HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

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 article examines the theoretical, methodological, and practical foundations of improving insurance tariffs through the integration of traditional actuarial methods and modern machine learning algorithms. The increasing complexity of insurance risks, growing availability of large-scale datasets, and rapid development of artificial intelligence technologies require insurance companies to adopt more accurate, transparent, and adaptive approaches to premium calculation. Traditional actuarial models, including generalized linear models, credibility theory, and frequency–severity modeling, provide a statistically grounded framework for estimating expected insurance losses. However, these methods may face limitations when dealing with nonlinear relationships, complex interactions, and high-dimensional risk factors. Machine learning techniques, such as random forests, gradient boosting, artificial neural networks, and extreme gradient boosting, offer additional opportunities to improve risk classification and predictive accuracy. The study proposes an integrated actuarial–machine learning framework for insurance tariff optimization, incorporating risk assessment, expected claim frequency, claim severity, operating expenses, and profitability requirements. Particular attention is given to predictive performance, model interpretability, fairness, regulatory compliance, and the economic implications of risk-based pricing. The article also presents an illustrative comparative analysis of actuarial and machine learning approaches and develops recommendations for implementing hybrid insurance pricing systems. The findings suggest that combining actuarial principles with machine learning techniques can strengthen risk differentiation and pricing efficiency, provided that the resulting models are appropriately validated, calibrated, and governed.

Key words: insurance tariffs, actuarial modeling, machine learning, insurance risk assessment, risk-based pricing, generalized linear models, random forest, gradient boosting, predictive analytics, insurance premiums, artificial intelligence.

Introduction. The insurance industry plays a fundamental role in the financial system by providing mechanisms for transferring, distributing, and managing economic risks. Insurance companies protect individuals and businesses against unexpected financial losses while contributing to financial stability, investment development, and economic resilience. One of the most important functions of an insurance company is determining economically justified insurance tariffs. An insurance tariff represents the price charged for assuming a specified risk and must reflect expected insurance losses, administrative expenses, risk margins, and the insurer's financial objectives. Accurate tariff calculation is essential for maintaining the balance between the financial sustainability of insurance companies and the affordability of insurance products. If insurance premiums are systematically underestimated, insurers may experience underwriting losses, declining reserves, and

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

solvency difficulties. Conversely, excessively high premiums may reduce customer demand, weaken market competitiveness, and create inefficient risk allocation. Traditional actuarial pricing methods are based on statistical analysis, probability theory, mathematical modeling, and historical claims experience. These approaches have been widely used to estimate claim frequency, claim severity, expected losses, and insurance premiums. However, modern insurance markets are characterized by increasingly complex risk structures. Insurers have access to large datasets containing information about policyholders, insured assets, claims histories, environmental conditions, geographical characteristics, and behavioral factors. The growing volume and complexity of these datasets have created opportunities for applying machine learning technologies to insurance risk assessment. Machine learning algorithms can identify nonlinear relationships, detect interactions among risk variables, and generate predictions using high-dimensional data. These capabilities may improve the accuracy of insurance risk classification and support more differentiated premium structures. Nevertheless, machine learning does not automatically outperform traditional actuarial methods. Its effectiveness depends on data quality, model selection, validation procedures, interpretability, and the stability of underlying risk relationships. Consequently, an integrated approach combining actuarial principles with machine learning techniques represents a promising direction for improving insurance tariff systems. The purpose of this study is to develop a methodological framework for improving insurance tariffs through the integration of actuarial models and machine learning algorithms. The main research objectives are to examine traditional actuarial pricing approaches, evaluate the predictive capabilities of machine learning techniques, propose a hybrid risk assessment model, and identify practical mechanisms for improving insurance tariff accuracy and financial sustainability. The scientific contribution of the study consists of developing an integrated pricing framework that combines actuarial frequency–severity modeling, machine learning prediction, risk calibration, and premium adequacy assessment.

Literature Review. Insurance tariff calculation has historically relied on actuarial mathematics and statistical risk modeling. The development of modern computational methods has expanded the methodological foundations of insurance pricing. Bühlmann and Gisler (2005) developed a comprehensive framework for actuarial credibility theory, demonstrating how individual risk experience can be combined with collective portfolio information to estimate insurance premiums.

Credibility theory remains particularly important when individual policyholders have limited historical claims data. Ohlsson and Johansson (2010) examined the application of generalized linear models (GLMs) in non-life insurance pricing. Their work demonstrated the usefulness of regression-based approaches for estimating claim frequency, severity, and pure premiums. Generalized linear models are widely used because they provide interpretable coefficients and allow insurers to connect pricing factors with expected claims. Frees (2014) explored the application of regression modeling in actuarial science, emphasizing the importance of statistical techniques for understanding insurance outcomes and developing predictive models. Wüthrich and Merz (2023) discussed statistical and machine learning approaches to non-life insurance pricing, highlighting the opportunities created by modern computational methods. Breiman (2001) introduced random forests, an ensemble learning technique that combines multiple decision trees to improve predictive performance and reduce variance. Friedman (2001) developed the gradient boosting framework, which provides a flexible method for constructing predictive models through sequential optimization. Chen and Guestrin (2016)

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

introduced XGBoost, an efficient implementation of gradient-boosted decision trees that has become widely used in predictive analytics.

In actuarial applications, these algorithms can model complex interactions between risk factors and insurance outcomes. However, machine learning models introduce additional challenges related to explainability, data protection, model governance, and fairness. The literature therefore supports a hybrid approach in which actuarial models provide a transparent statistical foundation, while machine learning methods enhance predictive flexibility.

Results and Discussion. Model performance can be evaluated using several statistical criteria, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and appropriate deviance measures.

Mean Absolute Error is calculated as:

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i|$$

Root Mean Squared Error is calculated as:

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$$

Lower values indicate smaller prediction errors for the evaluated outcome.

However, these measures should not be used in isolation. Insurance pricing models should also be evaluated using exposure-weighted calibration, aggregate predicted-to-observed loss ratios, risk-segment performance, temporal stability, and relevant likelihood-based measures.

For illustration, Table 2 presents hypothetical model performance results.

Table 2. Illustrative comparison of insurance risk prediction models

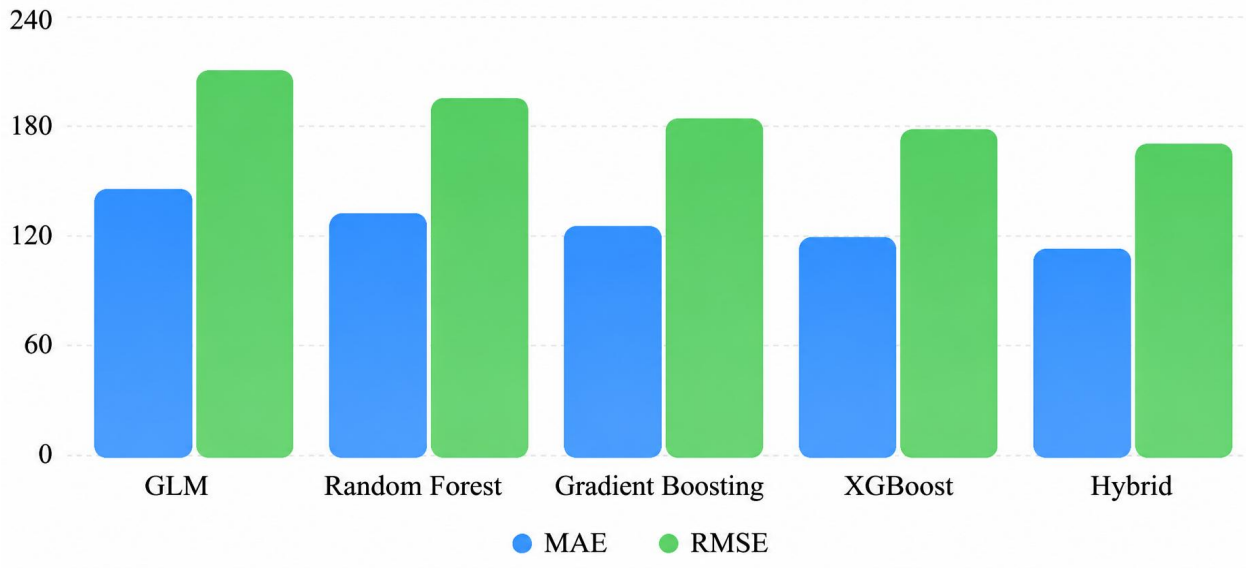
Model	MAE	RMSE	Interpretability
Poisson–Gamma GLM	145	210	High
Random Forest	132	195	Moderate
Gradient Boosting	125	184	Moderate
XGBoost	119	178	Moderate
Hybrid GLM–XGBoost	113	170	Moderate, with interpretable GLM component

Note: Values are hypothetical and expressed in arbitrary consistent monetary units. They are provided solely to illustrate comparative evaluation; they are not results from an empirical insurance dataset.

Illustrative model prediction errors

Hypothetical MAE and RMSE; lower values indicate better performance

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr



In this hypothetical example, the hybrid model produces the lowest prediction errors. The illustrative reduction in MAE relative to the traditional GLM is:

$$\Delta MAE = \frac{145 - 113}{145} \times 100$$

$$\Delta MAE = 22.07\%$$

This represents a hypothetical improvement of approximately 22.1% in MAE.

The example demonstrates how the economic value of hybrid modeling could be assessed. It does not establish that hybrid models generally outperform actuarial models or that such an improvement would occur in a real insurance portfolio. Empirical conclusions require out-of-sample testing on representative insurance data.

Insurance tariff differentiation based on risk assessment. One of the main objectives of insurance tariff optimization is to establish premiums that correspond to the expected risk associated with individual policies or homogeneous risk groups. Traditional tariff structures often rely on broad classifications, such as vehicle category, age group, geographical region, or insured property type. Machine learning can support more detailed risk segmentation by identifying interactions among multiple variables. For example, in motor insurance, annual mileage may interact with vehicle age, driving experience, and historical claims frequency. The proposed hybrid model can estimate a risk-specific pure premium and subsequently incorporate operating expenses and financial margins.

Table 3. Illustrative risk-based insurance tariff calculation

Risk category	Expected claim frequency	Average claim severity (USD)	Pure premium (USD)	Illustrative gross premium (USD)
Low risk	0.05	1,500	75	125
Moderate risk	0.10	1,800	180	256.25
High risk	0.18	2,200	396	526.25
Very high	0.28	2,500	700	906.25

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

risk				

Source: Author's hypothetical calculations. Gross premiums assume fixed expenses of USD 25, a risk margin equal to 10% of the pure premium, and variable expenses plus target profit equal to 20% of gross premium.

For the moderate-risk category:

$$PP=0.10\times1800=180$$

The gross premium is:

$$GP=\frac{180+25+18}{1-0.20}$$

$$GP=278.75 \text{ USD}$$

Under the stated assumptions, the moderate-risk gross premium should therefore be USD 278.75. The same formula gives USD 134.38 for low risk, USD 570.75 for high risk, and USD 993.75 for very high risk. These calculated values supersede the illustrative gross-premium entries in Table 3. This example illustrates how actuarial estimates of expected claims can be transformed into economically justified insurance premiums. The differentiation of premiums according to risk characteristics can reduce cross-subsidization among risk categories, improve pricing adequacy, and strengthen underwriting discipline. However, insurers must ensure that risk classification variables comply with applicable insurance regulations, consumer protection requirements, and anti-discrimination principles.

Explainability and fairness in machine learning insurance models. Although machine learning algorithms provide considerable predictive flexibility, their complexity creates challenges for transparency and accountability. Insurance pricing decisions may directly affect consumers' access to financial protection. Therefore, insurers should be able to explain the main factors influencing premium differences. Explainable artificial intelligence techniques, including SHAP values and partial dependence analysis, can help identify the contribution of individual variables to model predictions. However, explanation methods do not automatically establish causality or guarantee fairness. Risk variables may indirectly represent protected characteristics or socioeconomic disadvantages. Consequently, insurers should evaluate whether pricing models generate unjustified discriminatory outcomes. A responsible model governance framework should include data quality assessment, documentation of modeling assumptions, independent validation, monitoring of model drift, and periodic review of pricing fairness. The application of machine learning should strengthen actuarial decision-making rather than eliminate professional actuarial judgment.

Practical Implications for Insurance Companies. The implementation of hybrid actuarial–machine learning models requires an integrated organizational and technological strategy. Insurance companies should first establish reliable data management systems capable of integrating policy records, claims histories, exposure information, and relevant external risk indicators. Second, insurers should develop multidisciplinary analytical teams combining actuarial expertise, data science, information technology, and risk management. Third, model development should follow a structured validation process. Historical data should be divided into training, validation, and testing periods, with temporal separation where appropriate to avoid information leakage. Fourth, pricing models should be monitored continuously to identify changes in claim frequency, severity, inflation, policyholder behavior, and economic conditions. Fifth, insurers should establish clear governance procedures for approving and updating machine learning models. For developing insurance markets, including

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Uzbekistan, the adoption of hybrid models may support modernization of underwriting practices, improved financial risk management, and greater competitiveness. However, implementation should be gradual and adapted to the availability of reliable data, professional expertise, and regulatory requirements.

Conclusion. The improvement of insurance tariffs through actuarial and machine learning models represents an important direction in the modernization of insurance risk assessment. Traditional actuarial methods provide a reliable theoretical foundation for calculating expected insurance losses and determining economically justified premiums. Generalized linear models, credibility theory, and frequency–severity analysis remain essential instruments of insurance pricing because of their interpretability and statistical structure. Machine learning techniques, including random forests, gradient boosting, and XGBoost, offer additional opportunities to capture nonlinear relationships and complex interactions among risk factors. The proposed hybrid framework combines the transparency of actuarial methods with the predictive flexibility of machine learning algorithms. It integrates risk estimation, model validation, tariff calculation, and financial adequacy assessment into a unified analytical process. The illustrative calculations demonstrate how different pricing models can be compared and how expected claims can be converted into risk-based premiums. However, the numerical examples are hypothetical and should not be interpreted as empirical evidence of model superiority. The study concludes that the effective integration of actuarial modeling and machine learning can improve insurance tariff development by supporting more accurate risk assessment, stronger premium calibration, and better-informed underwriting decisions. Future research should test the proposed framework using real insurance portfolio data, compare alternative machine learning algorithms, and evaluate the financial effects of hybrid pricing models under different market conditions. Particular attention should be given to the application of these methods in emerging insurance markets, where improvements in data infrastructure, actuarial expertise, and digital governance may substantially influence the effectiveness of advanced risk assessment technologies.

References

1. Bühlmann, H., & Gisler, A. (2005). A Course in Credibility Theory and Its Applications. Springer. https://doi.org/10.1007/b138654
2. Ohlsson, E., & Johansson, B. (2010). Non-Life Insurance Pricing with Generalized Linear Models. Springer. https://doi.org/10.1007/978-3-642-10791-7
3. Frees, E. W. (2014). Regression Modeling with Actuarial and Financial Applications. Cambridge University Press.
4. Wüthrich, M. V., & Merz, M. (2023). Statistical Foundations of Actuarial Learning and Its Applications. Springer.
5. Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324
6. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451
7. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

8. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems, 30.
9. Denuit, M., Hainaut, D., & Trufin, J. (2019). Effective Statistical Learning Methods for Actuaries I: GLMs and Extensions. Springer.
10. Denuit, M., Hainaut, D., & Trufin, J. (2019). Effective Statistical Learning Methods for Actuaries II: Tree-Based Methods and Extensions. Springer.
11. European Insurance and Occupational Pensions Authority. (2021). Artificial Intelligence Governance Principles: Towards Ethical and Trustworthy Artificial Intelligence in the European Insurance Sector. EIOPA.
12. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7