



Mathematical Statistics
Stockholm University
Master Thesis **2024:11**
<http://www.math.su.se>

Prediction of Non-life Cancellation Rates with Machine Learning

Lei Sun*

September 2024

Abstract

The purpose of this thesis is to discuss estimation of cancellation rates for non-life insurances, including property, home, car insurances, through the application of machine learning techniques. Traditional methods often rely on a limited set of variables and may overlook other factors that significantly influence cancellation rates. By using machine learning techniques, particularly tree-based methods like Gradient Boosting Machines (GBM) and Generalized Linear Models (GLM), we aim to enhance the accuracy of these predictions by considering a broader range of influencing factors, such as insurance price and sales channel.

To address the challenge of imbalanced datasets, the Synthetic Minority Over-sampling Technique (SMOTE) was used to generate more balanced datasets. These balanced datasets enabled more effective model training. The thesis compares the performance of GLM and GBM models using both original and SMOTE-enhanced data. The results indicate that while SMOTE improves the identification of cancellations, it may also introduce noise, affecting overall predictive accuracy.

The study concludes that each model offers distinct advantages depending on the specific focus of the analysis. The findings underscore the potential of machine learning to refine the estimation of cancellation rates, offering valuable insights for the insurance industry to develop more effective strategies for customer retention and risk assessment.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: lisa.solen@gmail.com. Supervisor: Mathias Millberg Lindholm.