Abstract

This thesis explores the application of machine learning models for detecting fraudulent claims in motor insurance. It compares the effectiveness of several algorithms, including logistic regression, random forest, XGBoost, histogram-based gradient boosting, and multilayer perceptron (MLP). The study addresses the challenge of class imbalance in fraud detection, utilizing techniques such as Synthetic Minority Over-sampling Technique (SMOTE) and class weighting to enhance model performance. Real-world data provided by UNIQA pojišťovna a.s., including detailed information on insurance contracts and claims, serve as the basis for the empirical analysis. Among the models tested, XGBoost with SMOTE resampling and class weighting achieved the highest recall rate, detecting over 90% of fraudulent claims, while maintaining a reasonable level of precision. The feature importance analysis highlighted key predictors of fraud, such as claim amount, type of coverage or vehicle age. The findings underscore the potential of advanced machine learning techniques to improve the efficiency of fraud detection systems in the insurance industry.

JEL Classification	C49, G22, K42,
Keywords	machine learning, fraud detection, insurance, unbalanced data
Title	Machine Learning Methods in Motor Insurance Fraud Detection

Author's e-mail	85952280@fsv.cuni.cz
Supervisor's e-mail	barunik@fsv.cuni.cz