

A Comparison of Gradient Boosting Machines and Generalized Linear Models for Non-Life Insurance Pricing

Alexander Eriksson*

June 2021

Abstract

The pricing of an insurance contract is a crucial area for risk assessment in the insurance business. Actuaries typically apply statistical methods to perform this task, which is known as rate making. This study puts focus on claim frequency modelling, an important part within non-life insurance pricing, with the application of gradient boosting machines (GBMs). GBMs are a family of statistical learning methods which have grown in popularity because of their strong performance in numerous disciplines. By using simulated data, inspired by real data provided by the insurance company Hedvig, we compare GBMs with generalized linear models (GLMs), which typically are applied by pricing actuaries. In insurance pricing, transparency and interpretability of models are key aspects for actual business application. Therefore, we also investigate available tools for model interpretation and use them for interpreting the GBMs. We further explore the possibility of creating an improved GLM based on the potential insights provided by the GBM. We show that the GBM outperforms the GLM in terms of both prediction accuracy and ranking of the claim frequency risk if we are not aware of present interaction effects. By using variable importance, partial dependence plots and Friedman's H-statistic we also show that we can gain an understanding of how the GBM models work and how we can extract insights from them. Further, the insights provided by the GBM guided us in the right direction towards creating GLMs with improved model performance.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: alexandereriksson88@gmail.com. Supervisor: Filip Lindskog.