



Mathematical Statistics
Stockholm University
Master Thesis **2021:6**
<http://www.math.su.se>

Application and Comparison of Machine Learning and Traditional Methods to Insurance Pricing in Scarce Data Environments

Asmir Prepic*

June 2021

Abstract

Complex models and methods has received plenty attention over the recent years and various authors have shown the power of e.g. neural networks and random forests over traditional insurance pricing models. This thesis investigates the predictive power for a simulated insurance portfolio where there is less exposure among policyholders who have higher risk by utilising a synthetic minority oversampling technique (SMOTE) and comparing the predictive performance without application of SMOTE. In addition the same comparison is applied to a real insurance data set. The thesis shows that without SMOTE and where there is clearly less exposure among high risk customers compared to the rest of the portfolio, the traditional vanilla GLM outperforms the more complex models in predictive power. On the contrary, by utilizing SMOTE and oversampling the high risk policyholders such that the data is more balanced, neural networks, regression trees and random forests make better prediction based on the 10 fold cross validation technique.

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