

Article

Not peer-reviewed version

Machine Learning and Frequency–Severity Decomposition for Insurance Pricing

[Nguyet Nguyen](#) *

Posted Date: 1 April 2026

doi: 10.20944/preprints202603.2497.v1

Keywords: insurance pricing; pure premium; frequency–severity decomposition; machine learning; XGBoost; generalized linear models; Tweedie models; risk segmentation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Machine Learning and Frequency–Severity Decomposition for Insurance Pricing

Nguyet Nguyen

Mathematics & Statistics, Youngstown State University, Youngstown, 44555, USA; ntnguyen01@ysu.edu

Abstract

Insurance pricing plays a central role in risk management and financial decision-making, as accurate premium estimation directly impacts portfolio stability and profitability. This study investigates insurance pure premium estimation by integrating classical actuarial models with modern machine learning techniques. We compare the traditional frequency–severity decomposition framework with direct modeling approaches, including XGBoost and Tweedie models. For claim frequency, we evaluate Poisson-based models, generalized additive models, and XGBoost. For claim severity, we compare a Gamma generalized linear model with XGBoost. The results show that XGBoost significantly improves predictive performance for both components. Within the decomposition framework, the XGBoost–XGBoost model achieves the best overall prediction accuracy. However, lift-based analysis reveals that the XGBoost–Gamma model provides superior risk segmentation, highlighting a trade-off between prediction accuracy and risk ranking. Direct modeling approaches, while competitive, do not outperform the decomposition framework. Overall, the findings demonstrate that machine learning enhances predictive performance, but its effectiveness is maximized within the frequency–severity framework. The results further indicate that claim frequency is the primary driver of risk differentiation, while claim severity contributes more to prediction accuracy. These findings have important implications for risk management and pricing strategies in insurance portfolios.

Keywords: insurance pricing; pure premium; frequency–severity decomposition; machine learning; XGBoost; generalized linear models; Tweedie models; risk segmentation

1. Introduction

Accurate estimation of expected losses is a central challenge in automobile insurance pricing. Insurers must determine premiums that reflect the underlying risk of each policyholder while ensuring fairness, competitiveness, and regulatory compliance, as premium estimation directly affects portfolio stability and profitability.

A cornerstone of actuarial practice is the frequency–severity framework, which decomposes the expected loss—or pure premium—into two latent components: the expected number of claims and the expected claim size. By modeling these components separately, actuaries can employ statistical distributions tailored to the distinct stochastic properties of each process before recombining them for final loss estimation. This framework provides both interpretability and flexibility, making it fundamental to modern insurance ratemaking.

Generalized Linear Models (GLMs) have long served as the industry standard for actuarial modeling. Existing literature highlights the capacity of GLMs to provide an interpretable yet flexible structure for non-normal, skewed, and heteroscedastic insurance data [3,7,8]. While Poisson and Negative Binomial models are commonly used for claim frequency, Gamma models are widely applied to claim severity due to the strictly positive and right-skewed nature of loss amounts. Extensions such as Generalized Additive Models (GAMs) allow for nonlinear relationships while preserving model transparency [4,10].

Machine learning (ML) has emerged as a powerful tool in insurance analytics. Tree-based algorithms and gradient boosting methods are particularly effective in capturing complex interactions and nonlinear relationships that traditional parametric models may overlook. Empirical evidence suggests that models such as XGBoost can significantly improve performance in tasks ranging from claim frequency prediction to fraud detection [2,6,9].

Recent studies have explored both direct and component-based modeling strategies using machine learning. Direct modeling of total claim amounts, using methods such as Support Vector Regression (SVR), XGBoost, and neural networks, has demonstrated strong predictive accuracy for aggregate loss estimation [18]. Other studies incorporate machine learning within the classical frequency–severity framework, where gradient boosting models are compared with traditional GLMs [19]. These findings indicate that machine learning consistently improves claim frequency prediction, while results for claim severity remain less conclusive due to the high variability of loss amounts.

Despite these advances, direct modeling approaches do not explicitly account for the structural decomposition of insurance losses into frequency and severity components. Moreover, existing evidence on the relative performance of decomposition-based and direct modeling approaches remains inconclusive, particularly when evaluated under both predictive accuracy and risk segmentation criteria.

Classical actuarial models continue to play an important role due to their interpretability and regulatory transparency. Consequently, recent research has focused on hybrid approaches that integrate machine learning techniques within the frequency–severity framework [11,12]. These approaches aim to enhance predictive performance while preserving the structural foundations of actuarial pricing.

Alternative unified modeling approaches have also been proposed, most notably through the Tweedie distribution, which provides a compound Poisson–Gamma representation of aggregate losses [15,16]. While such models offer strong theoretical appeal, it remains unclear whether they can outperform decomposition-based methods in practical insurance applications.

Motivated by these considerations and the lack of consensus in the literature, this study investigates whether machine learning improves insurance pricing by replacing the classical decomposition framework or by enhancing its individual components. In particular, we evaluate model performance not only in terms of prediction accuracy but also in terms of the ability to identify high-risk policies.

To address these questions, we develop a comprehensive modeling framework that integrates classical actuarial models with modern machine learning techniques. We evaluate several models for claim frequency, including Poisson GLMs, spline-based models, and XGBoost. For claim severity, we compare a Gamma GLM with XGBoost. These models are combined within the frequency–severity decomposition framework to estimate pure premium. In addition, we examine direct modeling approaches, including XGBoost and Tweedie models, for predicting aggregate losses.

The contribution of this study is threefold. First, we demonstrate that machine learning methods significantly improve predictive performance for both claim frequency and claim severity, particularly in capturing nonlinear relationships and interactions. Second, we provide empirical evidence that the frequency–severity decomposition framework remains superior to direct modeling approaches, even when advanced machine learning methods are applied. Third, we identify a trade-off between prediction accuracy and risk segmentation: while models combining machine learning for both components achieve the best overall accuracy, models emphasizing frequency modeling provide superior identification of high-risk policies.

Overall, this study contributes to the literature on actuarial science and financial mathematics by providing a comprehensive comparison of classical and modern approaches to insurance pricing. The results highlight the continued relevance of the frequency–severity decomposition framework while demonstrating how machine learning can be effectively integrated to enhance predictive performance and risk differentiation.

The remainder of the paper is organized as follows. Section 2 describes the dataset and data preparation procedures; Section 3 presents the modeling framework and statistical methods; Section 4

reports the empirical findings; and Section 5 concludes with a discussion of the results and their implications for insurance pricing and risk management.

2. Data Description and Preparation

This study utilizes the French motor third-party liability (`freMTPL2`) datasets from the `CASdatasets` package [1]. The raw data comprise two files: `freMTPL2freq`, containing 677,991 policy-year records, and `freMTPL2sev`, containing 26,444 individual claim amounts. These datasets provide a comprehensive set of risk features, including driver demographics, vehicle attributes, and geographic factors.

Table 1 summarizes the variables included in the frequency dataset. These covariates represent standard rating factors used in automobile insurance pricing, including driver age, vehicle characteristics, geographic region, and exposure.

Table 1. Variables in the `freMTPL2freq` dataset.

Variable	Type	Description
<code>IDpo1</code>	Integer	Unique policy identifier.
<code>ClaimNb</code>	Integer	Number of claims during the exposure period.
<code>Exposure</code>	Numeric	Fraction of the year the policy was in force.
<code>Area</code>	Categorical	Geographic area classification.
<code>VehPower</code>	Categorical	Vehicle power category.
<code>VehAge</code>	Integer	Age of the vehicle (years).
<code>DrivAge</code>	Integer	Age of the driver (years).
<code>BonusMalus</code>	Numeric	Bonus–malus coefficient.
<code>VehBrand</code>	Categorical	Vehicle manufacturer or brand.
<code>VehGas</code>	Categorical	Fuel type (gasoline or diesel).
<code>Density</code>	Numeric	Population density of the insured area.
<code>Region</code>	Categorical	Administrative region of residence.

The severity dataset contains 26,444 claim-level observations and includes the policy identifier and the corresponding claim amount for each reported claim. Because the frequency data are recorded at the policy level while the severity data are recorded at the claim level, the two datasets must be reconciled before modeling.

2.1. Data Integration and Preprocessing

To reconcile the policy-level frequency data with the claim-level severity data, all individual claim amounts in `freMTPL2sev` were aggregated by policy identifier (`IDpo1`). This produced, for each policy, the total claim amount incurred during the exposure period and the number of associated claims. Policies with no reported claims were assigned a total loss of zero. The aggregated severity information was then merged with the frequency dataset to create a unified policy-level file containing exposure, claim counts, total incurred losses, and all rating variables.

For policies with at least one claim, an average claim severity variable was also computed for descriptive purposes. Pure premium values were obtained by dividing the total incurred loss by the policy's exposure. These derived quantities are used only for exploratory analysis in this section; the formal notation and modeling framework are introduced later in Section 3.

2.2. Exploratory Data Analysis

The raw insurance portfolio exhibits the extreme class imbalance and heavy-tailed distributions typical of motor liability risks. Figure 1 illustrates these characteristics using log-scaled axes to visualize the full range of the data.

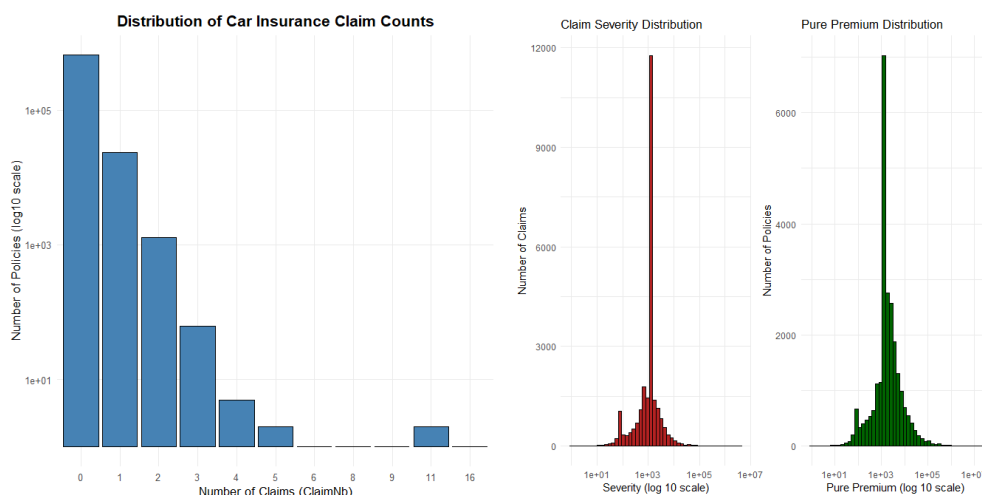


Figure 1. Log-scaled distributions of claim frequency (left), claim severity (center), and pure premium (right) based on the raw dataset.

The frequency distribution (Figure 1, left) is dominated by zero-claim policies, with a rapid decay in frequency as claim counts increase. Notably, the raw data contain rare observations of up to 16 claims per policy, appearing as isolated points in the extreme tail. Similarly, the claim severity and pure premium distributions (Figure 1, center and right) span several orders of magnitude. The severity plot reveals a significant concentration of claims around 1,000 EUR, but also shows an exceptionally long right tail reaching towards 10^6 EUR, representing catastrophic losses.

These visualizations highlight the necessity of the preprocessing steps described in Section 2.3. Specifically, the extreme sparsity of high claim counts and the high-leverage outliers in the severity tail motivate the use of capping (winsorization) to ensure the numerical stability and generalizability of the frequency and severity models.

2.3. Data Cleaning and Preparation

This section details the preprocessing steps implemented to ensure model stability and mitigate the influence of extreme observations. These procedures are critical for both the Poisson frequency models and the Gamma severity models, which are sensitive to high-leverage outliers.

Exposure Filtering Policies with extremely low exposure are prone to producing artificial variance in annualized claim rates. We excluded all records with an exposure below 0.1 years. This threshold ensures that the observations used for training represent a meaningful period of risk. As shown in Table 2, this step improved the mean exposure from 0.529 to 0.632 years.

Capping and Winsorization Given the extreme right-skewness observed in Figure 1, we applied capping to both frequency and severity components:

- **Claim Frequency:** Observations with more than 4 claims were capped at 4. Although the raw data contained counts as high as 16, these represented a negligible fraction of the portfolio ($< 0.01\%$) but could disproportionately influence the maximum likelihood estimation.
- **Claim Severity:** Individual claim amounts were winsorized at 100,000 EUR. This prevents catastrophic “black swan” events—such as the observed maximum loss of 4.07 million EUR—from distorting the Gamma GLM parameters and the subsequent pure premium calculations.

Table 2. Comparison of Dataset Statistics Before and After Cleaning.

Exposure Distribution			Claim Severity Quantiles (EUR)		
Statistic	Before	After	Quantile	Raw	Capped
Min	0.0027	0.1000	50% (Median)	1,172	1,172
Mean	0.5287	0.6318	95%	4,765	4,765
Max	2.0100	2.0100	99%	16,451	16,451
N	677,991	556,439	100% (Max)	4,075,400	100,000

2.4. Final Modeling Dataset Profile

The resulting dataset provides a stabilized foundation for estimating expected losses. Table 3 summarizes the response variables that will be utilized in Section 3. The disparity between the mean and median values across all metrics reinforces the inherent skewness that persists even after cleaning, necessitating the specialized modeling framework proposed in this study.

Table 3. Summary Statistics of the Final Modeling Dataset.

Variable	Mean	Median	Min	Max
ClaimNb (Capped)	0.045	0	0	4
Exposure	0.632	0.630	0.10	2.01
Total Claim Amount (EUR)	82.09	0	0	115,600
Average Severity (EUR)	76.46	0	0	100,000

3. Methodology

This study investigates alternative approaches to modeling insurance pure premium by comparing the classical frequency–severity decomposition with direct modeling of aggregate claim amounts. Let Y_i denote the total claim amount for policy i , $e_i > 0$ the exposure, and \mathbf{x}_i the associated covariates. The target of interest is the pure premium,

$$\pi_i = \frac{\mathbb{E}[Y_i | \mathbf{x}_i]}{e_i}.$$

3.1. Frequency–Severity Decomposition Framework

The classical actuarial framework expresses expected aggregate loss as:

$$\mathbb{E}[Y_i | \mathbf{x}_i] = e_i \lambda_i m_i,$$

where λ_i denotes the expected claim frequency per unit exposure and m_i denotes the expected claim severity conditional on at least one claim. Dividing by exposure, the pure premium is given by

$$\pi_i = \lambda_i m_i.$$

3.2. Claim Frequency Models

We evaluate multiple models for claim frequency and select the best-performing models for use in the decomposition framework.

3.2.1. Poisson Generalized Linear Model

The number of claims N_i is assumed to follow a Poisson distribution:

$$N_i \sim \text{Poisson}(\mu_i),$$

with mean $\mu_i = e_i \lambda_i$. Using a log-link function:

$$\log(\mu_i) = \log(e_i) + \mathbf{x}_i^\top \boldsymbol{\beta},$$

where $\log(e_i)$ is included as an offset.

3.2.2. Negative Binomial Model

To account for overdispersion, the Negative Binomial model is considered:

$$\text{Var}(N_i) = \mu_i + \alpha\mu_i^2,$$

with the same log-link structure as the Poisson model.

3.2.3. Poisson GLM with Natural Cubic Splines

To capture nonlinear relationships, selected predictors are modeled using natural cubic splines:

$$\log(\mu_i) = \log(e_i) + \beta_0 + \sum_{j=1}^4 f_j(z_{ij}) + \sum_{k=1}^p \gamma_k x_{ik}.$$

3.2.4. Generalized Additive Model (GAM)

The GAM extends this approach by estimating smooth functions:

$$\log(\mu_i) = \log(e_i) + \beta_0 + \sum_{j=1}^4 s_j(z_{ij}) + \sum_{k=1}^p \gamma_k x_{ik}.$$

3.2.5. XGBoost for Frequency

To capture complex nonlinearities and interactions, we apply XGBoost:

$$\hat{\mu}_i = \sum_{t=1}^T f_t(\mathbf{x}_i), \quad f_t \in \mathcal{F}.$$

3.2.6. Model Selection for Frequency

Based on predictive performance and risk segmentation metrics, the two best-performing models—Poisson GLM with splines and XGBoost—are selected for use in pure premium estimation.

3.3. Claim Severity Models

We consider two alternative approaches for modeling claim severity.

3.3.1. Gamma GLM

Claim severity Z_i is modeled using a Gamma distribution with log-link:

$$\log(\theta_i) = \mathbf{x}_i^\top \boldsymbol{\phi}.$$

3.3.2. XGBoost for Severity

To capture nonlinear relationships, we apply XGBoost to model the logarithm of severity:

$$\log(Z_i) = \sum_{k=1}^K f_k(\mathbf{x}_i) + \varepsilon_i.$$

3.4. Pure Premium Estimation

The estimated pure premium under the decomposition approach is:

$$\hat{\pi}_i^{\text{decomp}} = \hat{\lambda}_i \hat{m}_i.$$

We consider four model combinations:

- **Spline–Gamma:** Poisson spline frequency + Gamma severity,

- **Spline–XGBoost:** Poisson spline frequency + XGBoost severity,
- **XGBoost–Gamma:** XGBoost frequency + Gamma severity,
- **XGBoost–XGBoost:** XGBoost frequency + XGBoost severity.

3.5. Direct Modeling of Aggregate Loss

To provide a benchmark, we consider models that directly predict aggregate claim amounts.

3.5.1. Tweedie Model

The Tweedie distribution satisfies:

$$\text{Var}(Y_i | \mathbf{x}_i) = \phi \mu_i^p, \quad 1 < p < 2,$$

corresponding to a compound Poisson–Gamma model. The mean is modeled as:

$$\log(\mu_i) = \log(e_i) + \mathbf{x}_i^\top \boldsymbol{\beta}.$$

3.5.2. XGBoost for Total Loss

We also apply XGBoost to directly model total claim amounts:

$$\log(1 + Y_i) = f(\mathbf{x}_i).$$

3.6. Model Evaluation

Model performance is evaluated using both accuracy and risk segmentation metrics.

Prediction accuracy is assessed using Mean Squared Error (MSE), Mean Absolute Error (MAE), and Pearson correlation.

Risk segmentation is evaluated using decile-based analysis, where policies are ranked by predicted values and grouped into ten deciles. A well-performing model should exhibit strong concentration of losses in the highest-risk deciles.

4. Results

This section presents the empirical results for the claim frequency, claim severity, and pure premium models. We begin by evaluating the performance of alternative models for claim frequency, followed by claim severity, and then assess the combined frequency–severity decomposition and direct modeling approaches.

4.1. Claim Frequency Model Performance

Table 4 summarizes the performance of the claim frequency models.

Table 4. Performance Comparison of Claim Frequency Models.

Model	MSE	MAE	Correlation	Lift (Top 10%)
Poisson GLM	0.04654	0.08432	0.136	2.48
Poisson + Splines	0.04643	0.08415	0.144	2.76
Negative Binomial	0.04656	0.08445	0.136	2.48
GAM	0.04637	0.08418	0.147	2.62
XGBoost	0.04573	0.08349	0.185	3.06

Among all models, XGBoost achieves the best performance across all evaluation metrics, including the lowest MSE and MAE, as well as the highest correlation and lift. In particular, the lift in the top decile indicates a strong ability to identify high-risk policies, highlighting the effectiveness of machine learning in risk segmentation.

Among the parametric models, the Poisson model with spline terms and the GAM provide modest improvements over the standard Poisson GLM, suggesting that incorporating nonlinear effects enhances predictive performance. In contrast, the Negative Binomial model performs similarly to the Poisson GLM, indicating that overdispersion does not substantially affect predictive accuracy in this dataset.

Overall, these results demonstrate that machine learning models provide superior predictive performance and improved risk differentiation compared to classical parametric approaches.

Claim Capture Curves

Figure 2 presents the cumulative claim capture curves for the competing models. The XGBoost model consistently captures a larger proportion of claims in the highest-risk segments, followed by the Poisson model with spline terms and the GAM. The standard Poisson and Negative Binomial models exhibit weaker concentration of claims.

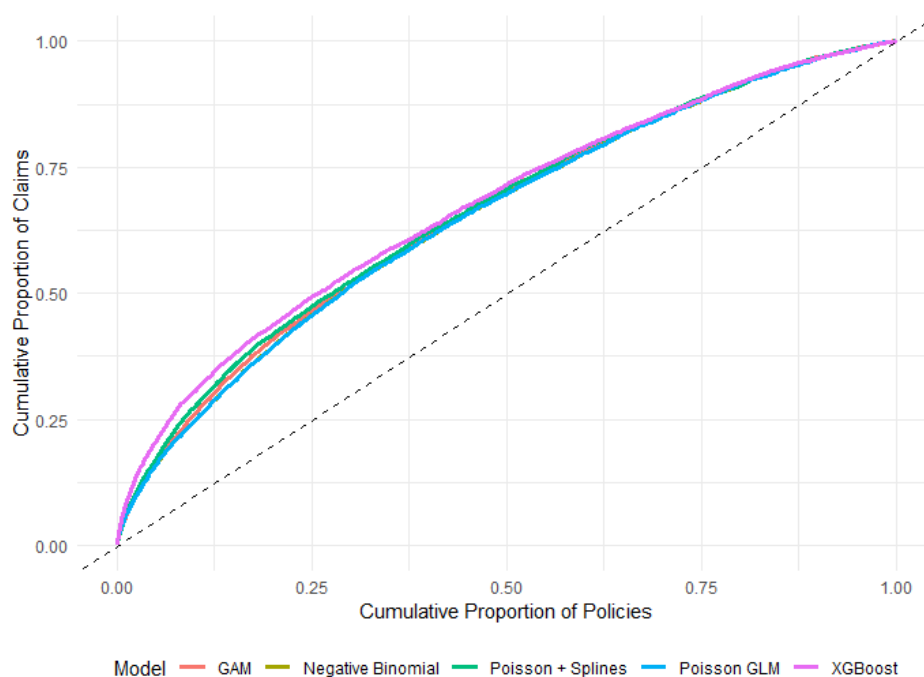


Figure 2. Comparison of Claim Capture Curves Across Frequency Models.

These results are consistent with the lift statistics and confirm that models incorporating nonlinear effects improve risk segmentation, with machine learning providing additional gains.

Decile Lift Analysis

Figure 3 presents the decile lift curves for the competing models. The XGBoost model achieves the highest lift in the top decile, followed by the Poisson model with spline terms and the GAM. Differences across the middle deciles are relatively small, reflecting the low overall claim frequency and limited variability in moderate-risk groups.

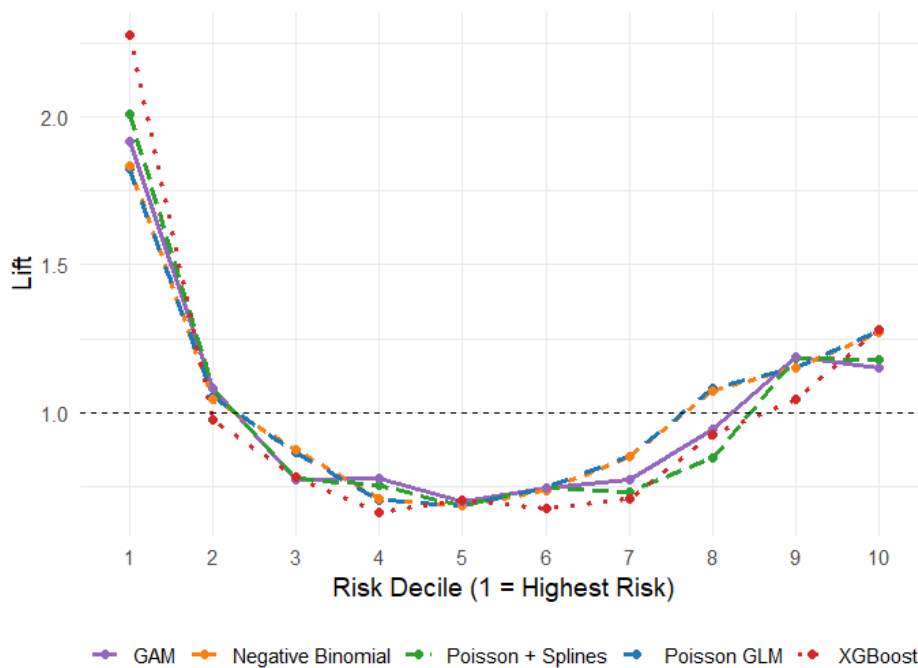


Figure 3. Decile Lift Chart for Frequency Models.

Interpretation of Model Components

Variable importance results from the XGBoost model (Figure 4) indicate that the bonus–malus score is the dominant predictor of claim frequency, followed by driver age, population density, and selected vehicle characteristics. These findings are consistent with actuarial intuition and highlight the importance of past claims experience and demographic factors in risk assessment.

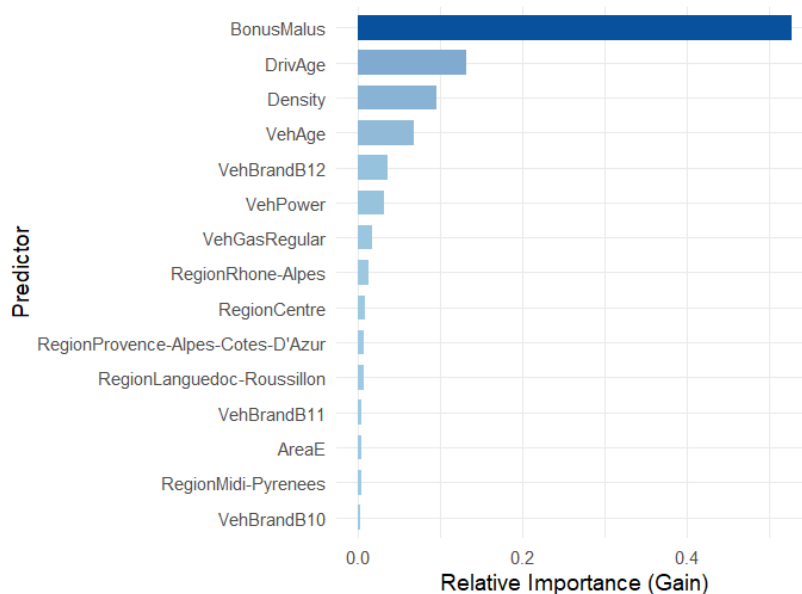


Figure 4. Variable Importance from the XGBoost Frequency Model.

The smooth functions estimated by the GAM (Figure 5) reveal nonlinear effects for several continuous predictors. In particular, the bonus–malus variable exhibits a strong nonlinear relationship with claim frequency, while vehicle age and population density display more moderate nonlinear patterns.

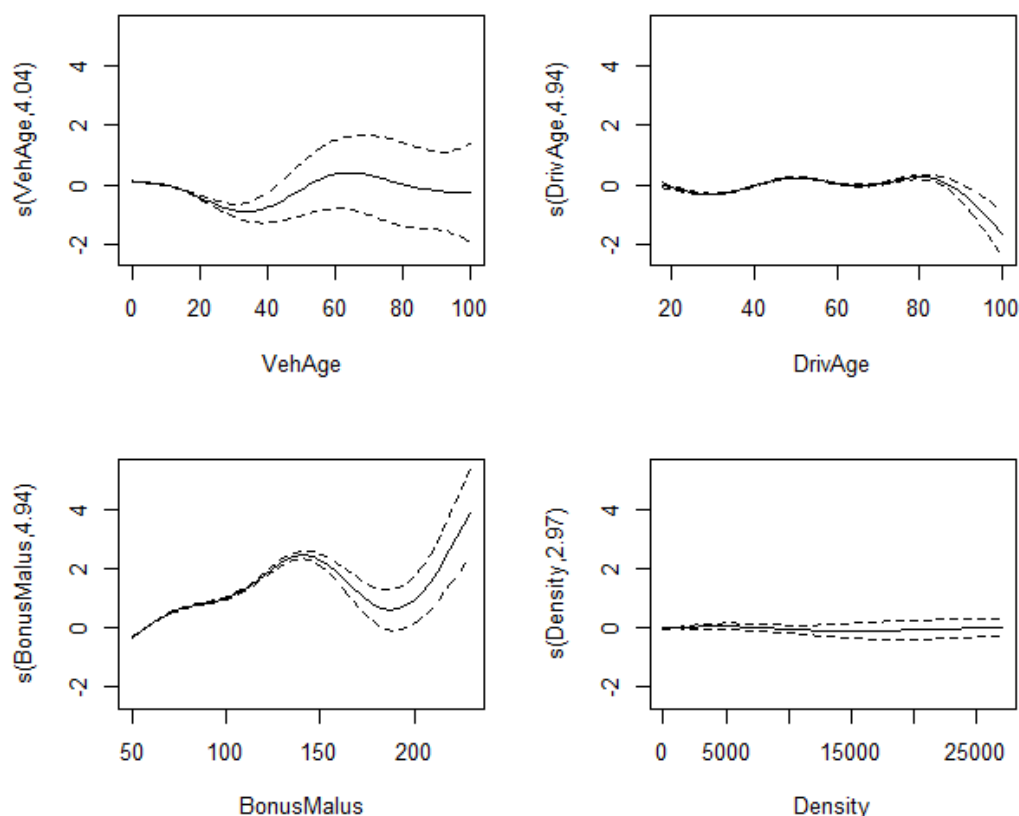


Figure 5. Estimated Smooth Effects from the GAM Frequency Model

Model Selection

The results indicate that XGBoost provides the strongest overall predictive performance, achieving the lowest error measures, the highest correlation, and the greatest lift in the highest-risk decile. In addition, the Poisson model with spline terms delivers competitive performance while retaining the interpretability and transparency of the generalized linear modeling framework widely used in actuarial practice.

To enable a direct comparison between a traditional actuarial approach and a modern machine learning method in subsequent analyses, both models are retained. The Poisson spline model serves as the representative GLM-based specification, while XGBoost represents the machine learning approach. This dual-model strategy allows us to evaluate not only predictive accuracy but also the trade-offs between interpretability and performance in insurance pricing applications. These findings justify the selection of XGBoost and spline-based GLMs for subsequent pure premium modeling.

4.2. Claim Severity Model Performance

Table 5 summarizes the performance of the claim severity models.

Table 5. Performance of Claim Severity Models.

Model	MSE	MAE	Correlation
Gamma GLM	29,341,925	1457.49	0.0161
XGBoost	3,842,868	1409.51	0.0461

The XGBoost model substantially outperforms the Gamma GLM across all evaluation metrics. In particular, it achieves a markedly lower MSE and MAE, along with a higher correlation with observed claim severity. These results indicate that machine learning methods are more effective in capturing the complex nonlinear relationships and interactions inherent in claim size data.

The Gamma GLM, while appropriate for modeling positive and right-skewed outcomes, assumes a log-linear relationship between covariates and severity. This assumption limits its ability to capture nonlinear effects and interactions, leading to inferior predictive performance.

Despite the improvement provided by XGBoost, the correlation remains relatively low, reflecting the inherent variability and stochastic nature of claim severity. This suggests that, although machine learning improves predictive accuracy, a substantial portion of severity variation remains unexplained.

These findings help explain the improved performance of decomposition models that incorporate XGBoost for severity, particularly in terms of prediction accuracy.

Figure 6 illustrates the relationship between actual and predicted claim severity. The Gamma GLM produces highly concentrated predictions, indicating strong shrinkage toward the mean and limited ability to capture variability in claim sizes. In contrast, the XGBoost model exhibits greater dispersion and improved alignment with observed values, highlighting its ability to capture nonlinear relationships and complex interactions.

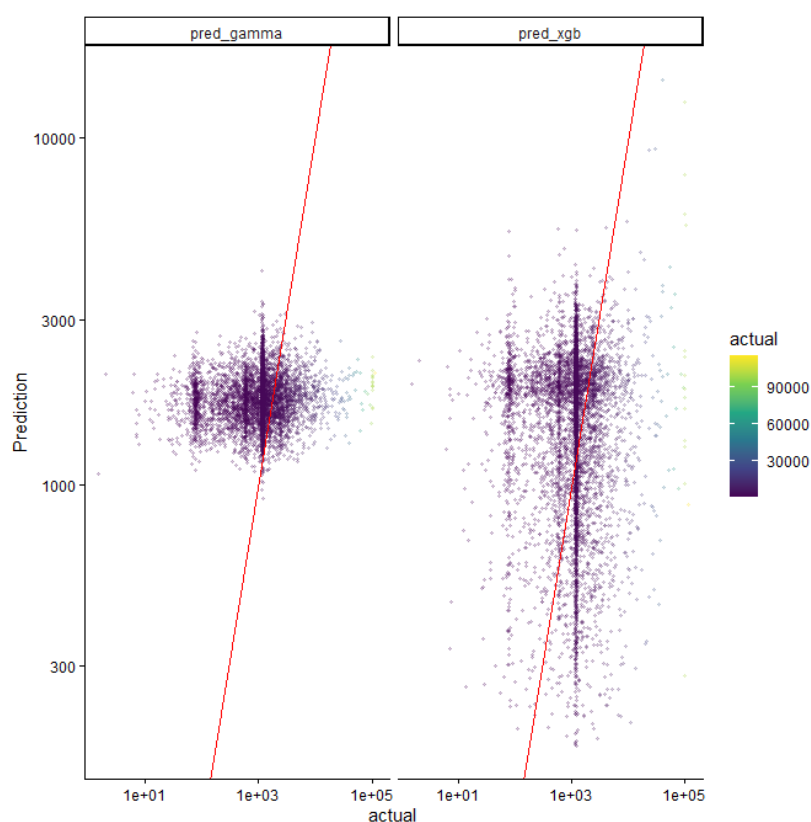


Figure 6. Comparison of actual and predicted claim severity for the Gamma GLM and XGBoost models on a log scale. Observations with zero values are excluded. Color represents the magnitude of observed claim severity.

However, both models display considerable scatter, underscoring the inherently stochastic nature of claim severity and the difficulty of accurately predicting individual claim amounts. Observations with zero or near-zero values were excluded from the log-scale visualization to avoid numerical issues associated with logarithmic transformation.

4.3. Results for Frequency–Severity Decomposition for Pure Premium

Table 6 presents the performance of the frequency–severity decomposition models using both error-based and ranking-based metrics.

Among the models considered, the XGBoost–XGBoost configuration achieves the best overall prediction accuracy, with the lowest MSE and MAE, as well as the highest correlation. This indicates that incorporating machine learning methods for both frequency and severity improves pure premium estimation.

Table 6. Performance of Frequency–Severity Decomposition Models.

Model	MSE	MAE	Correlation	Lift (Top 10%)
Spline–Gamma	15,553,426	244.57	0.0066	1.509
XGBoost–Gamma	15,548,539	243.19	0.0126	2.039
Spline–XGBoost	15,550,927	242.83	0.0112	1.452
XGBoost–XGBoost	15,547,197	241.64	0.0165	1.663

In contrast, the ranking-based results reveal a different pattern. The XGBoost–Gamma model achieves the highest lift, indicating superior performance in identifying high-risk policies. This suggests that improvements in modeling claim frequency have a greater impact on risk segmentation than enhancements in severity modeling.

The difference between error-based and lift-based metrics highlights an important trade-off. While the XGBoost–XGBoost model minimizes prediction error, the XGBoost–Gamma model is more effective in concentrating losses within the highest-risk segment. This indicates that different model configurations may be preferred depending on the objective, such as pricing accuracy versus risk classification.

The classical Spline–Gamma model yields the weakest performance across all metrics, although it remains a useful benchmark due to its interpretability.

Overall, these results demonstrate that machine learning improves predictive performance within the decomposition framework, but its impact differs between accuracy and risk segmentation.

Risk Segmentation Performance

Figure 7 illustrates the decile lift curves for the decomposition models. All models exhibit a strong concentration of losses in the highest-risk decile, indicating effective identification of high-risk policies.

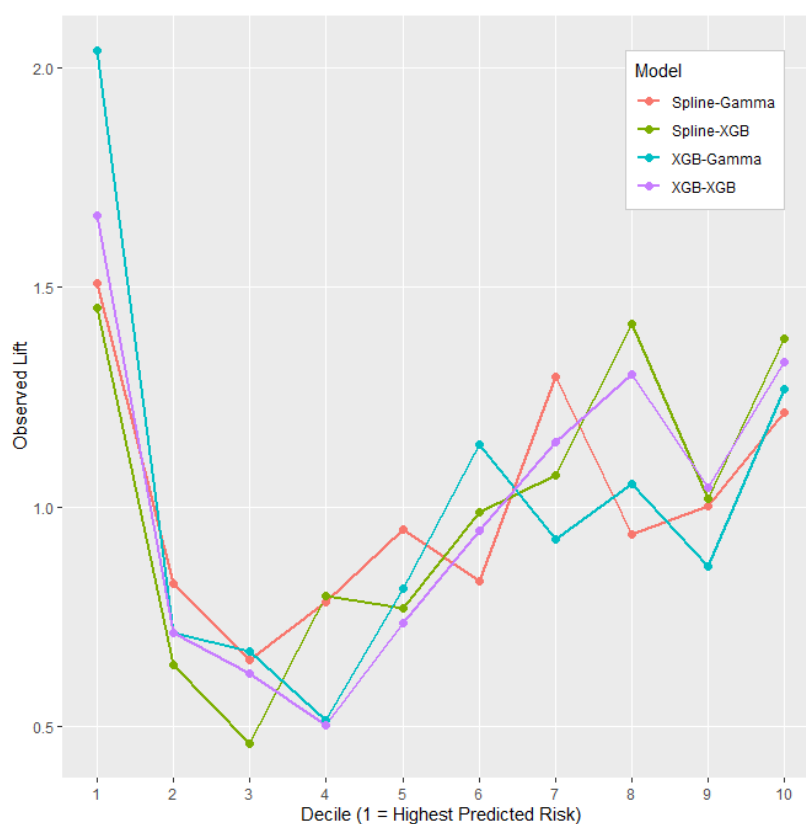


Figure 7. Decile lift curves for pure premium predictions across different model configurations. Decile 1 corresponds to the highest predicted risk.

Among the models, the XGBoost–Gamma configuration achieves the highest lift, confirming its superior performance in risk segmentation. In contrast, although the XGBoost–XGBoost model pro-

vides the best overall prediction accuracy, its lift is lower, reinforcing the trade-off between prediction accuracy and risk ranking performance.

These findings suggest that claim frequency plays a dominant role in risk differentiation, while claim severity contributes more to overall prediction accuracy. Consequently, improvements in frequency modeling have a greater impact on identifying high-risk policies than enhancements in severity modeling.

The non-monotonic behavior observed in lower-risk deciles reflects the inherent variability of insurance losses, particularly for policies with low exposure. These results provide empirical evidence supporting the continued use of frequency–severity decomposition, particularly when combined with machine learning methods.

4.4. Comparison Between Decomposition and Direct Modeling for Total Loss

To evaluate the effectiveness of direct modeling approaches, we compare the best-performing decomposition model (XGBoost–XGBoost) with direct XGBoost and Tweedie models for total claim amount prediction.

Table 7 summarizes the performance of decomposition and direct modeling approaches.

Table 7. Comparison of Decomposition and Direct Models for Total Claim Amount.

Model	MSE	MAE	Correlation
XGBoost–XGBoost (Decomposition)	1,641,356	138.99	0.0656
XGBoost (Direct)	1,644,657	150.43	0.0460
Tweedie GLM (Direct)	1,649,283	162.65	0.0329

The decomposition-based XGBoost–XGBoost model consistently outperforms both direct approaches across all evaluation metrics. It achieves the lowest MSE and MAE, as well as the highest correlation, demonstrating the effectiveness of modeling claim frequency and severity separately.

The direct XGBoost model provides competitive performance, indicating that flexible machine learning methods can capture nonlinear relationships in aggregate losses. However, it does not match the accuracy of the decomposition framework, suggesting that modeling aggregate losses directly may overlook important structural information captured by the frequency–severity approach.

The Tweedie model, despite its theoretical foundation as a compound Poisson–Gamma model, exhibits weaker performance across all metrics. This indicates that, although the data exhibit a compound structure, modeling aggregate losses within a single unified framework may be less effective than decomposing the problem into separate components.

The estimated Tweedie parameter ($\hat{p} = 1.52$) confirms that the data follow a compound Poisson–Gamma structure. However, the inferior predictive performance of the Tweedie model suggests that imposing a single functional relationship between covariates and aggregate loss may be overly restrictive. In contrast, the decomposition framework allows for distinct covariate effects on frequency and severity, resulting in greater modeling flexibility and improved predictive accuracy.

Overall, these results highlight the importance of structural decomposition in actuarial modeling. While machine learning enhances predictive performance within individual components, explicitly separating frequency and severity yields superior results. These findings provide empirical support for the continued use of the frequency–severity decomposition framework in modern insurance pricing.

5. Conclusions and Discussion

This study examines the integration of classical actuarial models and modern machine learning techniques for insurance pricing, with particular emphasis on comparing the traditional frequency–severity decomposition framework with direct modeling approaches. The results provide important insights into predictive performance and practical implications for risk management.

The empirical findings show that machine learning methods, particularly XGBoost, substantially improve predictive accuracy for both claim frequency and claim severity. The improvement is especially

evident for severity modeling, where the Gamma GLM tends to shrink predictions toward the mean and fails to capture the variability of claim sizes. In contrast, XGBoost effectively models nonlinear relationships and interactions, leading to lower prediction errors. Despite these improvements, correlation values remain relatively low, reflecting the inherently stochastic nature of claim severity.

Within the decomposition framework, the XGBoost–XGBoost model achieves the highest overall predictive accuracy, confirming that machine learning enhances pure premium estimation when applied to both components. However, improvements in error-based metrics are moderate, which highlights the high variability of insurance losses, particularly for policies with low exposure.

An important contribution of this study is the identification of a trade-off between prediction accuracy and risk segmentation. Although the XGBoost–XGBoost model minimizes prediction error, the XGBoost–Gamma model achieves the highest lift and provides superior identification of high-risk policies. This result indicates that claim frequency plays a dominant role in risk differentiation, while claim severity contributes more to overall prediction accuracy. Consequently, improvements in frequency modeling have a greater impact on ranking policies by risk level than enhancements in severity modeling.

The comparison between decomposition and direct modeling approaches provides further insight. While direct XGBoost and Tweedie models deliver competitive performance, they do not outperform the decomposition-based framework in this study. This finding contrasts with the results of [17], which suggest that direct modeling of aggregate losses can yield lower prediction error.

This discrepancy highlights an important distinction between general predictive modeling and actuarial applications. Direct modeling may perform well in settings where minimizing prediction error is the sole objective. However, in insurance pricing, the frequency–severity decomposition reflects the underlying structure of claim processes and enables separate modeling of distinct risk components. This structural advantage allows for greater flexibility in capturing heterogeneous risk factors and improves risk segmentation.

In addition, direct modeling imposes a single functional relationship between covariates and aggregate loss, which may be overly restrictive when frequency and severity are influenced by different drivers. In contrast, the decomposition framework accommodates distinct covariate effects for each component, resulting in improved predictive performance in practice.

From a practical perspective, these findings have important implications for insurance pricing and risk management. If the objective is to minimize prediction error, models that incorporate machine learning for both frequency and severity are preferred. If the goal is to identify high-risk policies for underwriting or portfolio management, models that emphasize frequency modeling are more effective.

The results are also consistent with prior studies comparing gradient boosting methods with classical GLMs within the frequency–severity framework [19]. As in previous work, machine learning significantly improves claim frequency prediction. However, while earlier studies report mixed results for severity modeling, the present analysis shows that XGBoost can outperform the Gamma GLM when appropriately specified and tuned. This suggests that the relative performance of machine learning for severity is context-dependent and may improve with richer data and more flexible model configurations.

Overall, this study demonstrates that machine learning enhances predictive performance, but its effectiveness is maximized when combined with the classical frequency–severity decomposition framework. The results indicate that theoretical advantages of direct modeling do not necessarily translate into superior performance in real-world actuarial applications. These findings provide empirical support for decomposition-based pricing approaches and offer a foundation for future research on hybrid models that balance predictive accuracy and risk segmentation in complex insurance environments.

Funding: This research received no external funding.

Data Availability Statement: The data used in this study are part of the CASdatasetsR package. Specifically, the freMTPL2freqand freMTPL2sevdatasets were used. These data are publicly available for research purposes and can be accessed via the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=CASdatasets> or the Zenodo repository (DOI: 10.57745/P0KHAG).

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AIC	Akaike Information Criterion
GAM	Generalized Additive Model
GLM	Generalized Linear Model
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
REML	Restricted Maximum Likelihood
NB	Negative Binomial
XGBoost	Extreme Gradient Boosting

References

- Dutang, C., and Charpentier, A. (2024). *CASdatasets: Insurance datasets*. R package version 1.2-0. DOI: [10.57745/P0KHAG](https://doi.org/10.57745/P0KHAG).
- Chen, T., and 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.
- de Jong, P., and Heller, G. Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press.
- Frees, E. W. (2014). *Regression Modeling with Actuarial and Financial Applications*. Cambridge University Press.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
- Henckaerts, R., Côté, M.-P., Antonio, K., and Verbelen, R. (2021). Boosting insights in insurance tariff plans with tree-based machine learning methods. *North American Actuarial Journal*, 25(2), 255–285.
- Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: From Data to Decisions*. Wiley.
- McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models* (2nd ed.). Chapman & Hall.
- Noll, A., Salzmann, R., and Wüthrich, M. V. (2018). Case study: French motor third-party liability claims. *SSRN Electronic Journal*. DOI: [10.2139/ssrn.3164764](https://doi.org/10.2139/ssrn.3164764).
- Ohlsson, E., and Johansson, B. (2010). *Non-Life Insurance Pricing with Generalized Linear Models*. Springer.
- Richman, R., and Wüthrich, M. V. (2019). Neural network embedding of the over-dispersed Poisson model. *Scandinavian Actuarial Journal*, 2019(6), 451–470.
- Wüthrich, M. V. (2020). Bias regularization in neural network models for general insurance pricing. *European Actuarial Journal*, 10(1), 179–202.
- Denuit, M., Marechal, X., Pitrebois, S., and Walhin, J.-F. (2007). *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. Wiley.
- Dickson, D. C. M., Hardy, M. R., and Waters, H. R. (2009). *Actuarial Mathematics for Life Contingent Risks*. Cambridge University Press.
- Jørgensen, B. (1997). *The Theory of Dispersion Models*. Chapman & Hall.
- Smyth, G. K., and Jørgensen, B. (2002). Fitting Tweedie's compound Poisson model to insurance claims data. *ASTIN Bulletin*, 32(1), 143–157.
- Moonoo, D.; Hosein, P. (2024). Predicting automobile insurance claim rate versus through severity and frequency predictions. In *Proceedings of the 2024 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*, Sharjah, United Arab Emirates, 2024; pp. 1–6. <https://doi.org/10.1109/ICTMOD63116.2024.10878218>.

18. Bekkaye, C.; Oukhoya, H.; Zari, T.; Guerbaz, R.; El Bouanani, H. (2025). Advanced strategies for predicting and managing auto insurance claims using machine learning models. *Statistics, Optimization & Information Computing*, 14(3), 1440–1457. <https://doi.org/10.19139/soic-2310-5070-2655>.
19. Clemente, C.; Guerreiro, G. R.; Bravo, J. M. (2023). Modelling motor insurance claim frequency and severity using gradient boosting. *Risks*, 11(9), 163. <https://doi.org/10.3390/risks11090163>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.