

ABSTRACT

The challenge addressed in this dissertation arises from the predominant use of the Generalised Linear Model (GLM) to create tariffs structures in the insurance market. With the advance of Machine Learning (ML) techniques, opportunities arise to improve the accuracy and predictive capacity of the models used in the insurance sector. Machine Learning models, such as the Gradient Boosting Machine (GBM), stand out for their robustness in capturing non-linear patterns and complex interactions between variables. However, despite the predictive superiority that these models tend to offer, the challenge of interpretability becomes a significant obstacle. This work aims to contribute to the debate on the incorporation of Machine Learning models in the insurance market, using as a case study the comparison between GLM and GBM in the modelling of claims frequency and severity. The aim is to compare the two models in terms of predictive effectiveness, assessing how each model responds to the dataset, and what ability is to accurately predict future results. It is important to note that in this dissertation we do not aim to create a tariff structure, but rather to analyse the relative performance of the two models. The intention is to provide a solid basis for comparison, enabling a better understanding of how Machine Learning models can complement or improve the GLM outputs. Through this analysis, we hope to contribute to a more informed discussion on the adoption of Machine Learning methods in the insurance market.

Keywords: GLM, GBM, Cluster, Non-Life Insurance, Automobile Insurance, Pricing Structure

RESUMO

O problema abordado nesta tese surge do uso predominante dos Modelos Lineares Generalizados (MLG) para a criação de tarifas no mercado segurador. Com o avanço das técnicas de Machine Learning (ML), surgem oportunidades para aprimorar a precisão e a capacidade preditiva dos modelos utilizados no setor segurador. Modelos de Machine Learning, como o Gradient Boosting Machine (GBM), destacam-se pela sua robustez na captura de padrões não-lineares e interações complexas entre variáveis, algo que o MLG, por ser linear na sua estrutura, não consegue captar com tanta eficácia. No entanto, apesar da superioridade preditiva que esses modelos tendem a oferecer, o desafio da interpretabilidade torna-se um entrave significativo. Este trabalho visa contribuir para o debate sobre a incorporação de modelos de Machine Learning no mercado segurador, utilizando como caso de estudo a comparação entre MLGs e GBMs na modelagem de frequência e severidade de sinistros. O objetivo desta dissertação é comparar os dois modelos em termos de eficácia preditiva, avaliando como cada um responde ao conjunto de dados, e qual a sua capacidade de prever resultados futuros com precisão. É importante salientar que esta dissertação não tem como objetivo criar uma tarifa, mas sim analisar o desempenho relativo dos dois modelos. A intenção é proporcionar uma base sólida de comparação, que permita entender melhor como os modelos de Machine Learning podem complementar ou melhorar o modelo GLM. Através desta análise, espera-se contribuir para uma discussão mais informada sobre a adoção de métodos de Machine Learning no mercado segurador.