Analytics Project
SCMA 648
Virginia Commonwealth University
Fall 2020

1 Summary

This project will give you the opportunity to apply the methods we have covered in class to a business problem. The project is intended to give you experience in applying analytics methods for data analysis in the context of a business case, using analytics software, and communicating results.

You are to form two-person teams with another member of the class. Send an email with your team members to the instructor at `jpbrooks@vcu.edu` by October 29th or indicate that you have no partner.

The project consists of 2 tasks:

- Broker segmentation
- Gross written premium prediction

Your team will submit a final report with the following structure: executive summary, problem introduction, description of preprocessing and methods, results, conclusions, and appendices.

The executive summary should be no more than one page and should introduce the problem and the people involved (in the context of the case), give an overview of the approach taken, summarize results, and make recommendations. Resist the temptation to save major results for later in the report - the executive summary should clearly and concisely state all major findings. The function of the remainder of the report is to provide further details supporting the recommendation in the executive summary.

The report should be a narrative written in sentence-paragraph form and not simply an itemized list. The report is limited to 10 pages (double spaced, 12 point font). The report should be written using complete sentences and correct grammar. Submit your reports in pdf format via Canvas. Turn in your code and data as separate files.

2 Analytics Tasks

For each of the tasks, generate additional features as measures of broker performance. The case description contains suggestions of additional features. Multiple sets of features can be generated based on using different time horizons. Also, you may wish to generate additional features that summarize performance over time. Describe the features that you generate in your report and describe which features (both provided and generated) that you use for which modeling steps. Be sure to apply appropriate treatment of categorical features.

2.1 Broker Segmentation

Apply a clustering method at least three years of information ending with 2019 to cluster brokers into five segments. Describe the characteristic properties of each segment. Provide a visualization of your broker segmentation using principal component analysis and describe the clusters in terms of the components.

2.2 Gross Written Premium Prediction

Based on at least three years of information ending with 2018, create a model to predict whether the gross written premium will increase or decrease for 2019. Be sure to apply appropriate treatment of categorical variables and to use proper practices in the empirical evaluation of predictive methods.

Apply at least two methods for the classification (e.g., classification trees and SVM) and compare the methods. The methods that you use do not necessarily need to be methods that we covered in class. You must use appropriate cross-validation techniques. Tune the parameters for at least one of the techniques. The methods should be tested using the same test data. Report on which features are important for prediction.

Adapt your model to predict whether the gross written premium will increase or decrease in 2020 based on the information available through 2019. Submit your predictions as a comma-delimited text file in the following format:

```
broker_id, prediction
254, 0.75
255, 0.55
256, 0.76
257, 0.24
...,...
```

The number in the second column for a broker indicates the probability that the gross written premium will increase in 2020. Missing values for gross written premium will be considered as zeros.

3 Important dates

Due Date	Task
October 29th	Project teams
November 24th	Final Report