20, subsample: 0.6, colsample_bytree: 0.95

Inference

- ExpectedSeverity and NumClaims are predicted and plugged into the Loss Cost formula with given constants
- ProfitLoading which can range from 0 to 0.6 was set to a value of 0.3

Problem 2

Model performance statistics

- Out of 2721 customers, I was awarded 577 (around 25%) customers. This is a lot of customers given there were 18 competitors in the market
- The highest loss I have endured is $559k
- I have endured 67 losses and the rest customers resulted profit. However, the total loss is $726k

Problem 3

Future Direction and Business Strategies

Business Strategies

- It is a good idea to overcharge to minimize risk.
- The goal of insurance premium modelling is to minimize risk as much as possible.
- One way of minimizing risk is to manually increase the premium of risky customers (customers with expected severity over a certain threshold)

Future directions to prevent bankruptcy

- I increased my profit loading from 0.3 to 0.5 and generated new premiums.
- The premiums csv was downloaded and my premiums were updated to new values.
- dirchlet_market.py script was run with new market values values to identify which customers got awarded to me.
- Upon analysis of profit it was found that I made a profit of $1,003,871