Can machine learning algorithms outperform traditionally used methods in insurance pricing?

Bor Harej
Prime Re Solutions
bor.harej@prs-zug.com

The comparison of different algorithms for insurance pricing exercise is a task that relies heavily on the data sample used. There are two options: real data and synthetic data. A critical issue with the real data is the lack of information of the exact underlying rate that we want to predict. That makes the comparison of different algorithms on real data less clear. We decide for an analysis on synthetic data to avoid this issue. Based on synthetic data sample, several different machine learning algorithms are calibrated to estimate the appropriate basic premium rate to cover expected claims: GLM, GAM, Random forest, XGBoost, Light GAM and Neural networks. We compare the predictions with the true underlying rates and try to find the best fit. Finally, we find remarkable results evaluating the profit of the insurers that would use a particular algorithm in comparison with others and the effect of winner's curse.