Comparing and Intereting Machine Learning Algorithms Estimating Technical Prices

Joakim Bilyk, Sebastian Cramer and Teis Blem *University of Copenhagen*

This document provides a practical example of the application of supervised machine learning algorithms to car insurance data using the mlr3 package. In non-life insurance pricing, a popular model is the frequency-severity model, which decomposes the price into the product of the probability of a claim arising and the expected claim size given a claim occurs. This paper argues that a tree-based model is well-suited to the frequency-severity model, as it can capture complex nonlinear relationships between risk factors and claims. To interpret the model's estimates, we used shapley values to gain insights into the relative importance of each risk factor. Finally, we used a decomposition technique to debias the price model and ensure it does not discriminate based on gender. Overall, our approach demonstrates the potential of machine learning to create more accurate and equitable pricing models in the insurance industry.

Keywords: mlr3, machine learning, regression, non-life insurance, estimating technical price, XGBoost, debiassing, bias, SHAP-values, interpretation of ML models

Contents

Getting familiar with the data

dsad (Lützen 2019)

Lützen, Jesper. 2019. *Diskrete Matematiske Metoder*. 2nd ed. Copenhagen, Denmark.