

Case Study for Insurance Modeling

Jeanne Fu

May 7, 2024

1 Introduction

This case study will focus on pricing strategies for a commercial auto line of business. The goal of the study is to incorporate Machine Learning to improve pricing accuracy, while maintain explainability.

Throughout the life cycle of a insurance pricing project, there are in general six steps, and we will discuss in more detail how to use ML in each step:

- Scoping (Set up a goal for the project, success criteria and/or metrics, resource, cost, etc.)
- Data Preparation (Data sourcing, EDA, data cleaning)
- Feature Engineering (Transformation, variable selection)
- Modeling (Benchmarking, feature importance ranking, model, validation)
- Implementation (Deploy model for business end-users, testing)
- Monitoring (and prepare for model refreshing)

This case study will be focusing on Data Preparation, Modeling and Implementation. I will discuss about potential risks and how to manage them in the end.

2 Data Preparation

Put together both internal and external data sources, and prepare for modeling data.

Internal data source:

- Exposure or premium data: account information, location, primary usage of vehicle, business industry
- Loss data: (Target, loss history)
- Vehicle information: vehicle weight, vehicle age, vehicle value (cost new), etc. May use external VIN decoding service if needed.
-

External data source:

- Driver information (credit history, police driving record, etc.)
- Credit history (for the business)

Things to do: - explore missing values - Check distribution

3 Feature Engineering

A lot of explorations could be done here. Include but not limited to:

- Encoding categorical variables
- Explore feature interaction
- Explore feature

4 Model Training and Validation

* Split Training, Test, Holdout (or Cross Validation)

* Selection of Target variables: Frequency and Severity VS. Pure Premium

Things to think about:

- Different coverages

- Outlier in pure premium
- Need to develop loss (alternative to)
- Need to trend

* AutoML for model Benchmarking

* AutoML for model variable importance (Shapeley Value)

5 Potential Risks and How to Manage Them

6 Conclusion