STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, Month 2026, pp 0–16.
Published online in International Academic Press (www.IAPress.org)

Strategies for Predicting and Managing Life Credit Insurance Claims Using Machine Learning Models

Hamza BOUFIKR*, Mohammed BENMOUMEN

Mohammed First University. Faculty of Science Oujda. Department of Mathematics. MASA, LaMAO Laboratory, Morocco

Abstract The increasing severity of credit life insurance claims has created a growing need for more effective methods for claim assessment and loss management. In this context, machine learning and statistical modeling provide useful tools for improving the prediction of claim amounts and supporting actuarial decision-making. Since enhancing customer service and pricing accuracy remains a central objective for insurers, advanced predictive methods are increasingly being used to extract meaningful insights from insurance portfolios.
This paper contributes to the literature on life credit insurance pricing by modeling total claim amounts using several complementary approaches, including Support Vector Regression (SVR), Extreme Gradient Boosting (XGBoost), Multi-Layer Perceptron (MLP), Tweedie Generalized Linear Models, and Quantile Regression. A comparative analysis is performed using standard predictive performance measures such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and, for quantile regression, the pinball loss.
The results show that the competing models provide different but complementary insights into claim severity. The machine learning models offer flexible predictive performance, the Tweedie model provides an interpretable actuarial framework adapted to non-negative and right-skewed claim amounts, and quantile regression reveals heterogeneous predictive behavior across the distribution of claims. Overall, the study highlights the value of combining machine learning methods with actuarially grounded statistical approaches in order to improve pricing analysis and support more effective loss management in life credit insurance.

Keywords Claim Amount; Machine Learning; Regression; Performance Criteria; Prediction; Life Credit Insurance; Claims Management.

DOI: 10.19139/soic-2310-5070-3545

1. Introduction

The scientific study of insurance risk has deep historical roots that predate modern statistical learning. Early foundations were established through the development of probability theory and life contingencies, where the first systematic analyses of mortality and longevity enabled insurers to quantify life-related risks. Pioneering demographic work such as Graunt's analysis of mortality patterns (1662) and Halley's construction of one of the first life tables for annuity pricing (1693) laid the groundwork for actuarial valuation. Subsequent advances in life insurance mathematics—particularly the formalization of present-value calculations for life annuities and assurances by de Moivre and later authors made it possible to translate uncertain lifetimes into financially consistent premiums and reserves. In the nineteenth century, classical mortality laws such as Gompertz (1825) and Makeham (1860) provided parsimonious parametric representations of death intensities and became central tools in life insurance pricing and reserving.

Parallel progress occurred in life insurance with the emergence of collective risk theory. The early twentieth century introduced key ideas for aggregate loss modeling and ruin probabilities, notably through the Lundberg framework and the later refinement of the Cramér–Lundberg model, which formalized the balance between premium

*Correspondence to: Hamza BOUFIKR (Email: hamzaboufikr@gmail.com).Mohammed First University. Faculty of Science Oujda. Department of Mathematics. MASA, LaMAO Laboratory, Morocco

ISSN 2310-5070 (online) ISSN 2311-004X (print)

inflow and stochastic claim outflow. These contributions established the theoretical basis for solvency analysis and capital adequacy, allowing insurers to study the probability of insolvency under realistic claim arrival and severity mechanisms. Later developments in aggregate claims computation and recursion techniques (e.g., Panjer recursion) strengthened the practical implementation of compound frequency–severity models, reinforcing the frequency–severity paradigm as one of the most influential actuarial decompositions for portfolio risk.

Credit life insurance is designed to protect lenders and borrowers by settling (all or part of) the outstanding loan balance in the event of the borrower's death (and sometimes disability, depending on contract design). While the product is conceptually simple, the underlying claim process is not: most contracts generate no claim during the coverage period, whereas the positive claim amounts—when they occur—can be highly heterogeneous and strongly right-skewed because they depend on borrower characteristics, loan size, loan term, and pricing conditions. In such settings, traditional linear regression approaches (which rely on Gaussian errors, linear effects, and often constant variance) may fail to capture heavy tails, nonlinear interactions, and regime like behavior between "no-claim" and "claim" outcomes.

Actuarial science has long addressed these challenges through structured probabilistic modeling. In particular, the frequency–severity framework decomposes the aggregate loss into a claim occurrence component (frequency) and a conditional positive amount component (severity), frequently estimated with Generalized Linear Models (GLMs) using appropriate distributions and link functions. This decomposition is attractive for insurance portfolios because it aligns with operational processes (underwriting, pricing, reserving) and provides interpretable risk factors. However, in modern credit life portfolios, the relationship between covariates (e.g., age at entry, loan characteristics, behavioral indicators) and claim cost can involve complex nonlinearities and interactions that are difficult to represent with a single parametric specification, especially when the dataset mixes numerical and categorical drivers and exhibits non-constant dispersion across the claim distribution.

At the same time, insurers are operating under increasing pressure to improve pricing accuracy, enhance claims management, and maintain financial stability under regulatory oversight (in Morocco, under the supervision of ACAPS). From a business perspective, better prediction of claim amounts supports fairer premiums, improved risk selection, tighter monitoring of loss experience, and more proactive customer service. From a risk-management perspective, accurate predictive models can reduce unexpected loss volatility by aligning technical pricing and portfolio steering with the true underlying risk profile—particularly important in credit-linked covers where exposures can change rapidly with credit cycles and borrower composition.

Motivated by these needs, this study proposes a data-driven framework for predicting and managing credit life insurance claim amounts using modern Machine Learning (ML) regressors. Using a real dataset provided by Wafa Insurance, the paper focuses on the response variable *total claim amount* and leverages a refined set of key predictors (including age at entry, loan amount, loan term, interest rate, number of children, sex, and smoker status) after preprocessing and normalization. Three nonlinear predictive models are investigated—Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Multi-Layer Perceptron (MLP)—with hyperparameters tuned via Randomized Search Cross-Validation and performance evaluated through standard error metrics (e.g., R^2, MSE, RMSE, MAE) complemented by diagnostic tools such as residual analysis and Taylor diagrams.

Beyond pure predictive accuracy, interpretability is also addressed through SHAP value analysis, which ranks and explains the contribution of each predictor to model outputs, helping bridge the gap between "black-box" learning and actuarial reasoning. In addition, the paper reconnects the ML approach to classical actuarial thinking by incorporating a Tweedie GLM and quantile regression alongside the ML regressors, thereby providing a coherent link between established actuarialpractice and flexible ML prediction.

The main contributions of this work are therefore fivefold: (i) delivering an applied, reproducible ML pipeline tailored to credit life claim amounts, (ii) providing a comparative assessment of three major nonlinear regressors on real insurer data, (iii) enriching model governance through interpretability via SHAP value analysis, (iv) establishing an actuarially grounded benchmark through a Tweedie GLM that naturally accommodates the zero-inflated and right-skewed structure of claim amounts, and (v) extending the analysis to the full conditional distribution through quantile regression, thereby providing insights into tail risk that are of direct relevance for insurance pricing and reserving.

2. Research Methodology

The collected dataset was sourced from **an Insurance Company**, a Moroccan life-insurance provider. The original dataset contains **2,000 rows** and **21 variables**, including demographic, financial, and behavioural features of insured borrowers. After removing irrelevant or incomplete columns and filtering for essential predictive variables, the refined dataset comprises **2,000 observations** with **7 explanatory variables**. In this dataset, the variables *age_entry*, *loan amount*, *loan_term*, *rate* and *children* are treated as numerical variables, whereas *sex* and *smoker* are considered categorical variables.

Variable	Type	Description
age_entry	Numerical	Age at loan inception.
loan_amount	Numerical	Borrowed principal amount.
loan_term	Numerical	Loan duration (months).
rate	Numerical	Periodic interest rate.
children	Numerical	Number of children.
sex	Categorical	Gender (M/F).
smoker	Categorical	Smoker indicator (Y/N).

Table 1. Description of the data set.

Modeling strategy for zero-inflated data: Given the highly zero-inflated structure of the response variable (98% zero claims), a natural alternative would be a two-part hurdle model, which explicitly separates the probability of claim occurrence from the severity of positive claims. However, the present study deliberately adopts a unified modeling strategy for two reasons. First, the Tweedie GLM with $1 < p < 2$ implicitly embeds a two-part structure through its compound Poisson–Gamma representation, jointly modeling the zero-mass probability and the conditional severity distribution within a single coherent framework . Second, the ML regressors (XGBoost, MLP, SVR) are applied directly to the full response variable, allowing them to learn the zero-heavy distribution from the data without imposing parametric assumptions on the claim occurrence process. A dedicated two-part implementation remains an important direction for future research.

2.1. Preprocessing of the data set

Prior to model training, the data set is preprocessed to ensure that all predictors are correctly scaled and consistently formatted. Let $\mathbf{x}_i$ denote the vector of explanatory variables for observation i:

$$\mathbf{x}_i = \big(\texttt{age\_entry}, \texttt{loan\_amount}, \texttt{loan\_term}, \texttt{rate}, \texttt{children}, \texttt{sex}, \texttt{smoker}\big). \tag{1}$$

The response variable is the total claim amount y_i, measuring the monetary value of losses associated with observation i.

2.1.1. Structure of the Response Variable The response variable y_i represents the total claim amount associated with contract i. In the analyzed portfolio, y_i is non-negative and exhibits a strongly zero-inflated distribution: out of the 2000 observations, only 40 contracts recorded a positive claim amount, while the remaining 1960 presented a zero outcome, corresponding to a zero-claim rate of approximately 98%. Far from being a statistical anomaly, this distributional feature is an intrinsic characteristic of credit life insurance portfolios. In such products, coverage is triggered exclusively by the borrower's death during the loan period — an event that remains rare over any fixed observation window, particularly when the insured population is relatively young or when contracts are observed over short durations. The resulting data structure combines a dominant point mass at zero with a right-skewed, heterogeneous distribution of positive claim amounts, a pattern that is well documented in the actuarial literature on aggregate loss modeling . This structure has direct implications for model evaluation: standard regression metrics such as R^2 or RMSE are sensitive to the mass of zeros and may overstate predictive performance if interpreted

without caution, since a degenerate model predicting zero for all observations would already achieve artificially favorable scores. Recognizing this, the present study adopts the Tweedie Generalized Linear Model as its actuarial benchmark, precisely because the Tweedie family — and in particular the compound Poisson–Gamma specification obtained for $1 < p < 2$ — provides a theoretically coherent framework for simultaneously capturing the probability of a zero claim and the conditional severity distribution over positive outcomes.

2.2. Data set splitting and normalization

To obtain robust estimates and assess generalization performance, the data set is randomly divided into two parts: 80% of the observations are used for training and the remaining 20% form the test set. The training sample is used to fit the models, while the test sample is reserved for evaluating their predictive accuracy on previously unseen data.

Because the predictors differ in units and scales, we standardize them using Min–Max normalization so that all transformed numerical variables lie within the interval $[0, 1]$:

$$X^{(i)}_{\text{normalized}} = \frac{X^{(i)} - X^{(i)}_{\min}}{X^{(i)}_{\max} - X^{(i)}_{\min}},$$

where $X^{(i)}_{\min}$ and $X^{(i)}_{\max}$ are respectively the minimum and maximum values of feature i, computed on the training set only and then applied consistently to the test set.

2.3. XGBoost theoretical model

XGBoost (eXtreme Gradient Boosting) is a gradient-boosted decision tree algorithm that models the prediction function as an additive ensemble of regression trees. For an input vector $\mathbf{x}_i$, the predicted response is

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{t=1}^{T} f_t(\mathbf{x}_i), \tag{2}$$

where each $f_t \in \mathcal{F}$ is a CART-type regression tree. The model is fitted by minimising the regularised empirical risk

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega(f_t), \tag{3}$$

with differentiable loss $l(\cdot, \cdot)$ and regulariser

$$\Omega(f_t) = \gamma T_t + \frac{1}{2}\lambda \sum_{j=1}^{T_t} w_{tj}^2, \tag{4}$$

which penalises the number of leaves T_t and squared leaf weights w_{tj}.

XGBoost is trained stage-wise using second-order gradient boosting. At iteration t, given current predictions $\hat{y}_i^{(t-1)}$, the added tree f_t is obtained by minimising a second-order Taylor approximation:

$$\tilde{\mathcal{L}}^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t(\mathbf{x}_i)^2 \right] + \Omega(f_t), \tag{5}$$

where

$$g_i = \left. \frac{\partial\, l(y_i, \hat{y})}{\partial \hat{y}} \right|_{\hat{y}=\hat{y}_i^{(t-1)}}, \qquad h_i = \left. \frac{\partial^2 l(y_i, \hat{y})}{\partial \hat{y}^2} \right|_{\hat{y}=\hat{y}_i^{(t-1)}}. \tag{6}$$

For a fixed tree structure (leaves indexed by j with instance sets I_j), the optimal leaf weight is

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}. \tag{7}$$

The prediction update after adding tree t is

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta\, f_t(\mathbf{x}_i), \tag{8}$$

where $\eta \in (0, 1]$ is the learning rate (shrinkage).

2.4. Support Vector Machine / Support Vector Regression (SVR)

Support Vector Machines rely on structural risk minimisation. Since our objective is claim severity prediction (a continuous response), we use Support Vector Regression (SVR). Given training data $\{(x_i, y_i)\}_{i=1}^n$ with $x_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$, SVR seeks a function $f(x) = w^\top x + b$ that is as flat as possible while allowing errors within an ε-insensitive tube. The primal optimisation problem is

$$\min_{w,b,\xi,\xi^*} \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{9}$$

subject to

$$\begin{aligned} y_i - (w^\top x_i + b) &\le \varepsilon + \xi_i, \\ (w^\top x_i + b) - y_i &\le \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* &\ge 0, \qquad i = 1, \dots, n. \end{aligned} \tag{10}$$

Nonlinear relationships are handled through the kernel trick, replacing inner products $x_i^\top x$ by a kernel function $K(x_i, x)$. The resulting predictor can be written as

$$f(x) = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i)\, K(x_i, x) + b, \tag{11}$$

where α_i, α_i^* are dual variables and only observations with $(\alpha_i^* - \alpha_i) \neq 0$ act as support vectors. Hyperparameters C, ε, and kernel parameters govern the bias–variance trade-off.

2.5. Multilayer Perceptron (MLP)

A multilayer perceptron (MLP) is a feed-forward neural network that learns a nonlinear mapping between predictors and a response through stacked affine transformations and nonlinear activations. Let $x \in \mathbb{R}^d$ denote the input vector. An MLP with L hidden layers defines

$$\begin{aligned} h^{(0)} &= x, \\ h^{(\ell)} &= \sigma^{(\ell)}\left(W^{(\ell)} h^{(\ell-1)} + b^{(\ell)}\right), \qquad \ell = 1, \dots, L, \\ f_\theta(x) &= W^{(L+1)} h^{(L)} + b^{(L+1)}, \end{aligned} \tag{12}$$

where $\sigma^{(\ell)}(\cdot)$ is a nonlinear activation (e.g. ReLU or $\tanh$) and $\theta = \{W^{(\ell)}, b^{(\ell)}\}_{\ell=1}^{L+1}$ is the set of trainable parameters.

Parameters are learned by empirical risk minimisation. For a loss function $\ell(\cdot, \cdot)$, the objective is

$$\min_\theta \ \frac{1}{n}\sum_{i=1}^{n} \ell(f_\theta(x_i), y_i) + \lambda\, \Omega(\theta), \tag{13}$$

where $\Omega(\theta)$ is a regularisation term (e.g. ℓ_2 weight decay). Optimisation is performed using gradient-based methods (SGD/Adam), with gradients computed efficiently via backpropagation.

2.6. Performance criteria for XGBoost, SVR and MLP models

The predictive performance of the three non-linear regression models—XGBoost, SVR and MLP—is assessed using R^2, MSE, RMSE, MAE, and the normalised mean squared error (NMSE). Let y_i be observed values, $\hat{y}_i$ predictions, and $\bar{y}$ the sample mean.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{14}$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}, \tag{15}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{16}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{17}$$

Pseudo-R^2 In quantile regression, the conventional coefficient of determination R^2 is not fully appropriate, since the model is estimated by minimizing the asymmetric absolute loss function rather than the sum of squared residuals. For this reason, a pseudo-R^2 measure is often employed to evaluate goodness-of-fit. This measure compares the loss achieved by the fitted quantile regression model with that of a benchmark model containing only an intercept. For a given quantile level $\tau \in (0,1)$, let $\hat{y}_i^{(\tau)}$ denote the predicted conditional quantile for observation i, and let $\hat{q}_\tau$ be the empirical τ-th quantile of the response variable obtained from the intercept-only model. The quantile loss function, also called the check function, is defined by

$$\rho_\tau(u) = u\big(\tau - \mathbf{1}(u < 0)\big),$$

where $\mathbf{1}(\cdot)$ is the indicator function. The pseudo-R^2 is then given by

$$R^2_{\mathrm{pseudo}}(\tau) = 1 - \frac{\sum_{i=1}^{n}\rho_\tau\left(y_i - \hat{y}_i^{(\tau)}\right)}{\sum_{i=1}^{n}\rho_\tau(y_i - \hat{q}_\tau)}.$$

This quantity measures the proportional improvement of the fitted quantile regression model relative to the intercept-only benchmark. A larger pseudo-R^2 indicates that the explanatory variables provide a better fit to the conditional quantile structure of the response variable. However, unlike the classical R^2 in ordinary least squares, pseudo-R^2 should not be interpreted as the proportion of total variance explained. Rather, it should be understood as a relative measure of improvement in quantile loss.

3. Results and Discussion

In this section, we first present the exploratory analysis of the dataset to support the interpretation of subsequent findings. We then proceed to the modelling stage, where algorithms are fitted and hyperparameters tuned using Randomized Search CV. Residual analysis contributes to risk reduction and improves the accuracy of claim amount predictions, so that estimated costs are more consistent with actual expenditures and can support efficient management of life credit insurance claims.

3.1. Explanatory Data Analysis

Figure 1 displays the Pearson correlation matrix for the numerical variables (`age_entry`, `loan amount`, `loan_term`, `rate` and `children`). Overall, correlations are of moderate magnitude, with a noticeable negative association between `loan amount` and `loan_term`, while remaining variables show weak linear relationships.

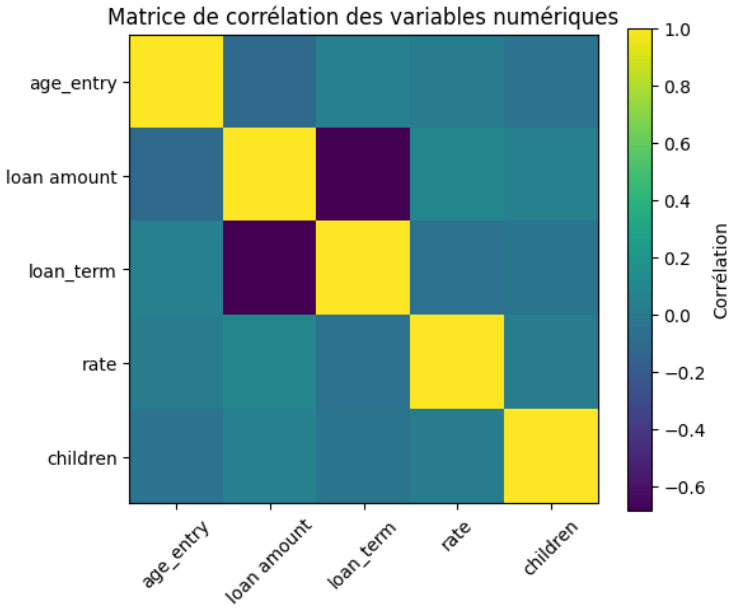


Figure 1. Correlation matrix of the numerical variables.

Figure 2 presents a heatmap of one-way ANOVA significance levels between the numerical variables and the categorical factors. More precisely, each cell reports the transformed significance level, $-\log_{10}(p\text{-value})$, obtained from testing whether the mean of a numerical variable differs across the categories of a given factor. Therefore, the figure should not be interpreted as a correlation plot, but rather as a graphical summary of the statistical evidence for between-group differences. Higher values indicate stronger evidence that the corresponding categorical factor is associated with variation in the numerical variable.

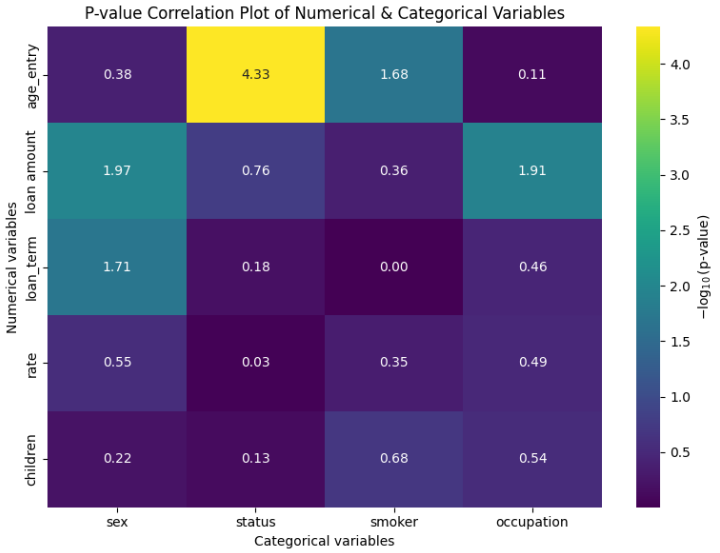


Figure 2. P-value correlation plot between numerical and categorical variables (colour scale in $-\log_{10}$(p-value)).

3.2. Hyperparameters with Modeling

3.2.1. Randomized Search Cross-Validation Results Randomized Search CV is employed to efficiently explore the hyperparameter space of each learning algorithm. Model accuracy is evaluated using R^2, MSE, RMSE and MAE, where relevant. This procedure yields near-optimal configurations for the XGBoost, SVR and MLP models, reported in Table 2.

Model	Parameters	Range	Best
XGBoost	Subsample	`[0.6, 0.8, 1.0]`	1.0
	reg-lambda	`[1.0, 1.5, 2.0]`	1.5
	reg-alpha	`[0, 0.1, 0.5, 1.0]`	1.0
	n-estimators	`[100, 200, 300, 400]`	400
	max-depth	`[3, 5, 7, 10]`	5
	learning-rate	`[0.01, 0.05, 0.1, 0.2]`	0.05
	colsample-bytree	`[0.6, 0.8, 1.0]`	0.6
SVR	Kernel	`['rbf', 'poly', 'sigmoid']`	rbf
	gamma	`['scale', 'auto', 0.01, 0.1, 1]`	auto
	epsilon	`[0.01, 0.1, 0.2, 0.5]`	0.01
	C	`[0.1, 1, 10, 100]`	10
MLP	Solver	`['adam', 'sgd']`	adam
	max-iter	`['adaptive', 'constant']`	adaptive
	hidden-layer-sizes	`[(50,), (100,), (50,50), (100,50)]`	(100,50)
	alpha	`[0.0001, 0.001, 0.01, 0.1]`	0.1
	activation	`['relu', 'tanh', 'logistic']`	tanh

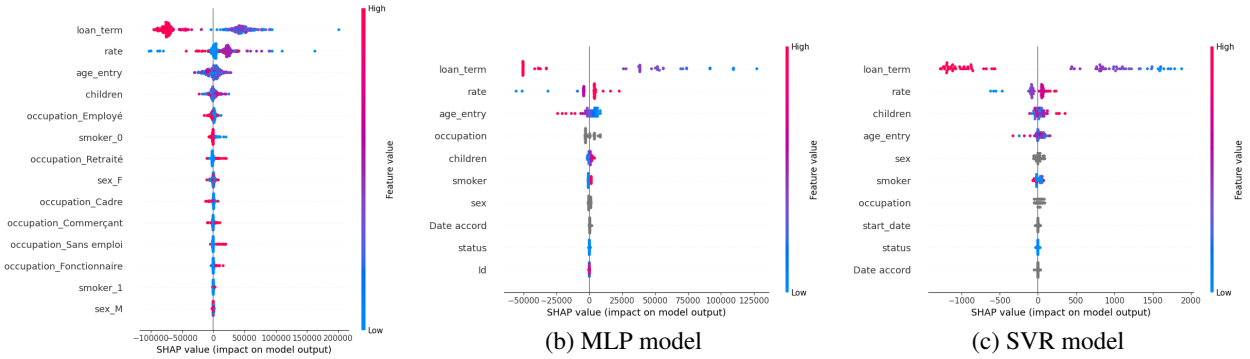
Table 2. The hyperparameters adjusted for the models in this study.

3.2.2. Hyperparameters Summary The hyperparameters are tuned using a blocked 5-fold cross-validation design to obtain robust and unbiased estimates of predictive performance. The data set is randomly partitioned into five folds, which are successively used for training and validation so as to limit sampling bias and preserve reproducibility. Table 2 summarizes the final hyperparameter values retained for each algorithm. For the XGBoost model, `subsample` denotes the fraction of observations used in each boosting round, `reg_alpha` is the L1 regularization term promoting feature selection, and `reg_lambda` is the L2 regularization term controlling overfitting. The parameter `n_estimators` corresponds to the number of trees, `max_depth` to their maximum depth, `learning_rate` to the step size of the boosting updates, and `colsample_bytree` to the proportion of predictors sampled at each tree.

3.2.3. SHAP Values Summary SHAP values quantify the contribution of each input variable to the model predictions, providing local explanations and a global ranking of feature importance.

3.2.4. Performance comparison of XGBoost, SVR and MLP In order to interpret and compare the three competing algorithms (XGBoost, MLP and SVR), Figure 3 indicates that contractual characteristics such as `loan_term` and `rate`, followed by `age_entry` and `children`, are dominant drivers of predicted claim amounts, whereas `sex` and `smoker` exert a more moderate but still non-negligible influence.

It is also worth noting that SVR exhibits a marked tendency toward under-dispersion: its predictions cluster in a narrow range around the mean, failing to reproduce the variability of observed claim amounts. This behavior, visible in the scatter plot of Figure 5, is a known limitation of SVR with RBF kernel when applied to highly skewed and zero-inflated targets, as the ε-insensitive loss penalizes deviations symmetrically and does not accommodate the asymmetric nature of the claim distribution.



(a) XGBoost model (b) MLP model (c) SVR model

Figure 3. SHAP summary plots for the XGBoost, MLP and SVR models.

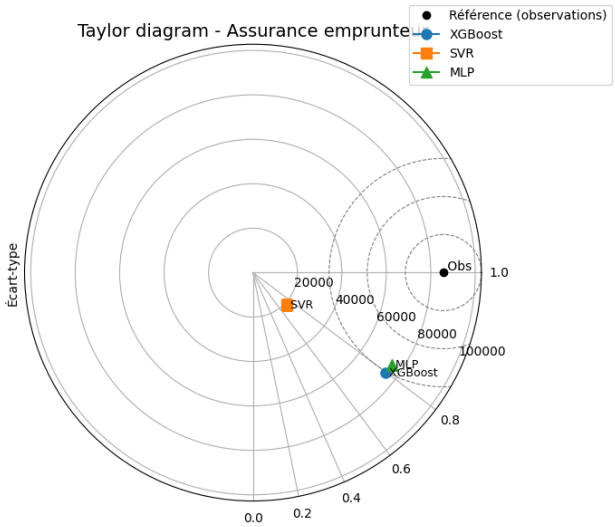


Figure 4. Taylor diagram comparing model performance.

3.2.5. Taylor Diagram The Taylor diagram in Figure 4 compares how well XGBoost, SVR and MLP reproduce the observed claim amounts. Models located closer to the reference "Obs" point have higher correlation with the data, a standard deviation close to the observations, and lower error. Both XGBoost and, in particular, MLP lie close to the reference point, while SVR is farther away with a smaller standard deviation, indicating weaker agreement and under-dispersed predictions.

3.3. Model Performance and Predictive Diagnostics

Referring to Table 3, XGB and MLP display similar and high goodness-of-fit on the training data ($R^2 \approx 0.84$), whereas SVR performs worse and exhibits substantially higher errors. On the validation set, the MLP model achieves the best generalisation ability, as reflected by the highest R^2 (0.681) and the lowest error metrics. XGB remains competitive but shows a slight deterioration in performance, while SVR continues to deliver the poorest results.

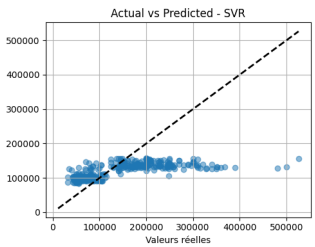
Model	Dataset	R^2	MSE	RMSE	MAE
XGB	Train	0.837	1474105612.2993	38394.0830	27950.0562
XGB	Valid	0.629	2715232997.2702	52107.8976	34485.3118
SVR	Train	0.177	7439870532.7435	86254.6841	58389.9576
SVR	Valid	0.229	5638870925.9305	75092.4159	51954.1253
MLP	Train	0.836	1482079406.6683	38497.7844	19209.4330
MLP	Valid	0.681	2333332808.6512	48304.5837	31839.6648

Table 3. Performance metrics for XGB, SVR and MLP models on training and validation sets.

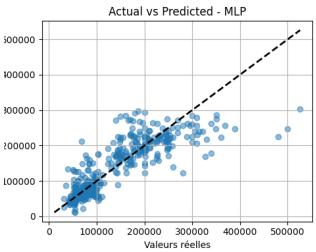
3.4. Prediction and managing life credit insurance claims

These three scatter plots compare observed claim amounts (x-axis) with predicted values (y-axis) for the MLP, SVR and XGBoost models; the dashed diagonal represents perfect predictions. MLP points are fairly aligned with the diagonal for small and medium claims; SVR forms a nearly horizontal band, reflecting low variability and underestimation of large claims; XGBoost follows the diagonal more closely than SVR but still shows dispersion for the upper tail.

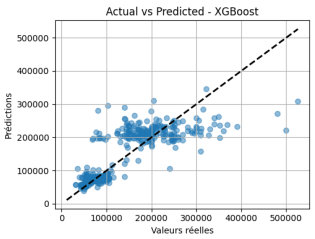
It should be noted that all three models exhibit increased dispersion for large claim amounts, as visible in the upper tail of Figure 5. This limitation is consistent with the sparse and heterogeneous nature of extreme losses in credit life insurance portfolios and motivates the complementary use of quantile regression in Section 3.6 for a more targeted assessment of tail behavior.



Actual vs Predicted – SVR



Actual vs Predicted – MLP



Actual vs Predicted – XGBoost

Figure 5. Actual versus predicted claim amounts for the MLP, SVR and XGBoost models.

3.5. Tweedie generalized linear model

The Tweedie GLM is introduced in this study as the primary actuarial benchmark against which the ML models are evaluated. Unlike the ML regressors, which optimize predictive accuracy through flexible nonlinear mappings, the Tweedie GLM provides a theoretically grounded and interpretable pricing framework that is standard practice in life insurance. Its inclusion allows a rigorous assessment of whether the added complexity of ML methods translates into meaningful gains in predictive performance relative to an established actuarial baseline.

The Tweedie generalized linear model (GLM) is well suited to insurance data because it accommodates non-negative and right-skewed responses and, for certain values of the variance power parameter, allows a positive probability mass at zero. This makes it particularly relevant for modeling claim amounts, aggregate losses, and pure premiums.

Let Y_i denote the response variable associated with observation i, and let $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik})^\top$ be the corresponding vector of explanatory variables. The Tweedie GLM assumes that, conditionally on $\mathbf{x}_i$, the response belongs to the class of exponential dispersion models with density

$$f(y_i; \theta_i, \phi) = a(y_i, \phi, p) \exp\left\{ \frac{y_i \theta_i - \kappa(\theta_i)}{\phi} \right\},$$

where θ_i is the canonical parameter, $\phi > 0$ is the dispersion parameter, p is the variance power parameter, and $\kappa(\cdot)$ is the cumulant function.

The fundamental property of the Tweedie family is the mean–variance relationship

$$\mathbb{E}(Y_i \mid \mathbf{x}_i) = \mu_i, \qquad \mathrm{Var}(Y_i \mid \mathbf{x}_i) = \phi \mu_i^p.$$

The systematic component of the GLM is defined by

$$\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta},$$

and the mean response is linked to the predictors through a link function g, so that

$$g(\mu_i) = \eta_i.$$

In actuarial applications, the logarithmic link is commonly adopted:

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

which yields

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

Under this specification, the regression coefficients have a multiplicative interpretation: a one-unit increase in a covariate x_{ij} multiplies the conditional mean by $\exp(\beta_j)$, all else being equal.

The Tweedie family contains several important distributions as special cases:

$p = 0$	Normal distribution,
$p = 1$	Poisson distribution,
$1 < p < 2$	Compound Poisson–Gamma distribution,
$p = 2$	Gamma distribution,
$p = 3$	Inverse Gaussian distribution.

The case $1 < p < 2$ is especially important in insurance modeling, since it corresponds to a compound Poisson–Gamma distribution. In this case, the response may be represented as

$$Y_i = \sum_{m=1}^{N_i} Z_{im},$$

where

$$N_i \sim \mathrm{Poisson}(\lambda_i),$$

and the severities Z_{im} are independent Gamma random variables. This representation is appealing in actuarial science because it naturally reflects a random aggregate loss composed of a random number of positive claim components.

For $p \neq 1$ and $p \neq 2$, the canonical parameter may be written as

$$\theta_i = \frac{\mu_i^{1-p}}{1-p},$$

and the cumulant function satisfies

$$\kappa(\theta_i) = \frac{\mu_i^{2-p}}{2-p}.$$

Consequently,

$$\mu_i = \kappa'(\theta_i), \qquad \mathrm{Var}(Y_i \mid \mathbf{x}_i) = \phi \kappa''(\theta_i) = \phi \mu_i^p.$$

Letting $y_1, \ldots, y_n$ denote independent observations, the log-likelihood is given by

$$\ell(\boldsymbol{\beta}, \phi, p) = \sum_{i=1}^{n} \log f(y_i; \mu_i, \phi, p), \qquad \mu_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta}).$$

The model parameters are then estimated by maximizing this log-likelihood, that is,

$$(\widehat{\boldsymbol{\beta}}, \widehat{\phi}, \widehat{p}) = \arg\max_{\boldsymbol{\beta}, \phi, p} \ell(\boldsymbol{\beta}, \phi, p).$$

In the context of life credit insurance, if Y_i denotes the claim amount associated with borrower i, the Tweedie GLM can be written as

$$Y_i \mid \mathbf{x}_i \sim \text{Tweedie}(\mu_i, \phi, p),$$

with

$$\log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik},$$

where the explanatory variables may include age at entry, sex, loan amount, loan term, interest rate, smoking status, number of children. The Tweedie specification is particularly attractive because it captures both the conditional mean structure and the heteroscedastic nature of insurance losses through the variance function

$$\text{Var}(Y_i \mid \mathbf{x}_i) = \phi \mu_i^p.$$

3.5.1. Training cross-validation and test hold-out evaluation framework Let the full data set be denoted by

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n},$$

where $x_i \in \mathbb{R}^p$ represents the vector of explanatory variables for observation i, and y_i denotes the corresponding response variable. To assess predictive performance in a rigorous way, the sample is first divided into two disjoint subsets: a training set $\mathcal{D}_{\text{train}}$ and a test hold-out set $\mathcal{D}_{\text{test}}$, with

$$\mathcal{D} = \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{test}}, \qquad \mathcal{D}_{\text{train}} \cap \mathcal{D}_{\text{test}} = \varnothing.$$

The training set is used for model estimation and internal validation, whereas the test hold-out set is reserved exclusively for the final out-of-sample evaluation. In order to obtain a stable estimate of model performance on the training data, K-fold cross-validation is applied within $\mathcal{D}_{\text{train}}$. More precisely, the training set is partitioned into K approximately equal-sized folds,

$$\mathcal{D}_{\text{train}} = \bigcup_{k=1}^{K} \mathcal{D}_k.$$

At iteration k, the model is trained on

$$\mathcal{D}_{\text{train}}^{(-k)} = \mathcal{D}_{\text{train}} \setminus \mathcal{D}_k,$$

and validated on the fold $\mathcal{D}_k$. This process is repeated for $k = 1, \ldots, K$, and the cross-validation performance is obtained by averaging the validation results over all folds.

If $\widehat{y}_i$ denotes the prediction of the fitted model for observation i, the predictive accuracy is assessed using the following criteria:

After cross-validation, the final model is refitted on the entire training set ($\mathcal{D}_{\text{train}}$) and then evaluated on the independent test hold-out set ($\mathcal{D}_{\text{test}}$). This procedure ensures that model selection is carried out using the training data only, while the test hold-out set provides an unbiased estimate of the model's generalization ability on unseen observations.

Table 4. Predictive performance of the Tweedie GLM on the training cross-validation and test hold-out sets

Metric	Train Cross-Validation	Test Hold-out
MAE	47,233.66	42,600.24
MSE	5,703,833,000.00	4,268,525,000.00
RMSE	74,924.12	65,333.95
R^2	0.3748	0.4167

3.5.2. Training and test performance Results A direct comparison of predictive performance reveals that the ML models outperform the Tweedie GLM on the test set. Specifically, the MLP achieves a validation RMSE of 48,304, and XGBoost a validation RMSE of 52,107, both substantially lower than the Tweedie test RMSE of 65,334. Similarly, the MLP validation MAE of 31,840 compares favorably against the Tweedie test MAE of 42,600. These results confirm that the ML regressors capture nonlinear patterns in the data more effectively than the parametric GLM benchmark, at the cost of reduced interpretability

The estimated variance power parameter $\hat{p} = 1.5$ indicates that the fitted model lies within the compound Poisson–Gamma regime $(1 < p < 2)$, which is well suited to capture the zero-inflated and right-skewed nature of the observed claim amounts. The corresponding dispersion parameter is estimated as $\hat{\phi} = 1.11$. Table 5 presents the estimated regression coefficients together with their multiplicative effects, given by $\exp(\beta_j)$, on the conditional mean claim amount.

These results suggest that the Tweedie GLM provides a stable and moderately accurate fit for claim amount prediction. The fact that the hold-out performance is slightly better than the training cross-validation performance may reflect favorable sample partitioning or variation between folds rather than over-optimism. Nevertheless, the R^2 values remain moderate, indicating that although the model captures an important part of the claim severity structure, a non-negligible share of variability is still left unexplained.

The Tweedie GLM is used in this study because it provides a theoretically grounded and actuarially relevant framework for modeling claim amounts, which are typically non-negative, right-skewed, and may include a mass at zero. In particular, for $1 < p < 2$, the Tweedie distribution corresponds to a compound Poisson–Gamma model, making it well suited for insurance pricing. Moreover, the Tweedie GLM offers an interpretable benchmark through its log-link specification, allowing the effects of risk factors on expected claim costs to be clearly quantified. Its inclusion is therefore essential for comparing the predictive performance of machine learning methods against a standard and widely accepted actuarial pricing model.

Table 5. Estimated Tweedie GLM coefficients

Variable	Coefficient (β)	Std. Error	$\exp(\beta)$	P-value
(Intercept)	13.4122	0.1542	668 131.08	¡ 0.001***
age_entry	-0.0024	0.0008	0.9976	0.002**
loan_term	-0.0266	0.0011	0.9737	¡ 0.001***
rate	-0.0715	0.0054	0.9310	¡ 0.001***
children	-0.0008	0.0031	0.9992	0.798
sex_M	-0.0003	0.0045	0.9997	0.947
smoker_1	0.0064	0.0052	1.0065	0.218
*Significativité : *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$*				
Paramètre de puissance Tweedie (p) : 1.5				

Significant Factors: Interest rate, Loan term, and Age at entry. They all show a negative correlation, i.e., $\exp(\beta) < 1$, with the predicted value.

Most Impactful: The Interest Rate. A 1-unit increase reduces the predicted mean by approximately 6.9% $(1 - 0.9310)$.

Non-Significant: Children, Gender, and Smoking status have no statistically reliable impact on the outcome in this specific dataset.

3.6. Quantile regression

Quantile regression provides a flexible alternative to classical mean regression by modeling the conditional quantiles of a response variable rather than its conditional expectation. This approach is particularly useful when the impact of explanatory variables differs across the distribution of the outcome, or when the response exhibits skewness, heavy tails, or heteroscedasticity.

Let Y_i denote the response variable and let $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ik})^\top$ be the vector of explanatory variables associated with observation i, for $i = 1, \ldots, n$. For a given quantile level $\tau \in (0, 1)$, the conditional quantile function of Y_i given $\mathbf{x}_i$ is defined by

$$Q_Y(\tau \mid \mathbf{x}_i) = \inf\{y \in \mathbb{R} : F_{Y|X}(y \mid \mathbf{x}_i) \geq \tau\},$$

where $F_{Y|X}$ denotes the conditional distribution function of Y given X.

The linear quantile regression model is specified as

$$Q_Y(\tau \mid \mathbf{x}_i) = \mathbf{x}_i^\top \boldsymbol{\beta}(\tau),$$

where $\boldsymbol{\beta}(\tau)$ is a vector of quantile-specific parameters. In contrast to ordinary least squares, the coefficients depend on τ, allowing the effect of a covariate to vary across different parts of the conditional distribution of the response.

The estimator $\widehat{\boldsymbol{\beta}}(\tau)$ is obtained by solving

$$\widehat{\boldsymbol{\beta}}(\tau) = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{k+1}} \sum_{i=1}^{n} \rho_\tau\big(Y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\big),$$

where $\rho_\tau(\cdot)$ is the check loss function defined by

$$\rho_\tau(u) = u\big(\tau - \mathbf{1}_{\{u<0\}}\big).$$

Equivalently,

$$\rho_\tau(u) = \begin{cases} \tau u, & \text{if } u \geq 0, \\ (\tau - 1)u, & \text{if } u < 0. \end{cases}$$

This asymmetric loss function assigns different weights to positive and negative residuals, thereby making quantile regression robust to outliers and suitable for asymmetric response distributions. In particular, when $\tau = 0.5$, the method yields median regression, which minimizes the sum of absolute deviations.

An important advantage of quantile regression is that it allows the marginal effect of a regressor to differ across quantiles. For a continuous explanatory variable x_j, the coefficient $\beta_j(\tau)$ measures the variation in the conditional τ-th quantile of Y associated with a one-unit increase in x_j, all else being equal. Consequently, quantile regression provides a richer description of the conditional distribution of the response than mean-based models.

Under suitable regularity conditions, the estimator is consistent and asymptotically normal:

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right) \xrightarrow{d} \mathcal{N}\big(0, \Sigma(\tau)\big),$$

where the covariance matrix $\Sigma(\tau)$ depends on the conditional density of the response at the quantile level τ. In empirical applications, inference is commonly based on bootstrap standard errors.

Overall, quantile regression constitutes a robust and semi-parametric framework for analyzing heterogeneous covariate effects across the entire conditional distribution of the response variable. It is particularly relevant in actuarial and financial applications, where the lower, central, and upper tails of the distribution may be of direct practical interest.

Table 6. Quantile regression performance across different quantiles

Quantile	MAE	RMSE	Pseudo-R^2	Pinball Loss
0.10	59867	86370	0.1909	7401.91
0.25	45157	70616	0.3146	14304.28
0.50	38929	60231	0.4311	19464.74
0.75	50300	70707	0.4134	17520.94
0.90	76137	101826	0.3759	11027.49

3.6.1. Quantile regression performance metrics By contrast, predictive performance deteriorates toward the tails. At the lower quantile $\tau = 0.10$, the model yields the lowest pseudo-R^2 (0.1909), suggesting limited explanatory power for small claim amounts. At the upper quantile $\tau = 0.90$, the model records the largest prediction errors, with an MAE of 76,137 and an RMSE of 101,826, despite a moderate pseudo-R^2 of 0.3759. This pattern is consistent with the typical structure of insurance claim data, where extreme losses are relatively sparse and highly heterogeneous, making the lower and upper tails more difficult to estimate accurately. Overall, these findings suggest that quantile regression is particularly effective for modeling the central region of the claim amount distribution,whereas predictive accuracy is weaker at extreme quantiles. This deterioration at the tails is consistent with the statistical properties of the portfolio: extreme losses are rare, highly heterogeneous, and not well captured by the available covariates alone. From an actuarial perspective, this result underscores the need for larger datasets or additional risk factors — such as medical underwriting data or loan-to-value ratios — to improve tail risk estimation in future work.

Quantile regression is used in this study because it provides a more comprehensive analysis of claim amounts across different points of the conditional distribution, rather than focusing only on the conditional mean. This is particularly relevant for insurance data, which are often skewed, heterogeneous, and affected by extreme values. Moreover, quantile regression is robust to outliers and allows the effects of explanatory variables to vary between low, median, and high claim levels, making it a valuable benchmark for evaluating predictive performance across the entire distribution.

It should be noted that the pseudo-R^2 values reported for quantile regression are not directly comparable to the R^2 values obtained from the ML models or the Tweedie GLM. While the classical R^2 measures the proportion of variance explained relative to a mean-based benchmark, the pseudo-R^2 measures the proportional reduction in the asymmetric quantile loss (pinball loss) relative to an intercept-only model. Consequently, pseudo-R^2 values are systematically lower and reflect a different aspect of predictive performance, namely the ability to correctly locate conditional quantiles of the distribution rather than to minimize squared errors. Regarding statistical inference, the quantile regression coefficients were estimated with bootstrap standard errors (500 replications). The results confirm that `loan_term` and `rate` remain statistically significant across all quantile levels, while `children`, `sex`, and `smoker` are non-significant at conventional levels, a finding consistent with the Tweedie GLM results reported in Table 5.

4. Conclusion and Future Research

This study developed a comparative framework for predicting and managing life credit insurance claim amounts using both machine learning and actuarial-statistical models. Based on the empirical analysis, the results show that the Multi-Layer Perceptron (MLP) achieved the best overall predictive performance among the machine learning models, with stronger generalization ability and lower validation errors than XGBoost and Support Vector Regression (SVR). XGBoost also produced competitive results and captured important nonlinear relationships, whereas SVR showed weaker predictive accuracy and a tendency to under-dispersion in its predictions. In addition, the SHAP analysis highlighted that contractual variables such as loan term and interest rate, followed by age at entry and number of children, were among the most influential drivers of predicted claim amounts.

Beyond the comparison of machine learning methods, the study also showed the relevance of complementary actuarial approaches. The Tweedie generalized linear model provided a theoretically grounded and interpretable

benchmark for insurance pricing, while quantile regression offered additional insights into the heterogeneous behavior of claim amounts across the conditional distribution. In particular, quantile regression performed best around the median, but its predictive quality deteriorated in the tails, especially for extreme high claims. Overall, these findings confirm that combining flexible machine learning tools with interpretable actuarial models can improve claim prediction, support more accurate pricing, and strengthen risk management in life credit insurance.

It is worth emphasizing that the extreme zero-inflation observed in this portfolio — where 98% of contracts recorded no claim over the observation period — represents both the central empirical challenge and the most practically relevant feature of this study. In credit life insurance, the rarity of the death event over any fixed observation window means that predictive models are, by construction, operating in a highly imbalanced setting.

This has two important implications. First, aggregate performance metrics such as R^2 should be understood as reflecting the models' ability to reproduce the overall claim distribution — including its dominant zero mass — rather than their precision in predicting individual positive claim amounts. Second, and more importantly from an operational standpoint, the identification of the small fraction of contracts likely to generate a claim is precisely where accurate modeling delivers the highest actuarial value, in terms of pricing adequacy, reserve sufficiency, and risk selection.

Future work should therefore prioritize methods specifically designed for this regime, such as two-part hurdle models , zero-inflated regression frameworks, or cost-sensitive learning approaches, which explicitly separate the modeling of claim occurrence from claim severity.

This has two important implications. First, aggregate performance metrics such as R^2 should be understood as reflecting the models' ability to reproduce the overall claim distribution — including its dominant zero mass — rather than their precision in predicting individual positive claim amounts. Second, and more importantly from an operational standpoint, the identification of the small fraction of contracts likely to generate a claim is precisely where accurate modeling delivers the highest actuarial value, in terms of pricing adequacy, reserve sufficiency, and risk selection. Future work should therefore prioritize methods specifically designed for this regime, such as two-part hurdle models , zero-inflated regression frameworks , or cost-sensitive learning approaches, which explicitly separate the modeling of claim occurrence from claim severity.

These results should, however, be interpreted in light of several limitations. The dataset comprises 2,000 contracts from a single Moroccan insurer, which may limit the external validity of the findings. Furthermore, the direct regression approach applied to the full claim amount — including zeros — does not fully exploit the two-part structure of the data, and a dedicated frequency-severity implementation remains an important avenue for future work. Finally, the small sample size may affect the stability of the MLP estimates in particular, as neural networks typically benefit from larger training sets.

Several directions may extend this work. First, future research could rely on larger and more heterogeneous datasets, including additional borrower, contractual, medical, or macroeconomic variables, in order to improve the prediction of extreme losses. Second, more advanced actuarial structures could be explored, such as two-part, hurdle, or frequency–severity models, to better distinguish between claim occurrence and claim size. Third, further comparisons could be conducted with other machine learning and deep learning techniques, including random forests, LightGBM, CatBoost, and recurrent or attention-based neural architectures. Fourth, model governance could be reinforced through richer explainability tools, fairness analysis, and stability tests across sub-portfolios and time periods. Finally, future studies may investigate hybrid frameworks that combine the interpretability of generalized linear models with the flexibility of machine learning algorithms, thereby providing a more robust foundation for pricing, reserving, and claims management in credit life insurance.

REFERENCES

1. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 785–794.
2. J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
3. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, 2009.

4. C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
5. H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 9, 1997.
6. A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.
7. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986.
8. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
9. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
10. K. E. Taylor, "Summarizing multiple aspects of model performance in a single diagram," *Journal of Geophysical Research: Atmospheres*, vol. 106, no. D7, pp. 7183–7192, 2001.
11. J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.
12. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. VanderPlas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
13. J. G. Cragg, "Some statistical models for limited dependent variables with application to the demand for durable goods," *Econometrica*, vol. 39, no. 5, pp. 829–844, 1971.
14. J. Mullahy, "Specification and testing of some modified count data models," *Journal of Econometrics*, vol. 33, no. 3, pp. 341–365, 1986.
15. R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," in *Advances in Neural Information Processing Systems (NeurIPS)*, 1991.
16. M. I. Jordan and R. A. Jacobs, "Hierarchical mixtures of experts and the EM algorithm," *Neural Computation*, vol. 6, no. 2, pp. 181–214, 1994.
17. E. W. Frees, *Regression Modeling with Actuarial and Financial Applications*. Cambridge University Press, 2014.
18. M. Denuit, X. Marechal, S. Pitrebois, and J.-F. Walhin, *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. Wiley, 2007.
19. S. A. Klugman, H. H. Panjer, and G. E. Willmot, *Loss Models: From Data to Decisions*, 4th ed. Wiley, 2012.
20. E. Ohlsson and B. Johansson, *Non-Life Insurance Pricing with Generalized Linear Models*. Springer, 2010.
21. M. V. Wüthrich and M. Merz, *Stochastic Claims Reserving Methods in Insurance*. Wiley, 2008.
22. T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.
23. P. K. Dunn and G. K. Smyth, "Randomized quantile residuals," *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 236–244, 1996.
24. B. Gompertz, "On the nature of the function expressive of the law of human mortality," *Philosophical Transactions of the Royal Society of London*, vol. 115, pp. 513–583, 1825.
25. W. M. Makeham, "On the law of mortality and the construction of annuity tables," *Journal of the Institute of Actuaries*, vol. 8, pp. 301–310, 1860.
26. R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.
27. G. K. Smyth and B. Jørgensen, "Fitting Tweedie's compound Poisson model to insurance claims data," *ASTIN Bulletin*, vol. 32, no. 1, pp. 143–157, 2002.
28. R. Koenker, *Quantile Regression*, Cambridge University Press, Cambridge, 2005.