Case Study for Insurance Modeling

Jeanne Fu

May 6, 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.

Toy data: https://www.kaggle.com/datasets/noordeen/insurance-premium-prediction/data

3 Feature Engineering

4 Model Training and Validation

5 Potential Risks and How to Manage Them

6 Conclusion