IMPROVING INSURANCE TARIFFS BASED ON ACTUARIAL AND MACHINE LEARNING MODELS FOR INSURANCE RISK ASSESSMENT

Sa'dullayev Sobirjon Zafar o'g'li

First-year Master's student in Mathematical Economics, Faculty of Mathematics, National University of Uzbekistan named after Mirzo Ulugbek

E-mail: sobirjonsadullayev240@gmail.com

Abstract. This thesis proposes a hybrid framework that improves insurance tariffs by combining actuarial and machine learning models. Generalized linear models form an interpretable frequency-severity tariff, while gradient boosting captures nonlinear effects and risk-factor interactions omitted from the baseline. The framework adds out-of-sample validation, explainability analysis, portfolio-level calibration and regulatory review. It supports more risk-adequate and transparent pricing, particularly in developing insurance markets such as Uzbekistan.

Keywords: insurance tariff; actuarial modeling; generalized linear models; machine learning; gradient boosting; risk assessment; explainable artificial intelligence.

Introduction. An insurance tariff is adequate when each premium reflects the expected loss of the insured risk. Underpricing weakens an insurer's financial stability, while overpricing can drive low-risk clients away and worsen adverse selection.

Generalized linear models (GLMs) remain a standard tool in non-life insurance pricing because their rating factors are transparent and easy to implement [1]. However, nonlinear effects and interactions must usually be specified manually. Tree-based machine learning methods can detect such patterns from data and can inform the design of tariff plans [2, 3].

Machine learning alone is rarely sufficient for tariff setting, because insurers must explain and justify their prices. This need is especially relevant in emerging markets such as Uzbekistan, where historical data may be limited and interpretability supports supervision. The aim of this thesis is to develop a framework that keeps the actuarial structure of the tariff and uses machine learning to correct its weaknesses.

Methodology. The baseline tariff models claim frequency and claim severity separately [1]. For policy i with exposure e_i and rating factors x_i, the number of claims follows a Poisson model and the average claim amount follows a Gamma model, both with a log link:

$$N_i \sim \text{Poisson}(e_i\lambda_i), \qquad \ln\lambda_i = x_i^\top\beta, \qquad \ln\mu_i = x_i^\top\gamma$$

The pure premium per unit of exposure is the product of expected frequency and expected severity. The gross premium adds a risk loading ρ and an expense ratio c:

$$\pi_i = \lambda_i\mu_i, \qquad P_i = \frac{\pi_i(1+\rho)}{1-c}$$

In the hybrid stage, a gradient boosting model [3] is trained with the GLM prediction as an offset, so that it learns only the structure the GLM has missed. This follows the idea of nesting actuarial models within machine learning models [4]. A calibration factor α then restores the balance between predicted and observed claims at portfolio level [7]:

$$\lambda_i^* = \alpha\,\lambda_i^{\mathrm{GLM}}\exp\{f(x_i)\}, \qquad \alpha = \frac{\sum_i N_i}{\sum_i e_i\,\lambda_i^{\mathrm{GLM}}\exp\{f(x_i)\}}$$

The same correction can be applied to severity. Models are compared on out-of-sample data using Poisson deviance for frequency, together with Gamma deviance for severity, the Gini index and actual-to-expected loss ratios by risk group:

$$D = \frac{2}{n}\sum_{i=1}^{n}\left[N_i \ln\frac{N_i}{e_i\lambda_i} - N_i + e_i\lambda_i\right]$$

The logarithmic term is taken as zero when $N_i = 0$. SHAP values are used to explain how each rating factor contributes to individual predictions [5].

Proposed Hybrid Tariff Framework

The hybrid approach is designed to balance predictive flexibility with the interpretability that tariff approval requires. The table below summarizes the expected trade-offs.

Criterion	GLM	Gradient boosting	Hybrid model
Interpretability	High	Low without additional tools	Medium to high
Nonlinear effects and interactions	Specified manually	Learned from data	Learned as a correction
Data requirements	Moderate	Higher	Moderate to higher
Regulatory justification	Straightforward	Difficult	Supported by the GLM base and SHAP

The framework can be implemented in six stages:

1. Collect and clean policy, exposure and claims data.
2. Build frequency and severity GLMs as the baseline tariff.
3. Train boosting corrections on the GLM offsets using cross-validation.
4. Calibrate the model and compare it with the baseline on out-of-sample data.
5. Use SHAP analysis to translate significant effects into understandable rating factors.
6. Review fairness and legal compliance, limit abrupt premium changes, and monitor actual loss ratios.

The fairness review is essential because a model can indirectly reproduce the effect of a prohibited characteristic through correlated variables [6]. Actuarial judgment therefore remains the final control over any data-driven tariff.

Conclusion. Combining actuarial and machine learning models offers a practical route to more risk-adequate insurance tariffs. The GLM keeps the tariff transparent, the boosting component captures omitted risk patterns, and calibration preserves portfolio-level balance. SHAP analysis and fairness review make the results easier to justify to management and regulators. Further research should test the framework on real insurance portfolio data from Uzbekistan and measure its effect on tariff accuracy, loss ratios and insurer profitability.

References

1. Ohlsson, E., & Johansson, B. (2010). *Non-Life Insurance Pricing with Generalized Linear Models*. Springer.
2. Henckaerts, R., Côté, M.-P., Antonio, K., & Verbelen, R. (2021). Boosting insights in insurance tariff plans with tree-based machine learning methods. *North American Actuarial Journal, 25*(2), 255–285.
3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
4. Schelldorfer, J., & Wüthrich, M. V. (2019). Nesting classical actuarial models into neural networks. Available at SSRN.
5. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774.
6. Lindholm, M., Richman, R., Tsanakas, A., & Wüthrich, M. V. (2022). Discrimination-free insurance pricing. *ASTIN Bulletin, 52*(1), 55–89.
7. Wüthrich, M. V., & Merz, M. (2023). *Statistical Foundations of Actuarial Learning and its Applications*. Springer.