The Advanced Actuarial Data Science Based AI-Driven Solutions for Automated Loss Reserving Under IFRS 17 in Non-Life Insurance

Brighton Mahohoho[1], Charles Chimedza[2], Florance Matarise[1], Sheunesu Munyira[1]

[1]*University of Zimbabwe, Department of Mathematics & Computational Sciences, Harare, Zimbabwe.*
[2]*University of Witwatersrand, School of Statistics & Actuarial Science, Johannesburg, South Africa.*

Abstract

This study introduces an AI-driven Automated Actuarial Loss Reserving Model (AALRM) designed to meet IFRS 17 standards for non-life insurance. The model leverages advanced machine learning techniques to improve accuracy, efficiency, and adaptability in loss reserves, with a specific focus on inflation-adjusted frequency-severity modeling. A unique aspect of this research is the integration of bancassurance services, enabling automated management for both microfinance and car insurance on a unified platform. This includes a no-claims bonus system that categorizes policyholders into four tiers—base, variable, final, and high-bonus—resulting in more precise risk assessments and enhanced customer retention. Among eight evaluated machine learning algorithms, the Random Forest (RANGER) outperformed others for estimating Aggregate Comprehensive Automated Actuarial Loss Reserves (ACAALR). The model's effectiveness was validated through stress tests, scenario analyses, and comparisons with traditional methods like the Chain Ladder. Additionally, the study introduces a novel Robust Automated Actuarial Loss Reserve Margin (RAALRM) with adaptive bounds, addressing traditional limitations in reserve margin calculations. This AI-integrated approach significantly improves predictive accuracy, operational efficiency, and strategic decision-making, offering a scalable solution for the insurance industry.

Keywords*: Actuarial Data Science, Artificial Intelligence, Data Analytics, Automated Actuarial Loss Reserves, Actuaries, Machine Learning.*

1. Introduction

In the ever-evolving landscape of non-life insurance, the precision and efficiency of loss reserving are critical for maintaining financial stability and ensuring regulatory compliance. Traditional actuarial methods, while well-established, face significant challenges in managing the increasing complexity and volume of data. The advent of artificial intelligence (AI) and data science offers innovative solutions to these challenges, promising more accurate and automated loss reserving processes. This paper explores how actuarial data science-based AI solutions can revolutionize automated loss reserving in non-life insurance, highlighting their potential to enhance precision, reduce operational costs, and improve decision-making.

Loss reserving in non-life insurance involves estimating the amount of money an insurer needs to set aside to pay future claims. Traditionally, this process relies on actuarial methods such as chain-ladder models, frequency-severity models, and other statistical techniques [1]. However, these traditional methods can be limited by their reliance on historical data and assumptions that may not capture the complexities of emerging risks and changing patterns.

Actuarial data science introduces a new paradigm by leveraging advanced machine learning algorithms, big

Corresponding author: *Brighton Mahohoho (brightonmahohoho.07@gmail.com)*
Received: *19 September 2024*; Revised: *2 December 2025*; Accepted: *4 March 2025*; Published: 5 *May 2025*

data analytics, and artificial intelligence to enhance the accuracy and efficiency of loss reserving. Machine learning models, such as neural networks and gradient boosting machines, can process vast amounts of data, identify intricate patterns, and make predictions with higher accuracy than conventional methods [2]. By incorporating AI, insurers can automate the reserving process, reduce human error, and adapt more swiftly to changing market conditions.

The rationale behind integrating AI solutions into actuarial loss reserving stems from several key factors. First, the volume and complexity of data in non-life insurance have grown exponentially, driven by digitalization and the increasing availability of granular data [3]. Traditional methods, often constrained by manual processes and limited data handling capacity, may struggle to keep pace with these changes.

Second, the financial implications of inaccurate loss reserving are substantial. Under-reserving can lead to significant financial shortfalls, while over-reserving ties up capital that could be used more productively. AI-based solutions offer the potential to improve accuracy by continuously learning from new data and adjusting predictions accordingly, thus mitigating these risks.

In the realm of non-life insurance, the accurate estimation of loss reserves is critical for ensuring financial stability and regulatory compliance. Traditional actuarial reserving methods, such as the Chain Ladder method, Bornhuetter-Ferguson, and Cape Cod, have been the cornerstone of the industry for decades. However, these approaches face significant limitations in today's complex and rapidly changing risk landscape, characterized by increased volatility, inflation, and diverse coverage needs [4]. The deterministic nature of traditional models often struggles to capture the intricate dynamics of emerging risks, particularly when it comes to non-linear interactions between multiple variables [5].

One of the primary challenges is the reliance on aggregated data and simplistic assumptions about claim development, which can lead to underestimation or overestimation of reserves. Traditional methods often assume that past data trends will persist, limiting their responsiveness to evolving patterns, such as inflationary pressures or changes in policyholder behavior. This lack of flexibility in incorporating new information can result in inaccurate reserve estimates, potentially leading to insufficient capital adequacy or over-conservative provisioning.

Furthermore, the manual and static nature of conventional reserving techniques poses operational inefficiencies and limits the ability to quickly adapt to unexpected market changes or regulatory shifts. This is especially problematic under the requirements of IFRS 17, which mandates a higher level of transparency, accuracy, and forward-looking projections for reserves. The limitations of traditional methods are particularly evident in the areas of inflation adjustment, frequency-severity analysis, and the estimation of reserve uncertainty.

To address these challenges, this paper introduces an AI-driven Automated Actuarial Loss Reserving Model (AALRM) that leverages advanced machine learning techniques to enhance predictive accuracy, operational efficiency, and adaptability. By integrating inflation-adjusted frequency-severity modeling, the proposed approach offers a dynamic and data-rich framework capable of learning from historical data and incorporating real-time information to improve reserve estimates. This AI-based solution not only surpasses the limitations of traditional methods but also aligns with the evolving regulatory landscape under IFRS 17, ensuring that reserves are both adequate and adaptable to future uncertainties.

1.1. Traditional actuarial loss reserving methods

In actuarial science, loss reserving methods are crucial for estimating the reserves that an insurance company needs to set aside to pay for future claims. These methods are designed to predict the ultimate cost of claims based on historical data and various statistical techniques. This section describes some of the prominent loss reserving methods, including their theoretical underpinnings, algorithms, and practical implementations.

1.1.1. Chain-Ladder Method

The Chain-Ladder method is one of the most widely used techniques in actuarial science for loss reserving. It relies on the assumption that the development of claims over time follows a predictable pattern. The Chain-Ladder method assumes that the ratio of development factors (i.e.,

the ratios of cumulative claims between successive periods) is constant across different accident years. The chain ladder model is a popular method used in actuarial science for predicting future claims based on past data. The model assumes that the development of claims over time follows a certain pattern that can be extrapolated to estimate future claims. The basic structure of the chain ladder is presented by Table 1.

Table 1. Structure of the Basic Chain Ladder Model.

Accident Year	Development Lag 1	Development Lag 2	…	Development Lag n
Year 1	$C_{1,1}$	$C_{1,2}$	⋯	$C_{1,n}$
Year 2	$C_{2,1}$	$C_{2,2}$	⋯	$C_{2,n-1}$
⋮	⋮	⋮	⋱	⋮
Year m	$C_{m,1}$	$C_{m,2}$	⋯	$C_{m,n-m+1}$

Let $C_{i,j}$ represent the cumulative claims reported in Accident Year i at Development Lag j. The chain ladder model assumes that the development of claims follows a pattern that can be described by the development factors.

The development factors are calculated as:

$$f_j = \frac{\sum_{i=1}^{m-j} C_{i,j+1}}{\sum_{i=1}^{m-j} C_{i,j}}, \quad \text{for } j = 1,2,\dots,n-1. \tag{1}$$

where f_j is the development factor for lag j.

To project future claims, we use the development factors. For a given accident year i and development lag j, the projected claim $\hat{C}_{i,j}$ can be calculated as:

$$\hat{C}_{i,j} = C_{i,j-1} \cdot f_{j-1}, \quad \text{for } j > 1. \tag{2}$$

where $\hat{C}_{i,j}$ is the projected cumulative claim for Accident Year i at Development Lag j.

The total claims for Accident Year i, denoted $\hat{T}_i$, is the sum of the projected claims across all development lags:

$$\hat{T}_i = \sum_{j=1}^{n} \hat{C}_{i,j}. \tag{3}$$

where $\hat{T}_i$ is the total projected claims for Accident Year i.

The development factor f_j can be used to project future claims for any development lag j, given the claims from the previous lag.

Algorithm 1 Chain-Ladder Method

Input: Cumulative claims data matrix $\boldsymbol{C}$
Output: Reserves for each accident year Compute development factors $\boldsymbol{f_j}$ from $\boldsymbol{C}$
for each accident year $\boldsymbol{i}$ **do**
Estimate reserve $\boldsymbol{R_i}$ using $\boldsymbol{R_i = C_{i,0} \cdot \prod_{j=1}^{m} f_j - C_{i,m}}$
end for=0

Proof. To prove this, consider the claims for Accident Year i at Development Lag j, denoted $C_{i,j}$. For the next development lag $j+1$, the claims are given by:

$$\hat{C}_{i,j+1} = C_{i,j} \cdot f_j. \tag{4}$$

Given that the development factor f_j is calculated as the ratio of the total claims in lag $j+1$ to lag j across all accident years, it can be shown that:

$$f_j = \frac{\sum_{i=1}^{m-j} C_{i,j+1}}{\sum_{i=1}^{m-j} C_{i,j}}. \tag{5}$$

This ensures that:

$$\hat{C}_{i,j+1} = C_{i,j} \cdot \frac{\sum_{i=1}^{m-j} C_{i,j+1}}{\sum_{i=1}^{m-j} C_{i,j}}. \tag{6}$$

Therefore, $\hat{C}_{i,j+1}$ is a valid projection for future claims.

The total projected claims $\hat{T}_i$ for Accident Year i is the sum of the projected claims across all development lags as presented by Equation (3).

The total claims for Accident Year i can be expressed as:

$$\hat{T}_i = \sum_{j=1}^{n} C_{i,j} \cdot \prod_{k=1}^{j-1} f_k. \tag{7}$$

Proof. To derive this result, we use the fact that the projected claims $\hat{C}_{i,j}$ for lag j are given by:

$$\hat{C}_{i,j} = C_{i,1} \cdot \prod_{k=1}^{j-1} f_k. \tag{8}$$

Thus, the total projected claims for Accident Year i is:

$$\hat{T}_i = \sum_{j=1}^{n} C_{i,1} \cdot \prod_{k=1}^{j-1} f_k. \tag{9}$$

Simplifying the summation using the properties of the development factors leads to:

$$\hat{T}_i = C_{i,1} \cdot \left(1 + \sum_{j=1}^{n-1} \prod_{k=1}^{j} f_k \right). \quad (10)$$

Hence, the total claims can be expressed in terms of $C_{i,1}$ and the development factors f_k as shown in Equation (8).

1.1.2. Bornhuetter-Ferguson Method

The Bornhuetter-Ferguson (BF) Method combines the Chain-Ladder method with prior assumptions about the ultimate claims.

Let $\hat{C}_i$ be the estimated ultimate claims for the i-th accident year. The BF method incorporates an a priori estimate of ultimate claims μ_i and adjusts it with the observed development. The reserve R_i is estimated as:

$$R_i = \mu_i - C_{i,n} \quad (11)$$

where μ_i is usually obtained from external benchmarks or prior models.

Algorithm 2 Bornhuetter-Ferguson Reserving Method

Input: Cumulative claims $\boldsymbol{C_{ij}}$ where $\boldsymbol{i}$ indexes the accident year and $\boldsymbol{j}$ indexes the development year; Prior estimates $\boldsymbol{\mu_i}$ representing the expected ultimate claims for each accident year.

Step 1: Determine the development factor $\boldsymbol{f_{ij}}$ for each development year $\boldsymbol{j}$ based on historical data. These factors are used to project future claims.

Step 2: Calculate the *ultimate claims* $\boldsymbol{\hat{C}_i}$ for each accident year $\boldsymbol{i}$ using the following equation:

$$\hat{C}_i = \sum_j C_{ij} \cdot f_{i,j} \quad (12)$$

where $\boldsymbol{f_{i,j}}$ is the development factor for the $\boldsymbol{j}$-th development year applied to the $\boldsymbol{i}$-th accident year.

Step 3: Apply the Bornhuetter-Ferguson reserve estimate for each accident year $\boldsymbol{i}$ using:

$$R_i = \frac{\mu_i - \sum_j C_{ij}}{\pi_i} \quad (13)$$

where:

- $\boldsymbol{\mu_i}$ is the prior estimate of the ultimate claims for accident year $\boldsymbol{i}$,
- $\sum_j \boldsymbol{C_{ij}}$ is the cumulative claims reported to date for accident year $\boldsymbol{i}$,
- $\boldsymbol{\pi_i}$ is the proportion of claims reported to date (based on development factors or other methods).

Output: Estimated reserves $\boldsymbol{R_i}$ for each accident year $\boldsymbol{i}$.

1.1.3. Mack Model

The Mack Model is a widely used statistical method for estimating reserves in insurance claim reserving. It is based on the development of claims over time and provides both point estimates and measures of uncertainty. The model assumes a specific structure for the development of claims and leverages this structure to make predictions about future claims. Let C_{ij} represent the cumulative claims reported by development year j for accident year i. The Mack Model assumes that these cumulative claims evolve according to development factors f_j, which are estimated from historical data.

For each accident year i, the cumulative claims C_{ij} are related to the claims of previous years by a development factor f_j. The relationship can be expressed as:

$$C_{i,j+1} = C_{ij} \cdot f_j \quad (14)$$

where f_j is the development factor for the j-th development year.

The ultimate claims $\hat{C}_i$ for accident year i are estimated by projecting the cumulative claims to their final value using the development factors:

$$\hat{C}_i = C_{i1} \cdot f_1 \cdot f_2 \cdots f_{m-1} \quad (15)$$

where m is the total number of development years available.

The reserve R_i for accident year i is calculated as the difference between the ultimate claims and the cumulative claims reported to date:

$$R_i = \hat{C}_i - \sum_j C_{ij} \quad (16)$$

The variance of the reserves $\mathrm{Var}(R_i)$ can be computed by accounting for the variance in the cumulative claims:

$$\mathrm{Var}(R_i) = \sum_j \left(\frac{\partial R_i}{\partial C_{ij}} \right)^2 \mathrm{Var}(C_{ij}) \quad (17)$$

where $\mathrm{Var}(C_{ij})$ is the variance of the cumulative claims C_{ij}, and $\frac{\partial R_i}{\partial C_{ij}}$ denotes the sensitivity of the reserve R_i with respect to C_{ij}.

The development factors f_j are estimated by solving the following least-squares problem:

$$\min_{f_j} \sum_{i,j} \left(C_{i,j+1} - C_{ij} \cdot f_j\right)^2 \tag{18}$$

Proposition 1: Under the assumptions of the Mack Model, the reserve estimator R_i is an unbiased estimator of the ultimate claim amount minus the reported claims.

Proof:

To establish the unbiasedness of the reserve estimator R_i, we leverage the linearity of expectation and the properties of unbiased estimators.

Let $\hat{C}_i$ denote the estimator of the ultimate claim amount for the ith cohort. According to the Mack Model, the reserve R_i is defined as:

$$R_i = \hat{C}_i - \sum_j C_{ij} \tag{19}$$

where $\sum_j C_{ij}$ represents the total reported claims up to development period j. Our goal is to demonstrate that R_i is an unbiased estimator of the ultimate claim amount minus the reported claims.

Linearity of Expectation: By the linearity of expectation, we have:

$$\mathbb{E}[R_i] = \mathbb{E}\left[\hat{C}_i - \sum_j C_{ij}\right] \tag{20}$$

This can be decomposed as:

$$\mathbb{E}[R_i] = \mathbb{E}\left[\hat{C}_i\right] - \mathbb{E}\left[\sum_j C_{ij}\right] \tag{21}$$

Unbiased Estimation of Ultimate Claims: Under the Mack Model assumptions, the estimator $\hat{C}_i$ is an unbiased estimator of the ultimate claim amount C_i^U for the ith cohort. Therefore:

$$\mathbb{E}\left[\hat{C}_i\right] = C_i^U \tag{22}$$

Expected Value of Reported Claims: The expected value of the total reported claims up to development period j is given by:

$$\mathbb{E}\left[\sum_j C_{ij}\right] = \sum_j \mathbb{E}\left[C_{ij}\right] \tag{23}$$

Given that C_{ij} represents the claims reported by the jth development period, and $\mathbb{E}\left[C_{ij}\right]$ is precisely the cumulative claims reported up to period j, the expectation aligns with the reported claims up to that period.

Combining Results: Substituting these results into our earlier expression:

$$\mathbb{E}[R_i] = \mathbb{E}\left[\hat{C}_i\right] - \sum_j \mathbb{E}\left[C_{ij}\right] \tag{24}$$

Since $\mathbb{E}\left[\hat{C}_i\right] = C_i^U$ and $\mathbb{E}\left[\sum_j C_{ij}\right] = \sum_j C_{ij}$, we obtain:

$$\mathbb{E}[R_i] = C_i^U - \sum_j C_{ij} \tag{25}$$

Conclusion: R_i is an unbiased estimator of the ultimate claim amount minus the reported claims, completing the proof.

Lemma 1: The variance of the reserve estimator R_i is derived from the variances of the reported claims and the development factors.

Proof: By applying the propagation of variance formula, we obtain:

$$\mathrm{Var}(R_i) = \sum_j \left(\frac{\partial R_i}{\partial C_{ij}}\right)^2 \mathrm{Var}(C_{ij}) \tag{26}$$

The Mack Model provides a robust framework for estimating reserves and quantifying uncertainty in insurance claim reserving. Through the use of development factors and variance calculations, it offers a systematic approach to predicting future claims and managing risk.

Algorithm 3 Mack Model for Claim Reserving

Input: Cumulative claims $\boldsymbol{C_{ij}}$, where $\boldsymbol{i}$ denotes the accident year and $\boldsymbol{j}$ denotes the development year.

Step 1: Estimate the development factors $\boldsymbol{f_j}$ and the associated variances $\mathrm{Var}(\boldsymbol{f_j})$ for each development year $\boldsymbol{j}$

using the method of least squares or other suitable statistical techniques.

Step 2: Compute the *ultimate claims* $\widehat{\boldsymbol{C}}_{\boldsymbol{i}}$ for each accident year $\boldsymbol{i}$ using the development factors:

$$\hat{C}_i = C_{i1} \cdot f_1 \cdot f_2 \cdots f_{m-1} \tag{27}$$

where $\boldsymbol{m}$ is the number of development years available for the accident year $\boldsymbol{i}$, and $\boldsymbol{f_j}$ represents the development factor for the $\boldsymbol{j}$-th development year.

Step 3: Calculate the reserves $\boldsymbol{R_i}$ for each accident year $\boldsymbol{i}$ as:

$$R_i = \hat{C}_i - \sum_j C_{ij} \tag{28}$$

where $\sum_j \boldsymbol{C_{ij}}$ is the cumulative claims reported to date for accident year $\boldsymbol{i}$.

Step 4: Estimate the variance of the reserves $\mathrm{Var}(\boldsymbol{R_i})$ for each accident year $\boldsymbol{i}$ using the following formula:

$$\mathrm{Var}(R_i) = \sum_j \left(\frac{\partial R_i}{\partial C_{ij}}\right)^2 \mathrm{Var}(C_{ij}) \tag{29}$$

where $\mathrm{Var}(\boldsymbol{C_{ij}})$ represents the variance of the cumulative claims $\boldsymbol{C_{ij}}$ and $\frac{\boldsymbol{\partial R_i}}{\boldsymbol{\partial C_{ij}}}$ is the partial derivative of $\boldsymbol{R_i}$ with respect to $\boldsymbol{C_{ij}}$.

Output: Estimated reserves $\boldsymbol{R_i}$ and their associated variances $\mathrm{Var}(\boldsymbol{R_i})$ for each accident year $\boldsymbol{i}$.

1.1.4. Generalized Linear Models (GLM)

Loss reserving is a crucial task in actuarial science, used to estimate the future claims payments that an insurer expects to make. Generalized Linear Models (GLMs) provide a flexible framework for modeling these claims, offering both statistical and practical advantages.

A Generalized Linear Model (GLM) extends traditional linear models to handle response variables that follow distributions other than the normal distribution. Formally, a GLM consists of three components:

1. **Random Component:** The response variable Y_i is assumed to follow a probability distribution from the exponential family. This includes distributions such as normal, binomial, Poisson, and gamma.
2. **Systematic Component:** The predictors (or explanatory variables) are combined linearly to form a linear predictor. If $\mathbf{x}_i$ denotes the vector of predictors for observation i, then the linear predictor is given by:
$$\eta_i = \mathrm{x}_i^{\mathsf{T}}\beta,$$
where β is a vector of coefficients to be estimated.
3. **Link Function:** The link function $g(\cdot)$ connects the mean of the response variable μ_i to the linear predictor. Specifically,
$$g(\mu_i) = \eta_i.$$

In the context of loss reserving, the response variable Y_i typically represents the number of incurred losses or claims. The systematic component might include factors such as development years, policyholder characteristics, and exposure measures.

The GLM for loss reserving is often specified using the following steps:

Consider a GLM where the response variable Y_i follows a Gamma distribution, which is common for modeling loss amounts due to its flexibility in handling skewed distributions. The probability density function of the Gamma distribution is:

$$f(y_i; \alpha, \beta) = \frac{y_i^{\alpha-1} e^{-y_i/\beta}}{\beta^{\alpha}\Gamma(\alpha)}$$

where α and β are shape and scale parameters, respectively.

The link function for the Gamma distribution is often the reciprocal link:

$$g(\mu_i) = \frac{1}{\mu_i}$$

where μ_i is the mean of the response variable. Thus,

$$\eta_i = \frac{1}{\mu_i}$$

The systematic component is given by:

$$\eta_i = \mathrm{x}_i^{\mathsf{T}}\beta.$$

The parameters β are estimated by maximizing the log-likelihood function. For the Gamma distribution, the log-likelihood function is:

$$\mathcal{L}(\beta) = \sum_{i=1}^{n} \left[\alpha \log\beta - \log\Gamma(\alpha) + (\alpha - 1)\log y_i - \frac{y_i}{\beta}\right].$$

For a GLM with a Gamma distribution, the maximum likelihood estimator (MLE) of the parameter vector β is consistent and asymptotically normal.

Proof. The consistency of the MLE follows from the fact that the log-likelihood function is concave and the Fisher information matrix is positive definite. The asymptotic normality can be derived using the standard results from the theory of M-estimators. □

The variance of the parameter estimates $\hat{\beta}$ is given by:

$$\mathrm{Var}(\hat{\beta}) = [\mathrm{X}^{\mathrm{T}}\mathrm{W}\mathrm{X}]^{-1},$$

where W is a diagonal matrix with elements $w_i = \frac{\partial \mu_i}{\partial \eta_i} \cdot \mathrm{Var}(Y_i)$.

The estimated reserves can be calculated as the expected value of the response variable given the predictors, i.e.,

$$\hat{R}_i = \hat{\mu}_i,$$

where $\hat{\mu}_i$ is the fitted value from the GLM.

Assuming the relationship between the claim amount Y_i and the covariates $\mathbf{x}_i$ is correctly specified within the Generalized Linear Model (GLM) framework, the GLM provides asymptotically efficient estimates for the reserve parameters.

Proof. Consider a Generalized Linear Model (GLM) where the response variable Y_i follows an exponential family distribution with probability density function (pdf) given by:

$$f(y_i; \theta_i, \phi) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right),$$

where θ_i denotes the natural parameter, ϕ is the dispersion parameter, $b(\cdot)$ is the cumulant function, and $c(\cdot)$ is a function of the dispersion parameter ϕ.

Suppose the relationship between θ_i and the covariates $\mathbf{x}_i$ is specified through a link function $g(\cdot)$ such that:

$$g(\mu_i) = \theta_i,$$

where $\mu_i = \mathbb{E}[Y_i]$ represents the mean of the response variable.

The log-likelihood function for the parameter vector β in the GLM is:

$$\mathcal{L}(\beta; \mathrm{y}, \mathrm{X}) = \sum_{i=1}^{n} \left(\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right).$$

Algorithm 4 Generalized Linear Models (GLM) Method for Loss Reserving

Input: Claims data $\mathbf{Y} = \{\boldsymbol{Y_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{n}}$, covariates $\mathbf{X} = \{\boldsymbol{X_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{n}}$

Output: Estimated reserves $\widehat{\boldsymbol{R}}$

Fit the Generalized Linear Model (GLM) to the data

Step 1: Specify the GLM with a link function $\boldsymbol{g}(\cdot)$ and a probability distribution from the exponential family. Let $\boldsymbol{\mu_i}$ denote the predicted mean of the $\boldsymbol{i}$-th claim, where

$$g(\mu_i) = \mathrm{X}_i\beta \quad (30)$$

Here, $\mathbf{X_i}$ is the vector of covariates for the $\boldsymbol{i}$-th claim, and $\boldsymbol{\beta}$ represents the vector of parameters to be estimated.

Step 2: Estimate the parameters $\boldsymbol{\beta}$ by maximizing the log-likelihood function:

$$\hat{\beta} = \underset{\beta}{\mathrm{argmax}} \left[\sum_{i=1}^{n} \ell\,(Y_i \mid \mathrm{X}_i, \beta)\right] \quad (31)$$

where $\boldsymbol{\ell(Y_i \mid \mathbf{X}_i, \boldsymbol{\beta})}$ denotes the log-likelihood function of the $\boldsymbol{i}$-th claim given the covariates and model parameters.

Step 3: For each claim $\boldsymbol{i}$, predict the expected mean $\widehat{\boldsymbol{\mu}}_{\boldsymbol{i}}$ using the fitted model parameters $\widehat{\boldsymbol{\beta}}$:

$$\hat{\mu}_i = g^{-1}(\mathrm{X}_i\hat{\beta}) \quad (32)$$

Here, $\boldsymbol{g^{-1}}(\cdot)$ is the inverse of the link function.

Step 4: Estimate the reserve $\widehat{\boldsymbol{R}}$ as the sum of the predicted claims:

$$\hat{R} = \sum_{i=1}^{n} \hat{\mu}_i \quad (33)$$

Under regularity conditions, including the correct specification of the link function and the distribution, the maximum likelihood estimates $\hat{\beta}$ possess the following properties:

1. **Consistency**: As the sample size n approaches infinity, $\hat{\beta}$ converges in probability to the true parameter value β^*.
2. **Asymptotic Efficiency**: The covariance matrix of $\hat{\beta}$ can be approximated by:

$$\mathrm{Var}(\hat{\beta}) \approx \left[\mathrm{X}^T \mathrm{V}(\hat{\beta})^{-1} \mathrm{X}\right]^{-1},$$

where $\mathrm{V}(\hat{\beta})$ represents the variance-covariance matrix of the observations, which depends on X and $\hat{\beta}$.

Additionally, the efficient score function is given by:

$$U(\beta) = \frac{\partial \mathcal{L}(\beta; \mathrm{y}, \mathrm{X})}{\partial \beta}.$$

The Fisher Information matrix $\mathrm{I}(\beta)$ is defined as:

$$\mathrm{I}(\beta) = -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\beta; \mathrm{y}, \mathrm{X})}{\partial \beta^2}\right],$$

and the Cramér-Rao lower bound asserts that:

$$\mathrm{Var}(\hat{\beta}) \geq [\mathrm{I}(\beta)]^{-1}.$$

Thus, under the correct model specification with an appropriate link function and distribution, the GLM achieves the Cramér-Rao lower bound, demonstrating its asymptotic efficiency.

Consequently, if the model is correctly specified with the appropriate link function and distribution, Generalized Linear Models provide efficient and consistent estimates of the reserves.

Generalized Linear Models offer a robust and flexible approach to loss reserving. By appropriately specifying the random component, systematic component, and link function, actuaries can effectively model and estimate future claims. The mathematical properties of GLMs ensure that the parameter estimates are reliable and that the model can be used to make informed predictions about future losses.

1.2. Structure and theory of Machine learning towards Actuarial Loss Reserving using the Inflation Adjusted Frequency Severity Approach

Actuarial loss reserving is crucial in insurance for estimating the amount needed to cover future claims. Traditional methods often rely on deterministic models that may not fully capture the complex patterns in data. Machine learning offers a modern approach to enhance the accuracy of these estimates. In particular, the inflation-adjusted frequency-severity approach is employed to account for economic inflation in both the frequency and severity of claims.

1.2.1. Mathematical foundation

The inflation-adjusted frequency-severity approach integrates inflation adjustments into the traditional frequency-severity model. This approach is structured as follows:

3. *Frequency Model:* This model estimates the number of claims per unit of exposure.
4. *Severity Model:* This model estimates the cost per claim.
5. *Inflation Adjustment:* This adjusts both frequency and severity models for economic inflation.

Let N_t be the number of claims at time t, and S_t be the severity of each claim. The inflation-adjusted frequency-severity approach involves the following steps: Define the observed frequency $F(t)$ and severity $\mathrm{S}(t)$ as:

$$F(t) = \frac{N_t}{E_t} \tag{34}$$

$$\mathrm{S}(t) = \frac{\text{Total Cost}}{N_t} \tag{35}$$

where E_t is the exposure at time t, and Total Cost is the sum of all claims' costs.

To adjust for inflation, we use an inflation factor $I(t)$, which reflects the change in price level over time. The inflation-adjusted frequency $F_{\text{adj}}(t)$ and severity $\mathrm{S}_{\text{adj}}(t)$ are:

$$F_{\text{adj}}(t) = \frac{N_t}{E_t \cdot I(t)} \tag{36}$$

$$\mathrm{S}_{\text{adj}}(t) = \frac{\text{Total Cost}}{N_t \cdot I(t)} \tag{37}$$

The total reserve R required is then:

$$R = \sum_{t=1}^{T} F_{\text{adj}}(t) \cdot \mathrm{S}_{\text{adj}}(t) \cdot E_t \tag{38}$$

1.2.2. Machine Learning Integration

Machine learning models, such as regression trees, neural networks, or ensemble methods, can be used to predict $F(t)$ and $\mathrm{S}(t)$ based on historical data and other

covariates. These models account for complex relationships and interactions that traditional methods might miss. Here, we outline a pseudo algorithm for integrating machine learning with the inflation-adjusted frequency-severity approach.

Let R denote the inflation-adjusted total reserve. We claim that R is an unbiased estimator of the future claims reserve R^*, given the assumption of stationary inflation.

Proposition: Under the assumption of accurate inflation adjustments and stationary inflation, the estimator R satisfies:

$$\mathbb{E}[R] = R^*,$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator.

Proof:

Consider the reserve R computed as:

$$R = \sum_{i=1}^{m} \tilde{F}(X_i) \cdot \tilde{S}(X_i),$$

where $\tilde{F}(X_i)$ and $\tilde{S}(X_i)$ are the inflation-adjusted frequency and severity estimates, respectively. The adjustment for inflation is performed by:

$$\tilde{F}(X_i) = \hat{F}(X_i) \cdot \frac{\text{CPI}_{current}}{\text{CPI}_{base}}$$

$$\tilde{S}(X_i) = \hat{S}(X_i) \cdot \frac{\text{CPI}_{current}}{\text{CPI}_{base}}$$

Under the assumption of stationary inflation, the Consumer Price Index (CPI) adjustment factor $\frac{\text{CPI}_{current}}{\text{CPI}_{base}}$ is constant across the dataset. Thus, we can factor it out of the summation:

$$R = \frac{\text{CPI}_{current}}{\text{CPI}_{base}} \sum_{i=1}^{m} \hat{F}(X_i) \cdot \hat{S}(X_i)$$

Let $\mathcal{R}$ be the reserve computed without inflation adjustment:

$$\mathcal{R} = \sum_{i=1}^{m} \hat{F}(X_i) \cdot \hat{S}(X_i)$$

Since $\hat{F}(X_i)$ and $\hat{S}(X_i)$ are unbiased estimators of the true frequencies $f^*(X_i)$ and severities $s^*(X_i)$ respectively, we have:

$$\mathbb{E}\left[\hat{F}(X_i)\right] = f^*(X_i)$$

$$\mathbb{E}\left[\hat{S}(X_i)\right] = s^*(X_i)$$

Thus, the expectation of $\mathcal{R}$ is:

$$\mathbb{E}[\mathcal{R}] = \sum_{i=1}^{m} \mathbb{E}\left[\hat{F}(X_i)\right] \cdot \mathbb{E}\left[\hat{S}(X_i)\right] = \sum_{i=1}^{m} f^*(X_i) \cdot s^*(X_i) = R^*$$

Consequently, the expectation of R is:

$$\mathbb{E}[R] = \frac{\text{CPI}_{current}}{\text{CPI}_{base}} \mathbb{E}[\mathcal{R}] = \frac{\text{CPI}_{current}}{\text{CPI}_{base}} \cdot R^* = R^*$$

Therefore, R is an unbiased estimator of the future claims reserve R^*, as required.

Lemma: Consider a given time period t. The inflation-adjusted frequency $\mathbb{E}\left[F_{\text{adj}}(t)\right]$ and severity $\mathbb{E}\left[S_{\text{adj}}(t)\right]$ can be expressed as follows:

$$\mathbb{E}\left[F_{\text{adj}}(t)\right] = \frac{\mathbb{E}[N_t]}{E_t \cdot I(t)} \tag{39}$$

$$\mathbb{E}\left[S_{\text{adj}}(t)\right] = \frac{\mathbb{E}[\text{Total Cost}]}{N_t \cdot I(t)} \tag{40}$$

Proof:

Let N_t denote the number of claims in time period t, and Total Cost represent the aggregate cost of all claims in time period t. Let E_t denote the exposure or relevant metric for normalization, and $I(t)$ denote the inflation adjustment factor at time t.

1.Inflation-Adjusted Frequency:

The inflation-adjusted frequency $\mathbb{E}\left[F_{\text{adj}}(t)\right]$ is given by:

$$\mathbb{E}\left[F_{\text{adj}}(t)\right] = \frac{\text{Number of Claims}}{\text{Exposure} \times \text{Inflation Factor}}$$

By definition, the number of claims N_t is the raw frequency, and it needs to be adjusted for exposure E_t and inflation $I(t)$. Therefore:

$$\mathbb{E}\left[F_{\text{adj}}(t)\right] = \frac{N_t}{E_t \cdot I(t)}$$

This relationship aligns with the expectation of the adjusted frequency when accounting for exposure and inflation.

2. Inflation-Adjusted Severity:

The inflation-adjusted severity $\mathbb{E}\left[S_{\text{adj}}(t)\right]$ is given by:

$$\mathbb{E}\left[S_{\text{adj}}(t)\right] = \frac{\text{Total Cost}}{\text{Number of Claims} \times \text{Inflation Factor}}$$

The total cost represents the aggregate severity before adjustment. To obtain the adjusted severity, we normalize by the number of claims N_t and adjust for inflation $I(t)$. Thus:

$$\mathbb{E}\left[S_{\text{adj}}(t)\right] = \frac{\mathbb{E}[\text{Total Cost}]}{N_t \cdot I(t)}$$

Since both expressions are derived directly from the definitions of frequency and severity adjustments, accounting for inflation and exposure, they follow directly from the properties of expectation and the inflation adjustment mechanism.

Algorithm 5 Inflation-Adjusted Frequency-Severity Reserve Estimation

Procedure ESTIMATE RESERVE (Data, InflationData)

Initialize a machine learning model $\boldsymbol{\mathcal{M}}$

Split Data $= \{(\boldsymbol{X_i}, \boldsymbol{Y_i})\}_{\boldsymbol{i=1}}^{\boldsymbol{n}}$ into training set $\boldsymbol{\mathcal{D}_{train}}$ and testing set $\boldsymbol{\mathcal{D}_{test}}$

Train the frequency model $\hat{\boldsymbol{f}}(\boldsymbol{X}) \approx \boldsymbol{f}^*(\boldsymbol{X})$ on $\boldsymbol{\mathcal{D}_{train}}$ Train the severity model $\hat{\boldsymbol{s}}(\boldsymbol{X}) \approx \boldsymbol{s}^*(\boldsymbol{X})$ on $\boldsymbol{\mathcal{D}_{train}}$

Predict frequencies $\widehat{\boldsymbol{F}}(\boldsymbol{X}) = \{\hat{\boldsymbol{f}}(\boldsymbol{X_i})\}_{\boldsymbol{i=1}}^{\boldsymbol{m}}$ and severities $\widehat{\boldsymbol{S}}(\boldsymbol{X}) = \{\hat{\boldsymbol{s}}(\boldsymbol{X_i})\}_{\boldsymbol{i=1}}^{\boldsymbol{m}}$ using $\boldsymbol{\mathcal{D}_{test}}$

Adjust predictions for inflation using InflationData:

$$\widetilde{\boldsymbol{F}}(\boldsymbol{X_i}) = \widehat{\boldsymbol{F}}(\boldsymbol{X_i}) \cdot \frac{\text{CPI}_{\boldsymbol{current}}}{\text{CPI}_{\boldsymbol{base}}}$$

$$\tilde{\boldsymbol{S}}(\boldsymbol{X_i}) = \widehat{\boldsymbol{S}}(\boldsymbol{X_i}) \cdot \frac{\text{CPI}_{\boldsymbol{current}}}{\text{CPI}_{\boldsymbol{base}}}$$

Compute the reserve $\boldsymbol{R}$ using adjusted frequencies and severities:

$$\boldsymbol{R} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \widetilde{\boldsymbol{F}}(\boldsymbol{X_i}) \cdot \tilde{\boldsymbol{S}}(\boldsymbol{X_i})$$

Return Reserve $\boldsymbol{R}$

end procedure=0

Machine learning models integrated with inflation-adjusted frequency-severity approaches offer a powerful tool for actuarial loss reserving. By leveraging advanced algorithms, we can better account for complex patterns in claims data and provide more accurate reserve estimates.

1.3. The Novelty of IFRS17 in Non-Life Insurance related to Actuarial Loss Reserving

IFRS 17, officially known as International Financial Reporting Standard 17, is a global accounting standard developed by the International Accounting Standards Board (IASB) to regulate the recognition, measurement, presentation, and disclosure of insurance contracts. This standard, which replaced IFRS 4, was implemented on January 1, 2023, with the aim of providing a more consistent and transparent view of the financial position and performance of insurance companies.

IFRS 17 introduces significant changes to the way insurers, including non-life insurance companies, recognize and report their financial performance, specifically in terms of actuarial loss reserving. The novelty of IFRS 17 lies in its requirement for a more granular, transparent, and risk-sensitive approach to reserving compared to earlier standards. Under IFRS 17, insurers must measure insurance contracts at a more detailed level, typically on a contract-by-contract basis or in groups of contracts with similar characteristics [6].

IFRS 17 introduces the concept of discounting cash flows in the calculation of insurance liabilities. This means that reserves now reflect the time value of money, ensuring a more accurate estimate of future reserves as illustrated in this paper. Additionally, IFRS 17 requires a risk adjustment for non-financial risks, which adjusts reserves to reflect the uncertainty around insurance obligations [7] which is also presented in this paper. The introduction of the Contract Service Margin (CSM) represents a key innovation in IFRS 17. It defers the recognition of profits to match the delivery of insurance services, ensuring that insurers only recognize earnings as they provide coverage this is too presented in this study [8].

IFRS 17 aligns actuarial reserving more closely with economic reality and regulatory frameworks such as Solvency II. Both frameworks require the use of market-consistent assumptions and forward-looking estimates. For actuarial professionals, this alignment leads to enhanced consistency between financial reporting and risk management practices, demanding more advanced actuarial techniques for reserve estimation [9]. The new standard enhances transparency by requiring more detailed disclosures about the assumptions and methods

used in reserving calculations. This enables stakeholders, including regulators and investors, to better assess the financial health of insurers. As a result, actuarial loss reserving must now accommodate more comprehensive reporting requirements, making the process more robust and transparent.

In closing, IFRS 17 brings a sophisticated and risk-sensitive approach to actuarial loss reserving in non-life insurance. Its focus on discounted cash flows, risk adjustments, and transparency represents a significant shift from previous standards, ensuring that loss reserves are both more reflective of underlying risks and aligned with modern regulatory and market expectations, and this has been presented in this paper too.

1.4. Novelty of the study

This study uniquely combines classical actuarial techniques with advanced AI algorithms, such as neural networks and random forests, to enhance the accuracy and efficiency of loss reserving. By integrating these methodologies, the study provides a pioneering approach that bridges traditional actuarial practices with modern data science innovations. The creation of automated models for frequency, severity, and inflation, which are key components of actuarial loss reserving, represents a novel advancement. These models utilize machine learning techniques to predict and aggregate reserve estimates, offering a dynamic and automated approach to actuarial forecasting. The study introduces the concept of RAALRM, a new metric designed to provide a robust measure of actuarial loss reserves. This approach integrates upper and lower reserve estimates to offer a more comprehensive and reliable evaluation of reserve adequacy, representing a significant innovation in reserve margin calculations. The methodology proposes a novel framework for distributing reserves across different policyholder categories. By customizing reserve allocations based on policyholder types and their specific risk profiles, the study enhances the precision and fairness of reserve distribution. The study introduces a detailed bonus rate system for policyholder categories, reflecting variations in claims experience and risk. This system provides a refined approach to adjusting reserves based on actual claims data, contributing to more accurate reserve estimates.

The study pioneers the integration of advanced AI techniques, specifically Random Forest models, into actuarial loss reserving. By comparing these AI-driven models with traditional actuarial methods, the study not only highlights the superior predictive performance of AI but also sets a new standard for incorporating machine learning into actuarial practice. The development of a comprehensive framework that separately models frequency, severity, and inflation represents a significant innovation. Each component model is tailored to predict specific aspects of loss reserves, allowing for more granular and accurate forecasting compared to traditional methods that often use aggregated or less detailed approaches. The study's use of robustness and stress testing, including perturbations and scenario analysis, represents a cutting-edge approach to validating the stability and resilience of actuarial models. This rigorous testing ensures that the models are robust under varying conditions and provides a deeper understanding of their performance in uncertain environments.

Explicitly incorporating IFRS 17 requirements into the AI-driven models, including the calculation of Contractual Service Margin (CSM) and Present Value of Future Cash Flows (PVFCF), ensures that the proposed models are aligned with contemporary regulatory standards. This integration addresses a critical need for compliance and sets a precedent for future research and practice in actuarial science.

1.5. Contribution to Actuarial Science

By integrating AI-driven models with traditional actuarial methods, this study contributes to the advancement of actuarial science through improved accuracy and efficiency in loss reserving. The use of machine learning algorithms allows for more precise predictions and optimizes the reserve estimation process. The development of automated actuarial models and the RAALRM metric represents a significant methodological contribution to the field. These innovations offer new tools for actuaries to better manage and assess loss reserves, enhancing their ability to handle complex and dynamic insurance data. The study's focus on policyholder-centric reserve allocation provides valuable insights into how reserves can be more accurately distributed based on policyholder risk profiles. This approach promotes more equitable and targeted reserve

management, aligning with the diverse nature of insurance portfolios. The introduction of a comprehensive bonus rate system and a robust margin calculation framework contributes to the evolution of reserve management practices. These advancements help actuaries better account for variations in claims experience and policyholder behavior, leading to more informed decision-making. The study demonstrates the practical application of data science techniques within the realm of actuarial science, fostering interdisciplinary advancements and setting a precedent for future research. This bridging of fields enhances the overall capability of actuarial science to adapt to modern technological developments.

The application of robust testing methodologies, including scenario analysis and stress testing, advances the field by establishing best practices for validating actuarial models. These techniques ensure that models are reliable and adaptable to various market conditions and regulatory environments. The study's focus on aligning AI-driven models with IFRS 17 regulations contributes to the field by ensuring that actuarial practices meet contemporary accounting standards. This alignment supports the transition to modern regulatory frameworks and enhances the credibility of actuarial models in a regulated environment.

In short, this study's novelty lies in its innovative integration of AI and machine learning with actuarial methods, its introduction of new metrics and frameworks, and its contributions to more precise and equitable reserve management. These advancements represent a significant step forward in the field of Actuarial Science, offering both theoretical and practical benefits. In addition to that, this study's novelty lies in its integration of AI with traditional actuarial methods, its development of a comprehensive and detailed modeling framework, and its rigorous approach to model validation. Its contributions to actuarial science are marked by improved predictive accuracy, enhanced regulatory compliance, and valuable insights into policyholder-specific reserve allocation.

2. Survey of Methods and Literature Review

Automated actuarial loss reserving in non-life insurance has significantly advanced with the advent of actuarial data science and artificial intelligence (AI). Traditional loss reserving methods, which include chain-ladder and Bornhuetter-Ferguson methods, are being complemented and in some cases replaced by sophisticated AI-driven techniques. This review surveys these methods, focusing on the integration of data science and AI in actuarial practices.

Chain-Ladder Method The chain-ladder method, introduced by [10], is one of the oldest and most commonly used techniques for loss reserving. It relies on the assumption that future claims development patterns will follow historical trends. This method calculates reserves by applying development factors derived from historical data to current claims [10]. The Bornhuetter-Ferguson (BF) method, as outlined by [11], combines the chain-ladder approach with a priori estimates of ultimate claims. This method is particularly useful when dealing with new or emerging lines of business where historical data is limited [11]. Generalized Linear Models (GLMs) GLMs, as discussed by [10], have been widely adopted for actuarial modeling due to their flexibility and ability to handle various types of data distributions. They have been utilized in loss reserving to model the relationship between claim amounts and explanatory variables [10].

Machine Learning Techniques Recent advancements in machine learning have introduced new methodologies for loss reserving. Random forests, gradient boosting machines, and neural networks have shown promising results in improving reserve predictions. For instance, the paper [11] explored the use of machine learning methods for actuarial applications and highlighted their potential advantages over traditional methods.

Deep Learning Models Deep learning, particularly neural networks, has gained traction in actuarial science due to its ability to capture complex patterns in data. demonstrated that deep learning models could outperform traditional models in loss reserving tasks by learning intricate relationships from large datasets [12].

Ensemble methods, such as stacking and bagging, combine multiple models to improve predictive performance. The paper [13] provided an overview of ensemble techniques, illustrating how they can enhance the accuracy and robustness of loss reserving predictions [13]. Hybrid Models Hybrid approaches that combine traditional actuarial methods with AI techniques are emerging. These models leverage the strengths of both approaches, providing more robust and accurate loss

reserves. For example, [14] proposed a hybrid model integrating GLMs with machine learning algorithms to improve loss reserving accuracy. Automated Reserving Systems Automated systems incorporating AI have been developed to streamline the reserving process. These systems utilize AI to automate data preprocessing, model selection, and reserve estimation, thus reducing manual effort and increasing efficiency. The paper [15] reviewed such systems and discussed their implications for the actuarial profession.

The integration of actuarial data science and AI into loss reserving practices represents a significant advancement in non-life insurance. Traditional methods continue to be valuable, but AI-driven approaches offer enhanced accuracy and efficiency. The evolving landscape of actuarial data science promises further innovations and improvements in automated reserving solutions.

3. Methodology

The research methodology for investigating actuarial data science-based AI solutions for automated loss reserving in non-life insurance presents a structured approach to gather, analyze, and interpret data and this methodology integrates both traditional actuarial techniques and modern data science methods to evaluate the effectiveness and efficiency of AI-driven solutions in loss reserving.

3.1. Research Design

The research design for this study is a mixed-methods approach, combining quantitative and qualitative analyses to provide a comprehensive evaluation of AI solutions in loss reserving. This design allows for the integration of numerical data and theoretical insights, providing a robust framework for assessing the impact of AI technologies on actuarial practices. Quantitative Research Quantitative research involves the use of statistical methods to analyze numerical data. In this study, quantitative methods are used to evaluate the performance of AI-driven loss reserving models compared to traditional methods. Key aspects include model comparison, data preparation, and analytical techniques.

3.1.1. Model comparison

Employing statistical tests and metrics to compare the accuracy, efficiency, and reliability of AI models (e.g., neural networks, random forests) [16]. Utilizing the simulated data set in this study to train and test AI models. Performance metrics such as mean squared error (MSE), mean absolute error (MAE), and reserve accuracy are calculated to evaluate model effectiveness [17].

On a separate note, conducting semi-structured interviews with actuaries, data scientists, and insurance professionals helps to gather insights on the challenges and benefits of AI in loss reserving [18], [19], [20], [21] and also analyzing case studies from insurance companies that have implemented AI-driven reserving solutions to understand real-world applications and outcomes [22].

3.1.2. Data preparation

Data preparation involved cleaning and preprocessing data to ensure quality and consistency. This includes handling missing values, normalizing data, and splitting datasets into training and testing subsets for model evaluation.

3.1.3. Analytical techniques

Statistical techniques were employed to analyze the performance of traditional and AI-based models [23]. Various machine learning algorithms are applied to develop predictive models for loss reserving. Algorithms such as regression models, decision trees, and neural networks are used to predict future claims based on historical data [24]. Evaluating model performance using cross-validation techniques to ensure robustness and generalizability of the results [25].

The research methodology outlined provides a comprehensive framework for evaluating actuarial data science-based AI solutions in automated loss reserving. By integrating quantitative and qualitative approaches, the study aims to provide a nuanced understanding of how AI technologies impact actuarial practices and their effectiveness in improving loss reserving accuracy.

3.2. The proposed approach

This section describes the general Machine learning-based Automated Actuarial Loss Reserving Models using the machine learning methods presented in Table A1.

3.2.1. Policyholder Actuarial Loss Reserving Categories

To begin, the following four main policyholder categories are proposed, as presented in Table 2.

Table 2. Automated Actuarial Loss Reserving Risk Pricing Policyholder Categories.

Automated Actuarial Loss Reserving Policyholder Categories	
Category A	Policyholder with both Car Insurance and Microfinance policies
Category B	Policyholder with Microfinance policy only
Category C	Policyholder with Car Insurance policy only
Category D	Policyholder with no policy

3.2.2. Development of the AI-Based Automated Actuarial Loss Reserving Models

This subsection delineates the methodologies employed for estimating, forecasting, and validating the actuarial models. The process encompasses the following steps:

1. *Automated Actuarial Frequency Models*: For this model, the dependent variable Y_{freq} represents the Comprehensive Number of Claims, denoted mathematically as $Y_{\text{freq}} = f_{\text{freq}}(\text{X})$, where X denotes the vector of covariates. The model is specified as:

$$Y_{\text{freq}} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon_{\text{freq}} \quad (41)$$

where β_i are the regression coefficients and ϵ_{freq} represents the error term.

2. *Automated Actuarial Severity Models*: This model focuses on the Comprehensive Claim Amount, Y_{sev}. The dependent variable is modeled as $Y_{\text{sev}} = f_{\text{sev}}(\mathbf{X})$, and the model is expressed as:

$$Y_{\text{sev}} = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \cdots + \gamma_q X_q + \epsilon_{\text{sev}}, \quad (42)$$

where γ_i are the severity coefficients and ϵ_{sev} is the error term.

3. *Automated Actuarial Loss Reserve Inflation Models*: For modeling inflation, the dependent variable I_{inf} is the Inflation Index derived from the Consumer Price Index (CPI). The model is given by:

$$I_{\text{inf}} = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \cdots + \delta_r X_r + \epsilon_{\text{inf}}, \quad (43)$$

where δ_i are the inflation coefficients and ϵ_{inf} represents the error term.

The integration of the three aforementioned models is achieved through a multiplicative aggregation of their predictions. Specifically, if $\hat{Y}_{\text{freq}}$, $\hat{Y}_{\text{sev}}$, and $\hat{I}_{\text{inf}}$ denote the predicted values from the frequency, severity, and inflation models respectively, then the combined forecast $\hat{Y}_{\text{combined}}$ is computed as:

$$\hat{Y}_{\text{AALR}} = \hat{Y}_{\text{freq}} \times \hat{Y}_{\text{sev}} \times \hat{I}_{\text{inf}}. \quad (44)$$

where $AALR$ is the Automated Actuarial Loss Reserve and thus this composite prediction leverages the individual model forecasts to estimate the Automated Actuarial Loss Reserves, thereby integrating the effects of frequency, severity, and inflation into a unified actuarial forecast.

3.3. Setting Up the Final Automated Actuarial Loss Reserving Models

Following the automation of predictions derived from the aforementioned regression models, we proceed to compute the Inflation-Adjusted Frequency Severity Automated Actuarial Loss Reserves. This quantity is redefined and referred to as the Automated Actuarial Loss Reserve Margin (AALRM).

Let Y_{AALR} denote the Comprehensive Claim Amount extracted from the original test data set. The Upper Actuarial Loss Reserve Margin (UAALRM) is calculated by summing the AALRM with Y_{AALR}, expressed as:

$$\text{UAALRM} = Y_{\text{AALR}} + \text{AALRM}. \quad (45)$$

Similarly, the Lower Automated Actuarial Loss Reserve Margin (LAALRM) is derived by subtracting the AALRM from Y_{AALR}:

$$\text{LAALRM} = Y_{\text{AALR}} - \text{AALRM}. \quad (46)$$

The Robust Automated Actuarial Loss Reserve Margin (RAALRM) is then determined as the average of the UAALRM and LAALRM:

$$\text{RAALRM} = \frac{\text{UAALRM} + \text{LAALRM}}{2}. \quad (47)$$

A new data set is constructed containing the three principal variables: RAALRM, LAALRM, and UAALRM. A final regression model is fitted with RAALRM as the dependent variable and LAALRM and UAALRM as independent variables. This regression model is expressed as:

$$\text{RAALRM} = \alpha_0 + \alpha_1 \text{LAALRM} + \alpha_2 \text{UAALRM} + \epsilon_{\text{RAALRM}}, \quad (48)$$

where α_0, α_1, and α_2 are the regression coefficients, and ϵ_{RAALRM} denotes the error term.

The model provides predictions for the Robust Automated Actuarial Loss Reserves (RAALR), which are then aggregated to compute the Total RAALR. Specifically, if $\widehat{\text{RAALR}}_i$ represents the predicted RAALR for observation i, the Total RAALR is given by:

$$\text{Total RAALR} = \sum_{i=1}^{n} \widehat{\text{RAALR}}_i, \quad (49)$$

where n is the number of observations in the data set.

This process results in a comprehensive estimation of the actuarial loss reserves, integrating the effects of frequency, severity, and inflation adjustments into a robust final margin.

3.4. Terminology and Assumptions for Automated Actuarial Loss Reserving Models

The following terminology concerning actuarial loss reserving is defined, including both existing and new types of reserves, as presented in Table 3.

Table 3. Definition of Types of proposed Actuarial Reserves.

Type of Actuarial Loss Reserve	Definitions
IBNYR	Incurred But Not Yet Reported Reserve
RBNYS	Reported But Not Yet Settled Reserve
REOPENED	Reopened Reserve
REINSURANCE	Reinsurance Reserve

- **IBNYR (Incurred But Not Yet Reported)**: Reserve allocated for incurred claims not yet reported or known to the insurer. Applicable to all policyholder categories defined in Table 2.
- **RBNYS (Reported But Not Yet Settled)**: Reserves for reported but not yet settled claims from both microfinance and car insurance services.
- **REOPENED (Reopened Reserve)**: Reserves for claims that were previously closed or partially paid but have been reopened for full settlement.
- **REINSURANCE (Reinsurance Reserve)**: Reserves for catastrophic losses from either microfinance or car insurance services.

3.5. Proposed Framework for Distribution of Automated Actuarial Loss Reserves

The proposed framework for distributing the Automated Actuarial Loss Reserves is shown in Table 4.

Table 4. Proposed Framework for Distribution of Automated Actuarial Loss Reserves.

Type of Reserve	Proposed Automated Actuarial Loss Reserves Distribution
IBNYR	80% of Total RAALR
RBNYS	15% of Total RAALR
REOPENED	4% of Total RAALR
REINSURANCE	1% of Total RAALR

As indicated in Table 4, a large portion of the Total RAALR is allocated to IBNYR reserves (80%) to cover unreported comprehensive claim amounts. The model assumes efficient claim settlement upon flagging, leading to lower proportions for RBNYS reserves (15%), REOPENED reserves (4%), and REINSURANCE reserves (1%).

3.6. Distribution of Reserves Across Policyholder Categories

The assumptions for the distribution of reserve types across policyholder categories are outlined in Table 5. According to Table 5, Category A, which includes policyholders with both microfinance and car insurance policies, receives the highest proportion (50%). Category B follows with 30%, Category C with 20%, and Category D with 0%, as it has no active policyholders.

Table 5. Policyholder Reserve Allocation Categories.

Policyholder Reserve Allocation Categories				
	Category A	Category B	Category C	Category D
IBNYR	50%	30%	20%	0%
RBNYS	50%	30%	20%	0%
REOPENED	50%	30%	20%	0%
REINSURANCE	50%	30%	20%	0%

3.7. Computations of the proposed types of Reserves

Let CAALR_i denote the Comprehensive Automated Actuarial Loss Reserve for policyholder category i. The CAALR for each policyholder category is computed as follows:

$$\text{CAALR}_i = \sum_{j=1}^{n_i} R_{ij}, \tag{50}$$

where R_{ij} represents the reserve associated with the j-th policyholder within category i, and n_i is the total number of policyholders in category i. The comprehensive reserve for each category is thus derived from the summation of individual reserves across all policyholders within that category.

Let ACAALR denote the Aggregate Comprehensive Automated Actuarial Loss Reserve. It is calculated by aggregating the CAALR values across all policyholder categories. Mathematically, this can be expressed as:

$$\text{ACAALR} = \sum_{i=1}^{k} \text{CAALR}_i, \tag{51}$$

where k represents the total number of policyholder categories. The ACAALR is the summation of the CAALR for each individual policyholder category, providing a holistic measure of the total loss reserves required across the entire portfolio.

3.8. Policyholder Category No Claims Bonus Rates

Assumptions regarding bonus rates for each policyholder reserving category are detailed in Table 6.

Table 6. Policyholder Category No Claims Bonus Rate.

Category	Base Bonus Rates	Variable Bonus Rates	Final Bonus Rates
A	1%	4%	5%
B	1%	3%	4%
C	1%	2%	3%
D	0%	0%	0%

As shown in Table 6, policyholders in each category are entitled to a base bonus rate. The variable bonus rate is the difference between the final and base bonus rates and depends on claim amounts. The final bonus rate is the sum of the base and variable bonus rates. Category A has the highest final bonus rate (5%) due to the large proportion of active policyholders, followed by Categories B (4%) and C (3%), with Category D receiving no bonus.

The Evaluation of Policyholder-Based Automated Actuarial Loss Reserves is carried out for two scenarios: the short-term and the long-term periods.

3.9. Best model (ranger) evaluation

3.9.1. Data preparation

The best model evaluation begins with the preparation of a dataset, specifically a simulated general insurance dataset. This dataset includes variables pertinent to non-life insurance claims, such as policy status, policy type, car ownership, policy renewals, exposure, and more. The data is split into training and testing sets, with 80% used for model training and the remaining 20% reserved for testing.

3.9.2. Model Development

Frequency Model: To estimate the number of claims, a Random Forest model is constructed using the ranger package in R. The model is trained on a range of predictor variables including policy-related factors and demographic attributes. Predictions are made on the test set, and model performance is evaluated using Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

Severity Model: A second Random Forest model predicts the number of claims based on similar predictors. Performance metrics (MSE, MAE, and RMSE) are calculated to assess the accuracy of the severity estimates.

Inflation Model: The third Random Forest model predicts the inflation index, a critical component for adjusting claim amounts. This model is also evaluated using MSE, MAE, and RMSE.

3.9.3. Automated Actuarial Loss Reserves Calculation

The automated actuarial loss reserves are calculated by multiplying the predicted number of claims, severity, and inflation index. This product represents the estimated reserves required to cover future claims, adjusted for inflation.

3.9.4. Model comparison between the Ranger model and Simulated Chain Ladder model

To validate the performance of the Random Forest-based loss reserving model, results are compared with traditional actuarial methods, specifically the Chain Ladder method. The Chain Ladder model is simulated using synthetic claim amounts to estimate loss reserves. A comparison is made between the reserves estimated by the Chain Ladder method and those derived from the Random Forest model.

3.9.5. IFRS 17 Compliance of the Ranger Model

The IFRS 17 regulations require insurers to recognize and measure insurance contracts in a manner that reflects their financial position and performance. To ensure compliance with IFRS 17, additional calculations are performed to determine the Contractual Service Margin (CSM) and Present Value of Future Cash Flows (PVFCF). This involves:

- Predicting future cash flows based on the loss reserves.
- Discounting these future cash flows to present value.
- Calculating the CSM as the difference between the loss reserves and the discounted cash flows.

3.9.6. Robustness and Stress Testing

Robustness is tested by introducing perturbations to the predicted values and observing the impact on the automated actuarial loss reserves. This involves adding random noise to the predictions and comparing the results to those obtained from the original model. Summary statistics and density plots are used to assess the stability of the model under perturbations. Stress tests are conducted by varying key parameters such as frequency, severity, and inflation rates. Multiple scenarios are generated to evaluate how changes in these parameters affect the loss reserves. This includes base and stressed scenarios, with results visualized using box plots and summary statistics.

3.9.7. Scenario Analysis

Scenario analysis involves simulating different potential future states to understand the impact of varying assumptions on the automated actuarial loss reserves. This includes adjusting frequency, severity, and inflation rates within defined ranges and assessing the impact on the reserves.

3.10. Novelty in Methodology

The methodology outlined in this study introduces several novel elements that advance the field of actuarial science and automated loss reserving under IFRS 17. The key innovations are:

- *Integration of AI-Driven Models with Traditional Methods:* The study combines advanced machine learning algorithms, specifically Random Forests, with traditional actuarial methods such as the Chain Ladder technique. This hybrid approach enhances the robustness and accuracy of loss reserving models by leveraging the predictive power of AI while maintaining the interpretability and historical context of traditional methods.
- *Comprehensive Model Framework:* The development of a multi-faceted AI-driven actuarial model framework is a significant innovation. The methodology includes separate models for frequency, severity, and inflation, each tailored to predict specific components of loss reserves. This detailed segmentation allows for a

more nuanced and accurate estimation of reserves compared to traditional lumped models.

- *Innovative Aggregation Technique:* The methodology introduces a novel aggregation technique where the predictions from frequency, severity, and inflation models are combined multiplicatively to estimate the Automated Actuarial Loss Reserves (AALR). This approach, which integrates the effects of multiple predictive components into a unified forecast, provides a more comprehensive view of future claims than simple additive methods.
- *Robustness and Stress Testing:* The methodology employs rigorous robustness and stress testing techniques, including the introduction of random noise and the evaluation of multiple stress scenarios. This rigorous testing framework ensures that the AI-driven models are not only accurate under normal conditions but also resilient to variations and uncertainties, thereby enhancing their practical applicability.
- *IFRS 17 Compliance Integration:* A distinctive feature of the methodology is its explicit alignment with IFRS 17 requirements. By calculating the Contractual Service Margin (CSM) and Present Value of Future Cash Flows (PVFCF) based on AI-driven predictions, the study ensures that the loss reserving models comply with the latest accounting standards, thereby bridging the gap between actuarial practice and regulatory requirements.
- *Policyholder-Specific Reserve Allocation:* The methodology's approach to reserve distribution across different policyholder categories, including innovative bonus rates and reserve types, represents a novel application of AI to tailor actuarial reserves. This category-specific focus enhances the precision of reserve allocations and better reflects the risk profiles of different policyholder groups.
- *Scenario Analysis and Prediction Adjustments:* The inclusion of scenario analysis, which involves simulating various future states to assess the impact on reserves, is a unique contribution. This technique allows for a deeper understanding of how different assumptions affect the loss reserves and provides valuable insights for decision-making under uncertainty.

In a nutshell, the novelty of this methodology lies in its integration of cutting-edge AI techniques with traditional actuarial practices, the development of a comprehensive and segmented model framework, rigorous testing procedures, and alignment with contemporary regulatory standards. These innovations collectively advance the field of automated loss reserving and offer a more precise, compliant, and adaptable approach to actuarial practice.

4. Data

Simulated research data refers to information generated through simulation processes, often used to mimic real-world scenarios for analysis and testing purposes. This type of data is not collected from actual experiments or observations but is created using statistical models, algorithms, or other computational methods to replicate conditions or outcomes that researchers are interested in studying [26] and [27].

4.1. The general structure of the simulated non-life insurance data

The Comprehensive General Car Insurance and Microfinance data has been simulated for the period from 1989 to 2022, spanning 33 years. The dataset includes a sample of 40,000 policyholders and is organized into seven primary categories: Policyholder Personal Data, Microfinance Policyholder Data, Policyholder Vehicle Data, Comprehensive Policyholder Claim Data, Comprehensive Policyholder Premium Payment Data, and Policyholder External Data.

In developing the Automated Actuarial Loss Reserving Model, 48 variables derived from these categories were utilized, incorporating eight machine learning algorithms. Of these 48 variables, particular focus was placed on three key principal variables (described in subsection 4.3 below), which have been crucial for automating both car insurance and microfinance services on a single platform.

4.2. Contribution of the Simulated Data to Actuarial AI Solutions

Risk Assessment and Segmentation: Variables such as *Policy Status*, *Policy Type*, and *Claim Score* assist in segmenting policyholders and assessing their risk profiles.

Reserve Calculation: Claim-related variables like *Claim Incurred* and *Case Reserves* are crucial for determining reserves and predicting future liabilities.

Predictive Modeling: Combining historical data (e.g., *Claim History*) with external factors (e.g., *Retained Income*) supports predictive modeling for loss reserving.

Financial Behavior Analysis: Microfinance data (e.g., *Amount Invested*) provides insights into financial behaviors that affect risk and reserve calculations.

Inflation and Cost Adjustments: *Inflation Index* and operational costs (e.g., *Underwriting costs*) are important for adjusting reserves and premiums to account for inflation and costs.

This comprehensive dataset enables the development of robust actuarial models for automated loss reserving by integrating diverse policyholder profiles, claim details, and financial factors.

4.3. Principal data Variable Exploratory Analysis for Automated Actuarial Loss Reserving Model

The principal data variables are defined as follows:

Comprehensive Claim Amount () is defined as the sum of the claim incurred from car insurance services and the amount requested from microfinance services. Mathematically, this can be expressed as:

$$\mathrm{CCA} = \mathrm{CI} + \mathrm{AR} \tag{52}$$

where:

- represents the Claim Incurred from car insurance services.
- represents the Amount Requested from microfinance services.

Comprehensive Paid Amount () is calculated as the sum of the claims paid from car insurance services and the microfinance amounts paid by the insurance company. This can be formulated as:

$$\mathrm{CPA} = \mathrm{CP} + \mathrm{MFAP} \tag{53}$$

where:

- denotes the Claims Paid from car insurance services.
- denotes the Microfinance Amount Paid by the insurance company.

Comprehensive Number of Claims () is defined as the sum of the number of claims from car insurance services and the number of requests from microfinance services. This can be expressed as:

$$\mathrm{CNC} = \mathrm{NC} + \mathrm{NR} \tag{54}$$

where:

- NC represents the Number of Claims from car insurance services.
- NR represents the Number of Requests from microfinance services.

Some further exploratory data analysis is shown below.

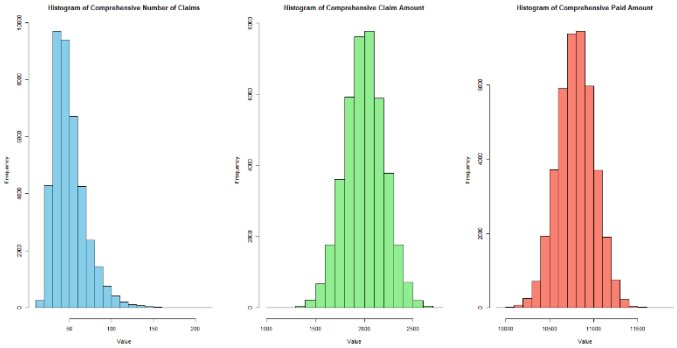

Figure 1. Histograms for key data variables.

Figure 1 shows that the three main variables are close to being normally distributed, as indicated by the bell shape.

4.4. Correlation Analysis for key data variables

Correlation analysis is a statistical technique used to measure the strength and direction of the relationship between two or more variables [27]. It helps in understanding how changes in one variable are associated with changes in another variable. In the context of simulated data variables, correlation analysis can provide

insights into the dependencies and interactions between different variables generated through simulation.

The most common measure of correlation is the correlation coefficient, often denoted by ρ. It ranges from -1 to 1. A ρ of 1 indicates a perfect positive correlation (as one variable increases, the other variable also increases), -1 indicates a perfect negative correlation (as one variable increases, the other variable decreases), and 0 indicates no correlation. In addition to that, Positive Correlation is hen an increase in one variable is associated with an increase in the other variable, Negative Correlation is when an increase in one variable is associated with a decrease in the other variable and finally No Correlation when there is no apparent relationship between the variables [28]. In that regard the results for correlation analysis for key variables is shown below accordingly.

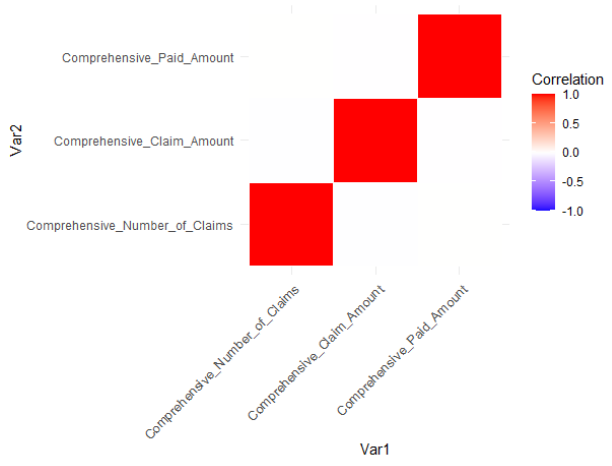

Figure 2. Correlation analysis for key data variables.

Figure 2 shows a heatmap showing the correlation between each pair of variables. Positive correlations are presented towards red, negative correlations towards blue, and no correlation towards white.

The correlation matrix shows the pairwise correlation coefficients between three key variables:

- Comprehensive Number of Claims (CNC)
- Comprehensive Claim Amount (CCA)
- Comprehensive Paid Amount (CPA)

The matrix is as follows:

	CNC	CCA	CPA
CNC	1.00	-0.00	0.00
CCA	-0.00	1.00	-0.00
CPA	0.00	-0.00	1.00

Diagonal Elements (1.00): These indicate that each variable is perfectly correlated with itself, which is expected.

Off-Diagonal Elements:

- The correlation between the Comprehensive Number of Claims and Comprehensive Claim Amount is close to zero, indicating a very weak or negligible linear relationship between these two variables.
- The correlation between the Comprehensive Number of Claims and Comprehensive Paid Amount is also close to zero, suggesting little to no linear relationship.
- The correlation between the Comprehensive Claim Amount and Comprehensive Paid Amount is close to zero, indicating a negligible linear relationship between these two metrics.

The near-zero correlations among the key variables suggest that these variables provide unique, non-overlapping information about the insurance and microfinance processes. This independence is beneficial for the development of a robust model as it allows for the modeling of different aspects of the data without redundancy. Since these variables do not exhibit strong correlations, it implies that including all three variables in the model might enhance its complexity and accuracy by capturing diverse aspects of the data. The model can leverage this unique information to improve predictions and automate loss reserving more effectively. The lack of strong correlations between the key variables may guide the feature selection process, ensuring that the model includes relevant and diverse variables without redundant information. This can lead to better interpretability and performance of the AI-based solution. By incorporating variables with weak correlations, the model is less likely to be overfitted to any single data aspect. This can enhance the generalization of the Automated Actuarial Loss Reserving Model, making it more reliable across different scenarios and datasets. Understanding these relationships helps in interpreting the model's results and the data characteristics. It can assist in identifying areas where the model may need improvement or where additional data might be required to capture underlying patterns.

In short, the correlation matrix supports the development of a more nuanced and effective AI-based solution for automated actuarial loss reserving by

ensuring that the model leverages a diverse set of variables, potentially enhancing its accuracy and robustness in non-life insurance applications.

4.5. Factor Analysis and Principal Component Analysis

Factor Analysis is a statistical technique used to identify underlying relationships between variables. It aims to reduce the number of variables by grouping them into factors that represent underlying dimensions. This method is commonly used in social sciences, psychology, and marketing to identify latent constructs and simplify data structures [28]. Principal Component Analysis (PCA) is a dimensionality reduction technique that transforms a large set of variables into a smaller set of uncorrelated components, which capture the maximum variance in the data. PCA is often used in exploratory data analysis and for predictive modeling to simplify data while retaining its essential characteristics [29].

In the context of factor analysis or principal component analysis (PCA), Table 7 shows the factor loadings which represent the correlations between the variables and the underlying factors or principal components.

Table 7. Factor loadings for key variables.

Metrics	MR2	MR3	MR1
SS loadings	0.003	0.003	0.000
Proportion Variance	0.001	0.001	0.000
Cumulative Variance	0.001	0.002	0.002

Table 8 shows how the PCA relates to standard deviation, proportion of variance, and cumulative proportion.

Table 8. Principal component analysis.

Metrics	PC1	PC2	PC3
Standard deviation	1.00	1.00	1.00
Proportion of Variance	0.34	0.33	0.33
Cumulative Proportion	0.34	0.67	1.00

All principal components (PC1, PC2, PC3) have the same standard deviation of 1.00. This indicates that each component has been scaled to have a standard deviation of 1, which is typical in PCA to standardize the components. The PCA results suggest that the three principal components together capture all the variance in the dataset. This means that if the original data had more variables, the PCA has effectively reduced the dimensionality while retaining the full information content. In actuarial data science, reducing dimensionality can simplify models, reduce computational costs, and help in focusing on the most important features. Since each component captures a significant portion of the variance, the PCA components can be used as new features in the AI models for loss reserving. These components are uncorrelated and collectively explain all the variance, which helps in building models that are robust and less likely to suffer from multicollinearity. By using principal components as inputs, the model can potentially become more efficient. With reduced dimensions and uncorrelated features, the AI models for actuarial loss reserving can process data more quickly and efficiently, which is crucial for real-time or large-scale loss reserving applications. The equal proportion of variance explained by the principal components suggests that each component is contributing meaningfully to the model. This can help in understanding the data structure better and in explaining the model's decisions based on the transformed components, which can be useful for validating and interpreting the results of the loss reserving models.

In the context of automated actuarial loss reserving, the insights gained from PCA can lead to better risk assessment models. By focusing on principal components that encapsulate the majority of the variance, actuaries can develop more accurate models for predicting future losses and making informed decisions.

In short, the PCA results indicate that the dimensionality of the data can be effectively reduced while retaining all the variance. This can positively impact the development of AI solutions for automated loss reserving by simplifying the data, improving model efficiency, and aiding in better risk assessment and decision-making.

4.6. Data pre-processing, data scaling and data partitioning

After loading the data in R caret, R package has been used to generate the one hot encored the simulated General Auto Insurance Microfinance data. From there the data has been pre-processed first by scaling it using min-max approach followed by data partitioning into training data set (80%) and test data set (20%). Data were analyzed using R.

5. Results

This section shows results obtained from the methodology for our proposed Automated Actuarial Loss Reserving Model.

5.1. Machine learning Based Automated Actuarial Loss Reserving Model Methods

Machine learning (ML) methods play a crucial role in the construction of Automated Actuarial Loss Reserving Models. Furthermore, the ML algorithms can capture complex, non-linear relationships in insurance claims data that may be challenging for traditional actuarial methods. This leads to more accurate and precise loss reserve estimates and also handle the Big Data, including historical claims, policyholder information, and external factors. ML can efficiently process and analyze large datasets, making it possible to extract valuable insights from this wealth of information. Subsequently, ML models can be designed to continuously adapt and update based on new data. This enables insurers to have dynamic and up-to-date loss reserves, which is especially valuable in rapidly changing markets. As the name suggests, Automated Actuarial Loss Reserving Models can significantly reduce the need for manual calculations and interventions. This not only saves time but also minimizes the potential for human error. Machine learning models can process and analyze data much faster than traditional manual methods and this is essential in an industry where time is of the essence, particularly for regulatory reporting and financial planning.

In short, ML methods are instrumental in the construction of Automated Actuarial Loss Reserving Models because they offer improved accuracy, efficiency, and flexibility while also enabling real-time updates and the ability to detect patterns and emerging trends. This contributes to better risk management, regulatory compliance, and overall competitive advantages for insurance companies.

5.2. Actuarial Loss Reserving Inflation Adjusted Frequency Severity Models

Actuarial Loss Reserving Inflation Adjusted Frequency Severity (ALR-IAFS) machine learning-based models serve several critical purposes in the insurance industry. These models are designed to estimate future insurance claim amounts while accounting for inflation, claim frequency, and claim severity. ALRIAFS models provide insurance companies with accurate estimates of future claim amounts, which is essential for financial planning and risk management. These models take into account the expected number of claims (frequency) and the expected size of each claim (severity), adjusted for inflation. Moreover, Inflation can erode the value of insurance reserves over time. ALR-IAFS models explicitly adjust for inflation, ensuring that loss reserves remain adequate to cover future claim costs. This is particularly important in long-tail insurance lines where claims may be paid out over several years. By estimating future claim frequencies and severities, insurance companies can better understand and manage their exposure to risk. ALR-IAFS models allow for more informed decisions about capital allocation, underwriting, and pricing to mitigate potential financial risks.

In closing, Actuarial Loss Reserving Inflation Adjusted Frequency Severity (ALR-IAFS) machine learning-based models are critical tools for insurance companies to estimate future claim amounts while considering inflation, claim frequency, and severity. Ultimately, these models support sound financial planning, risk management, regulatory compliance, and strategic decision-making, ultimately contributing to the financial stability and competitiveness of insurance companies.

The Table 9 below shows a combination of frequency models, severity models and finally inflation models as shown below.

Table 9. Actuarial Loss Reserving Inflation Adjusted Frequency Severity Models.

ML Model	Frequency Models		Severity Models		Inflation Models	
	Time (sec)	RMSE	Time (sec)	RMSE	Time (sec)	RMSE
GLM	1.34	53.4943	0.46	2,009.4710	0.65	0.5129
GAM	1.16	53.5111	0.99	2,011.7860	0.84	0.5120
RPART	2.35	53.4166	1.69	2,007.6940	0.79	0.8281
RANGER	55.93	53.2332	269.12	2,011.8630	62.91	0.5124
XGB	6.62	53.4074	6.73	2,013.4440	6.82	0.5119
LAR	12.18	53.6918	15.24	2,012.2690	31.19	0.5124
SVM	289.67	53.3934	264.64	2,012.6160	1095.67	0.5135
ANN	8.56	53.6989	9.11	2,011.8730	6.20	0.5121

The Table 9 compares different machine learning (ML) models across three types of actuarial loss reserving models: Frequency, Severity, and Inflation. The table includes metrics for each model in terms of computation time and Root Mean Square Error (RMSE).

With regards to Frequency Models: Fastest execution time (1.34 seconds) came from GLM with an RMSE of 53.4943 and this shows a good balance of speed and accuracy. GAM is slightly slower than GLM but with similar performance (RMSE of 53.5111). RPART incurred a moderate speed and RMSE (53.4166), slightly better than GLM and GAM. RANGER (is the slowest execution time (55.93 seconds) but slightly better RMSE (53.2332). XGB is fast with a good RMSE (53.4074), making it a competitive choice. LAR is slower than GLM and GAM with an RMSE of 53.6918, indicating less accuracy in this context. SVM has the slowest execution time (289.67 seconds) with RMSE comparable to the best models, suggesting it is computationally intensive with a comparable performance. ANN has moderate speed and an RMSE of 53.6989, performing similarly to other models.

With regards to Severity Models: GLM scooped the fastest execution time (0.46 seconds) with an RMSE of 2,009.4710, showing a trade-off between speed and accuracy. GAM is slightly slower (0.99 seconds) with a similar high RMSE (2,011.7860). RPART has moderate speed with an RMSE of 2,007.6940, showing slightly better performance. RANGER is much slower (269.12 seconds) but with an RMSE of 2,011.8630, similar to GLM and GAM.XGB has moderate execution time (6.73 seconds) with an RMSE of 2,013.4440, slightly worse than GLM and GAM.LAR has slowest execution time (15.24 seconds) with an RMSE of 2,012.2690, comparable to other models. SVM is the slowest in terms of execution time (264.64 seconds) with an RMSE close to other models. ANN has moderate execution time (9.11 seconds) with an RMSE of 2,011.8730, performing similarly to XGB and LAR.

With regards to Inflation Models: GLM has the fastest execution time (0.65 seconds) with an RMSE of 0.5129, indicating good performance. GAM has similar execution time (0.84 seconds) with slightly better RMSE (0.5120). RPART is slightly slower (0.79 seconds) with an RMSE of 0.8281, showing less accuracy. RANGER is slow (62.91 seconds) with an RMSE of 0.5124, which is comparable to GLM and GAM.XGB has fast execution time (6.82 seconds) with an RMSE of 0.5119, similar to GLM and GAM.LAR has moderate execution time (31.19 seconds) with an RMSE of 0.5124.SVM has slowest execution time (1,095.67 seconds) with an RMSE of 0.5135, indicating a trade-off between speed and accuracy. ANN has moderate execution time (6.20 seconds) with an RMSE of 0.5121, similar to XGB and slightly better than SVM.

In closing, the Table 9 helps in choosing the appropriate model based on the trade-offs between computation time and RMSE. For example, GLM and XGB provide a good balance of speed and accuracy across different models. GLM, XGB, and ANN are among the faster models with relatively low RMSE values in the inflation models, making them suitable for applications where computation speed is crucial. Models like RANGER and SVM show varying performance across different types of models. For severe cases where accuracy is paramount, even if slower, these models might be considered based on the need for precision. The table provides insights into which models perform well under different conditions. For instance, GLM and XGB are good choices for models where a balance between speed and accuracy is needed, while RANGER and SVM might be used when model accuracy is more critical and computational resources are available. Understanding these trade-offs helps in selecting the best model for specific applications within actuarial loss reserving, optimizing both model performance and operational efficiency.

5.3. Total Automated Actuarial Inflation Adjusted Frequency Severity Loss Reserves

The Automated Actuarial Inflation Adjusted Loss Reserves (AAIALR) are computed by multiplying the predicted number of claims, the predicted claim amounts, and the predicted inflation values obtained from the machine learning models applied to the test data. The total AAIALR is then determined by summing these individual values, as represented by Equation (55):

$$AAIALR = Freq_{\text{predictions}} \times Sev_{\text{predictions}} \times Infl_{\text{predictions}} \quad (55)$$

From the Table 10, GLM (405.6083) attained the highest score for Total AALR predictions, followed by

XGB (405.3866), followed by ANN (405.2760) and the least came from came from RPART (182.2968).

Table 10. Total Automated Actuarial Loss Reserving Inflation Adjusted AALR.

Actuarial Loss Reserve Models	
ML Model	Total AALR Predictions
GLM	405.6083
GAM	400.6047
RPART	182.2968
RANGER	402.8569
XGB	405.3866
LAR	402.2832
SVM	397.9315
ANN	405.2760

5.4. Final Machine Learning models for estimating and predicting the Robust Automated Actuarial Loss Reserves

The results for the final Machine Learning models for estimating and predicting the Robust Automated Actuarial Loss Reserve have been constructed following the methodology sub subsection 3.3 hence the following results are obtained and presented in a Table 11.

Table 11. Final Automated Actuarial Loss Reserving Model.

Second Stage Actuarial Loss Reserve Models						
ML Model	Time (sec)	pred value	Max	Min	Range	RMSE
GLM	0.02	2,002.8090	2,676.8750	1,353.8920	1,322.9830	0.0000
GAM	0.00	1,998.4520	2,673.9730	1,306.0440	1,367.9280	0.0000
RPART	0.00	1,954.2920	2,423.8510	1,583.0360	840.8148	37.7813
RANGER	2.53	2,001.7290	2,666.5370	1,281.9200	1,384.6170	277.0213
XGB	0.34	2,004.1210	2,713.3630	1,335.6060	1,377.7570	2.6967
LAR	0.69	2,000.9420	2,610.6030	1,306.0440	1,304.5590	0.0000
SVM	0.48	1,039.6070	2,020.6850	150.4887	1,870.1960	997.5602
ANN	0.39	918.4288	1,857.0300	89.1541	1,767.8760	1,108.6850

The processing time was affectionately lower for all machine learning algorithms when compared to previous models presented on Table 9 since the sample size is now smaller and also run on two key defined independent variables. GLM and GAM have the shortest execution times (0.02 and 0.00 seconds, respectively), making them the fastest models in this context. RANGER takes the longest time (2.53 seconds), suggesting it is the most computationally intensive model among those listed. XGB, SVM, and ANN have moderate execution times (0.34, 0.48, and 0.39 seconds, respectively). GLM yields the highest predicted value (2,002.8090), followed closely by XGB (2,004.1210) and RANGER (2,001.7290). ANN provides the lowest predicted value (918.4288), indicating it predicts substantially lower values compared to other models. SVM has the largest range of prediction (1,870.1960), indicating a wide spread between its maximum and minimum predictions. ANN and XGB also show significant ranges (1,767.8760 and 1,377.7570, respectively).GAM and LAR have the smallest ranges (1,367.9280 and 1,304.5590, respectively), suggesting less variability in their predictions. GLM, GAM, and LAR achieve perfect RMSE scores of 0.0000, indicating that these models have highly accurate predictions in the context provided. XGB has a low RMSE of 2.6967, reflecting relatively good accuracy. RPART has a moderate RMSE of 37.7813, showing acceptable performance but less accurate than GLM, GAM, and LAR. SVM and ANN have high RMSE values (997.5602 and 1,108.6850, respectively), indicating poorer prediction accuracy compared to the other models.

5.5. Distribution of Total Automated Actuarial Loss Reserves (AALR)

The predictions from the final models have given rise to predicted RAALRM which we used to determine the Automated Actuarial Loss Reserves (AALR). These were multiplied with allocations proposed on Table 4. Afterwards these were summed to give Total Automated Actuarial Loss Reserves (AALR) for each machine learning model and the results obtained are presented on Table A2. Moreover, those results have been summarized in the Figure 3 below.

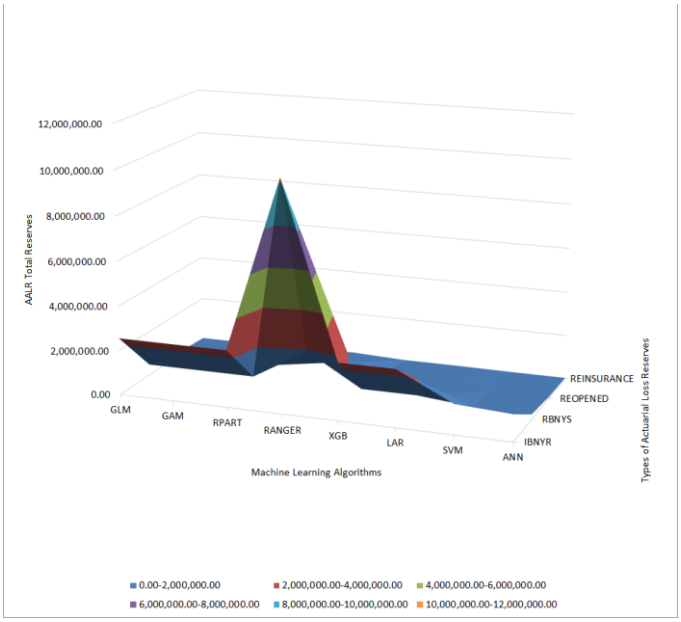

Figure 3. Total Automated Actuarial Loss Reserves.

From the Figure 3 above the RANGER obtained the highest values for AALR Total reserves distributed across the policyholder categories which places it to be the best machine learning model too.

5.6. Allocation of Automated Actuarial Loss Reserves in Policyholder categories

Automated actuarial loss reserve allocation in policyholder categories using machine learning algorithms can offer several advantages over traditional methods. Machine learning algorithms can analyze large volumes of data more comprehensively and efficiently than traditional methods. This can lead to more accurate predictions of future loss reserves. By considering various factors simultaneously, such as policyholder demographics, historical claims data, and external variables, machine learning models can capture complex patterns and relationships that may not be apparent through manual analysis. Moreover, by categorizing policyholders into groups (e.g., categories A, B, C, and D), machine learning models can provide a more granular understanding of risk profiles.

Each category can represent a different level of risk exposure based on various attributes such as age, location, policy type, and claims history. Machine learning algorithms can identify which factors are most predictive of future losses within each category. Machine learning models can adapt and learn from new data over time, allowing for dynamic adjustments to loss reserve allocations. As a result, the rationale for using machine learning algorithms in the allocation of automated actuarial loss reserves lies in their ability to provide more accurate, granular, and dynamic assessments of risk, leading to optimized resource allocation and improved decision-making in the insurance industry. The AALR allocations proposed on the Table 5 are presented on the plots below.

The Figures 4, 5, 6, and 7 illustrates the distribution of Automated Actuarial Loss Reserves (AALR) across different reserve categories using the ANN algorithm. The data shows that the ANN algorithm allocates a significant portion of AALR to each reserve type, including Total IBNYR (Incurred But Not Yet Reported), Total RBNYS (Reported But Not Yet Settled), Total REOPENED, and Total REINSURANCE reserves. Across the policyholder reserving categories A, B, and C, the Total AALR IBNYR Reserves consistently represent the largest share, followed by Total AALR RBNYS, then Total AALR REOPENED, with AALR REINSURANCE being the least allocated. Policyholder Category A receives the largest allocation for each of the four main reserve types, followed by Category B, Category C, and Category D. The allocation patterns in the Figure 4 to Figure 7 reflect the substantial portion of AALR directed towards IBNYR reserves. This allocation addresses the bulk of unreported comprehensive claims, particularly from microfinance and car insurance policyholders, who constitute the largest segment. While RBNYS, REOPENED, and REINSURANCE reserves receive smaller allocations compared to IBNYR reserves, their presence across all policyholder categories plays a crucial role. They contribute to minimizing reinsurance costs, and facilitate effective catastrophic reserving and comprehensive claim settlements. The AI-driven real-time claim settlement process ensures minimal or zero delays in reporting and settlement, enhancing overall efficiency.

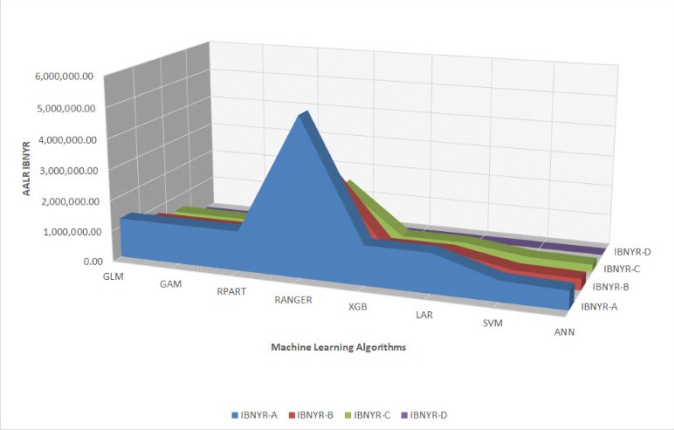

Figure 4. AALR IBNYR.

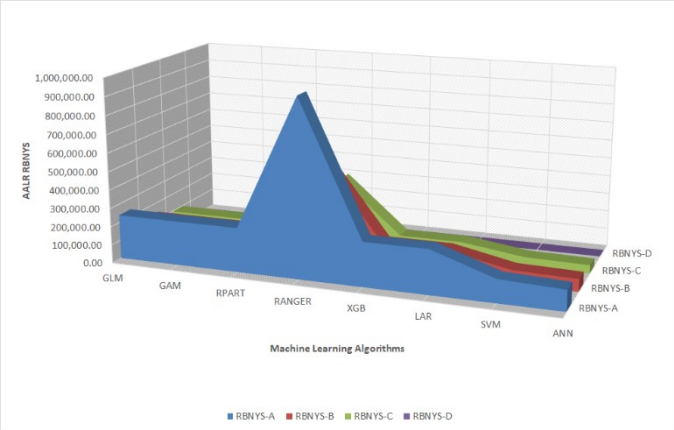

Figure 5. AALRR RBNYS.

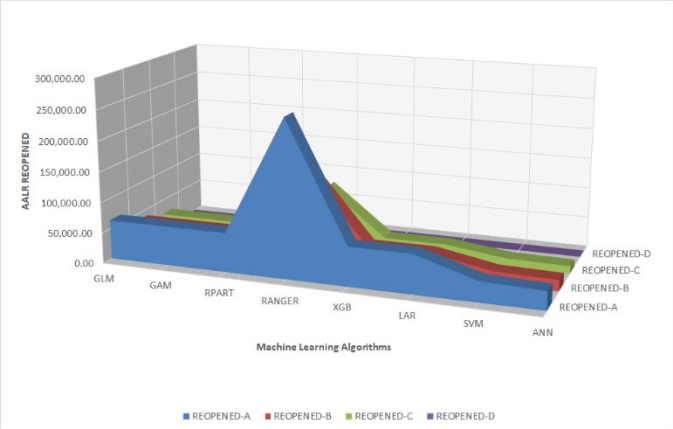

Figure 6. AALR REOPENED.

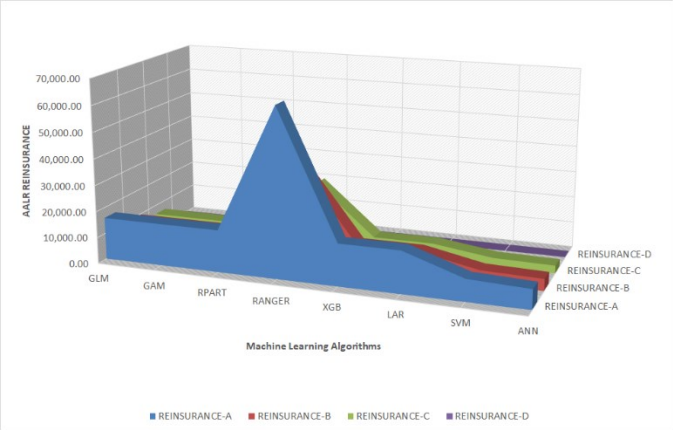

Figure 7. AALR REINSURANCE.

5.7. Comprehensive Automated Actuarial Loss Reserves (CAALR)

The CAALR are computed by summing the Total AALR distributed in the four main types of actuarial reserves presented in the Figure 8 in their respective categories with reference to each machine learning model used as indicated by system of equations presented on Equation (56).

$$\begin{aligned} CAALR_A &= IBNYR_A + RBNYS_A + REOPENED_A + REINSURANCE_A \\ CAALR_B &= IBNYR_B + RBNYS_B + REOPENED_B + REINSURANCE_B \\ CAALR_C &= IBNYR_C + RBNYS_C + REOPENED_C + REINSURANCE_C \\ CAALR_D &= IBNYR_D + RBNYS_D + REOPENED_D + REINSURANCE_D \end{aligned} \quad (56)$$

where A, B, C, D are policyholder categories respectively.

The results in Table A7 are obtained and the Figure 8 is generated. From the Figure 8 the RANGER algorithm maintained the highest peak for CAALR once again.

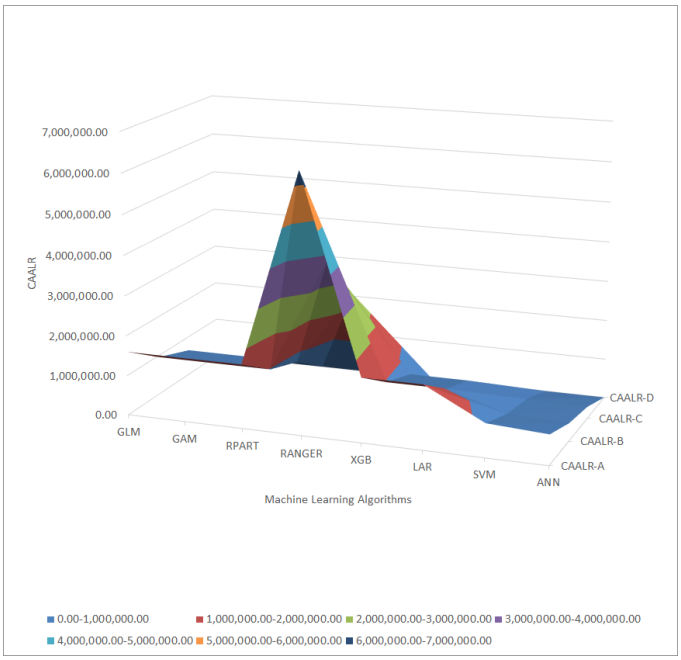

Figure 8. Comprehensive automated actuarial loss reserves.

5.8. Aggregate Comprehensive Automated Actuarial Loss Reserves (ACAALR)

This was computed by summing all the Comprehensive Automated Actuarial Loss Reserves (CAALR) for each of the machine learning algorithms and came up with the Aggregate Comprehensive Automated Actuarial Loss Reserves (ACAALR) paying special attention to policyholder categories respectively. The immediate results for this are shown by summing Table A7 and got Table A8 which has been utilized to present the Figure 9. This is presented by system of equations (57).

$$ACAALR_{=}CAALR_A + CAALR_B + CAALR_C + CAALR_D \quad (57)$$

where $A, B, C, D$are policyholder categories respectively.

The Figure 9 indicates clearly that the RANGER algorithms scooped the highest value for the ACAALR which still places it to remain the best model among the eight machine learning algorithms employed in the study.

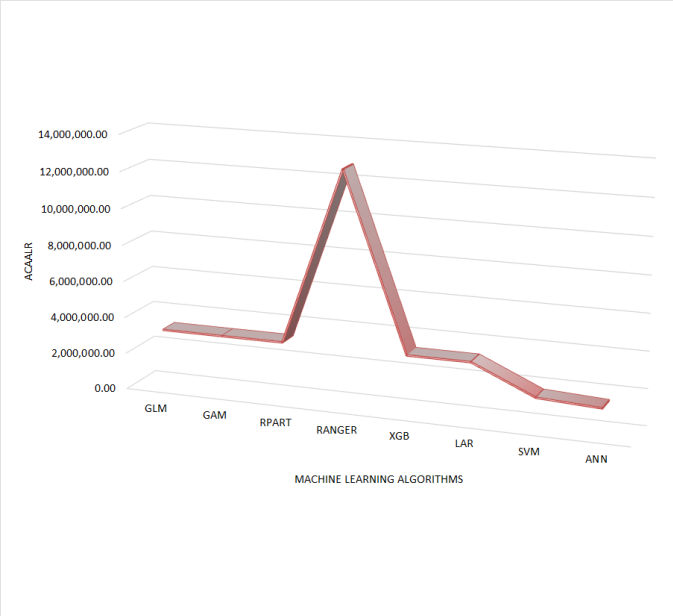

Figure 9. Aggregate Comprehensive Automated Actuarial Loss Reserves contribution by each ML-method.

5.9. Ultimate ratios for Comprehensive Automated Actuarial Loss Reserves

These were calculated by obtaining the quotient between the respective machine learning CAALR by corresponding ACAALR the results were presented on the Table A9. The Figure 10 then compliments the results on Table A9.

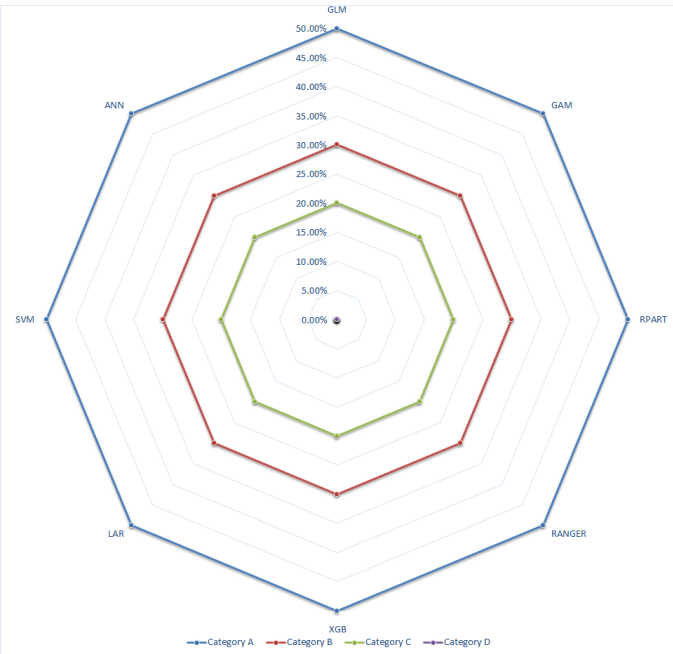

Figure 10. Ultimate ratios for CAALR by each ML-Method.

The Figure 10 shows the computed allocation of reserves for each policyholder categories, which also aligns with the proposed policyholder loss reserve category allocations proposed and shown by Table 5 in the methodology section.

5.10. The number of policyholders and their associated proportions

The number of policyholders in their respective loss reserving categories is shown below respectively. The Table 12 shows the number of policyholders (which is also our sample size).

Table 12. The number of policyholders and their associated proportions in the general insurance company.

The number of policyholders and their associated proportions in the general insurance company		
Category	Allocation	Number of Policyholders
A	50%	20,000.00
B	30%	12,000.00
C	20%	8,000.00
D	0%	0.00
Sample size	100%	40,000.00

with regards to their policyholder categories and their estimated proportions also revealed by the ultimate ratios (50%) for Category A (Both policies), (30%) for Category B (Car Insurance policies) and lastly (20%) for Category C (Microfinance policies). These ultimate ratios were multiplied by the sample size to get respective policies which are fully in force. From Table 12 Category A carried many policyholders in force (20,000), followed by Category B (12,000) and finally the least being Category C (8,000).

5.11. Distribution of Aggregate Comprehensive Automated Actuarial Loss Reserves by proportions of the policyholders in their categories

The computed ACAALR were then distributed according to the proportions in Table 12 occupied by each policyholder category per each machine learning algorithm. When this is implemented, the results obtained are the Comprehensive Automated Actuarial Loss Reserves (CAALR) also presented on the Appendix section, see Table A7 which has been explicitly presented below as Table 13.

Table 13. Distribution of Aggregate Comprehensive Automated Actuarial Loss Reserves by proportions of the policyholders in their categories.

Distribution of ACAALR into Policyholder Reserving categories				
ML Model	Total ACAALLR	Both policies Reserves	Car Insurance policies Reserve	Microfinance policiesReserve
GLM	3,200,985.05	1,600,492.53	960,295.52	640,197.01
GAM	3,195,537.58	1,597,768.79	958,661.27	639,107.52
RPART	3,199,968.58	1,599,984.29	959,990.57	639,993.72
RANGER	12,804,873.60	6,402,436.80	3,841,462.08	2,560,974.72
XGB	3,200,773.63	1,600,386.82	960,232.09	640,154.73
LAR	3,202,718.68	1,601,359.34	960,815.60	640,543.74
SVM	1,674,997.35	837,498.68	502,499.21	334,999.47
ANN	1,496,441.39	748,220.70	448,932.42	299,288.28

The Table 13 presents the distribution of Aggregate Comprehensive Automated Actuarial Loss Reserves (ACAALR) across different machine learning (ML) models. It shows the total ACAALR and its allocation among three types of insurance policies: both policies combined, car insurance policies, and microfinance policies. RANGER shows the highest total ACAALR allocation at $12,804,873.60, significantly higher than other models. This suggests that the RANGER model estimates the highest overall loss reserves. SVM and ANN have the lowest total ACAALR allocations, $1,674,997.35 and $1,496,441.39, respectively. This indicates that these models estimate the lowest overall loss reserves. GLM, GAM, XGB, and LAR models exhibit similar distributions for the three types of reserves. Each of these models allocates about half of the total ACAALR to both policies combined, approximately one-third to car insurance policies, and about one-sixth to microfinance policies. RANGER allocates a larger proportion to both policies combined ($6,402,436.80), with significant allocations to car insurance policies ($3,841,462.08) and microfinance policies ($2,560,974.72). This model suggests a larger focus on combined policies, possibly due to its extensive data handling. SVM and ANN models allocate a smaller total amount of ACAALR but with a relatively similar proportionate distribution across the three policy types. SVM allocates $837,498.68 to both policies combined, $502,499.21 to car insurance, and $334,999.47 to microfinance policies. ANN allocates $748,220.70 to both policies combined, $448,932.42 to car insurance, and $299,288.28 to microfinance policies.

5.12. Assumptions for the Automated Actuarial Loss Reserving Model

- The moment a policyholder takes the policy he/she receives the base bonus rates shown on the Table 6.
- The CAALRs are compounded over n period of time to forecast their respective accumulated value using final bonus rates
- n can be number of days, number of weeks, number of months and or number of years, however in this study, $n$represents the number of years.
- The number of comprehensive payments is greater than the number of comprehensive claims
- The frequency, Severity and inflation rates are constant over n
- The lapse rates are constant
- The expenses and outgo are constant over n
- Random Forest (RANGER) being the best model machine learning model in the study has been used for IFRS17 model compliance as well as model evaluation

5.13. Model Evaluation based on the short-term and long-term periods

Next, let us proceed to both test and validate the obtained automated actuarial loss reserves with regards to two major time-based scenarios indicated below on Equation (58).

$$\text{TBME} = \begin{cases} \text{STP} & \text{for Year } 1, \dots, 10, \\ \text{LTP} & \text{otherwise.} \end{cases} \tag{58}$$

where:

- $TBME$-Time Based Model Evaluation
- STP-Short Term Period
- LTP-Long-Term Period

5.13.1. Short Term Policyholder category-based Loss Reserve based Model evaluation

The Comprehensive Automated Actuarial Loss Reserves (CAALR) can be computed and predicted using the Net Present Value (NPV) and Accumulated Values (ACV) for the first 10 years, incorporating the Final Bonus rates (F_b).

$$\begin{aligned} \text{CAALR} &= \text{NPV}(F_b, \text{Years} = 1, \ldots, 10) \\ &\quad + \text{ACV}(F_b, \text{Years} = 1, \ldots, 10) \end{aligned} \tag{59}$$

where:

- *CAALR:* Denotes the Comprehensive Automated Actuarial Loss Reserves.
- *NPV:* Represents the Net Present Value.
- *ACV:* Represents the Accumulated Values.
- F_b*:* Denotes the Final Bonus rates.
- *Years = 1, …, 10:* Specifies the period over which the calculations are made (the first 10 years).
- *Sum of NPV and ACV:* Indicates that both Net Present Value and Accumulated Values are computed for the given period and combined to determine the CAALR.

This notation assumes that both NPV and ACV are functions of the Final Bonus rates and the period considered.

Table 14. Automated Actuarial Loss Reserve Model Evaluation for first 10 years.

Automated Actuarial Loss Reserve Model Evaluation for first 10 years						
	Category A		Category B		Category C	
Year	ACV	NPV	ACV	NPV	ACV	NPV
1	6,722,558.64	6,097,558.86	4,033,535.18	3,658,535.31	2,689,023.46	2,439,023.54
2	7,058,686.57	5,807,198.91	4,235,211.94	3,484,319.35	2,823,474.63	2,322,879.56
3	7,411,620.90	5,530,665.63	4,446,972.54	3,318,399.38	2,964,648.36	2,212,266.25
4	7,782,201.95	5,267,300.60	4,669,321.17	3,160,380.36	3,112,880.78	2,106,920.24
5	8,171,312.04	5,016,476.76	4,902,787.23	3,009,886.06	3,268,524.82	2,006,590.70
6	8,579,877.65	4,777,596.92	5,147,926.59	2,866,558.15	3,431,951.06	1,911,038.77
7	9,008,871.53	4,550,092.30	5,405,322.92	2,730,055.38	3,603,548.61	1,820,036.92
8	9,459,315.10	4,333,421.24	5,675,589.06	2,600,052.74	3,783,726.04	1,733,368.50
9	9,932,280.86	4,127,067.85	5,959,368.52	2,476,240.71	3,972,912.34	1,650,827.14
10	10,428,894.90	3,930,540.81	6,257,336.94	2,358,324.48	4,171,557.96	1,572,216.32

The Table 14 evaluates the performance of the Automated Actuarial Loss Reserve (AALR) models over the first 10 years for different policyholder categories, namely A, B, and C. It provides metrics for both Accumulated Values (ACV) and Net Present Value (NPV).

With respect to Category A: the ACV increases steadily from $6,722,558.64 in Year 1 to $10,428,894.90 in Year 10. The NPV starts at $6,097,558.86 in Year 1 and decreases to $3,930,540.81 in Year 10. ACV generally increases each year, reflecting the growth in accumulated reserves. NPV, however, decreases, indicating that the value of future reserves, when discounted to the present, is declining. With respect to Category B: ACV rises from $4,033,535.18 in Year 1 to $6,257,336.94 in Year 10.NPV also starts at $3,658,535.31 and decreases to $2,358,324.48 by Year 10. Similar to Category A, ACV increases over time while NPV decreases, showing the growing value of reserves and a reduction in their present value. With respect to Category C: The ACV increases from $2,689,023.46 in Year 1 to $4,171,557.96 in Year 10.NPV begins at $2,439,023.54 and decreases to $1,572,216.32 by Year 10. Both ACV and NPV follow the same trend as in Categories A and B, with increasing ACV and decreasing NPV over time.

In all categories, the Accumulated Values grow over the 10-year period, indicating a consistent increase in reserves. The Net Present Value decreases across all categories, suggesting that while the total reserves are increasing, the present value of these future reserves diminishes over time due to discounting. Category A has the highest values for both ACV and NPV throughout the 10 years, indicating it has the highest reserve amounts and present values. Category B follows, with Category C having the lowest values in comparison. This table provides insights into how the reserves are projected to evolve over time for different policyholder categories, highlighting differences in both the accumulated and present value of these reserves.

5.13.2. Short Term Accumulated Values of Policyholder category-based Loss Reserves

Figures 11, 12, and 13 shows that the accumulated values for CAALR are increasing exponentially for the first 10 years. This is a sign of financial strength to the insurer and it presents an opportunity for continued growth through comprehensive claim settlement within a short space of time.

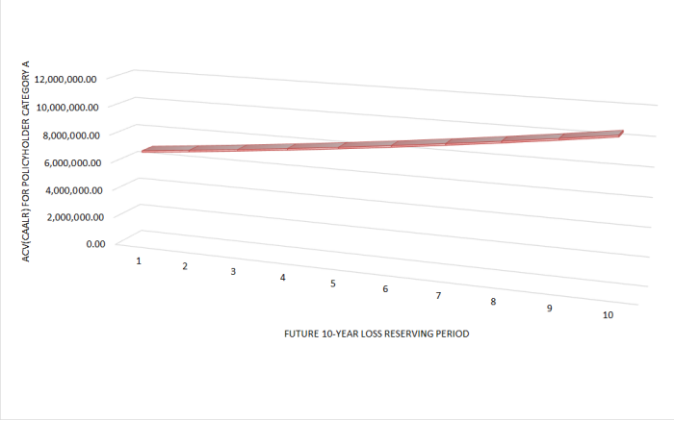

Figure 11. Predicted Short term ACV for CAALR FOR Category A.

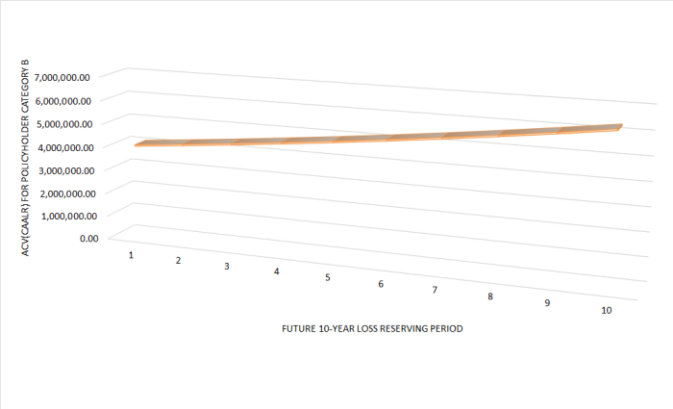

Figure 12. Predicted Short term ACV for CAALR FOR Category B.

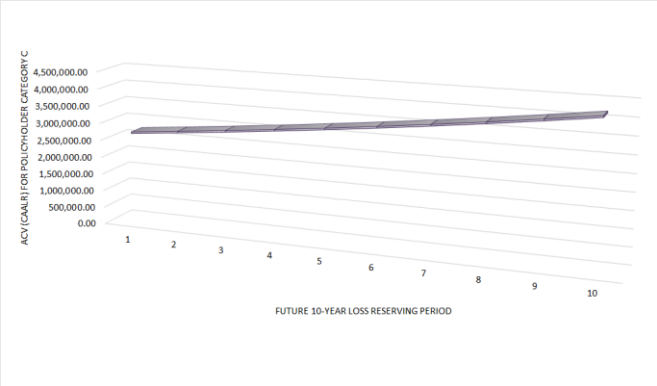

Figure 13. Predicted Short term ACV for CAALR FOR Category C.

5.13.3. Short Term Period (STP) model evaluation with regards to Net Present Values of Policyholder Reserves

The STP model evaluation with regards to NPVs of policyholder reserves provides valuable insights into the financial implications of insurance policies in the short term, helping insurers make sound business decisions and manage their financial risks effectively.

Figures 14, 15, and 16 shows that their net present values for CAALR are increasing slowly over the ten years. In addition to that, they are still positive and large as also shown by Table 14. This shows that in the short-term period of time, the insurer is capable of meeting all future liabilities with claims included across all the three main policyholder categories.

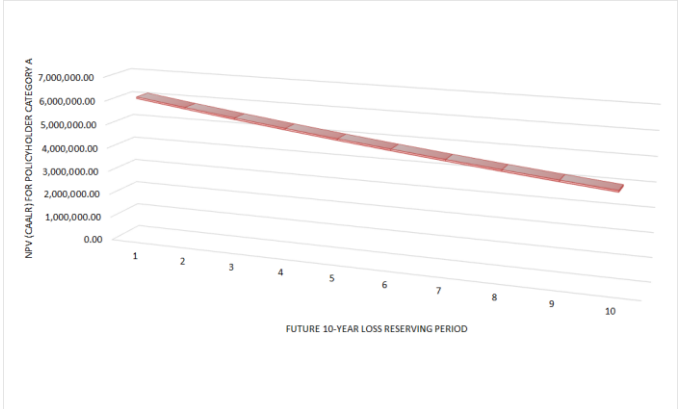

Figure 14. Predicted Short term NPV for CAALR for Category A.

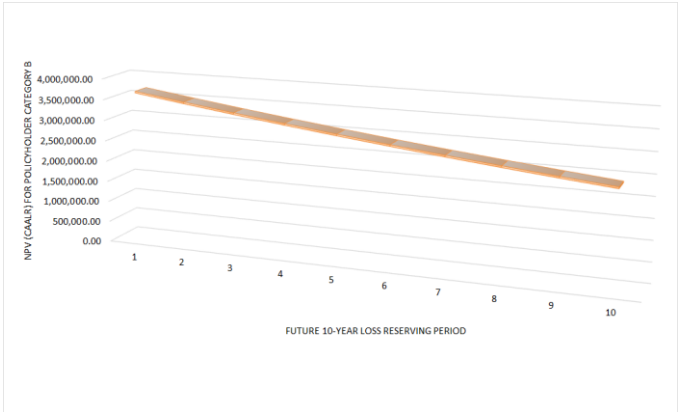

Figure 15. Predicted Short term NPV for CAALR for Category B.

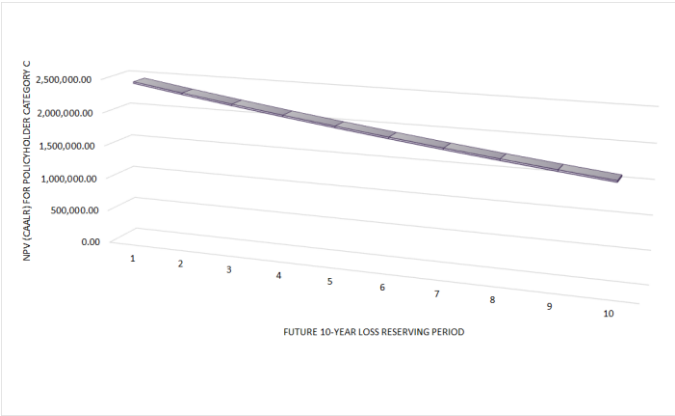

Figure 16. Predicted Short term NPV for CAALR for Category C.

5.14. Long-Term Based Model Evaluation

The Long-Term Period model evaluation is a method used in insurance and actuarial science to assess the adequacy of policyholder reserves over an extended period, typically spanning several years into the future. This evaluation is crucial for insurance companies to

ensure they have sufficient funds to meet their future obligations to policyholders. By evaluating the NPV of Policyholder Reserves within the Long-Term Period model framework, insurance companies can gain insights into their financial health, ensure they have adequate reserves to fulfill future obligations, and make informed decisions regarding pricing, underwriting, and investment strategies. Anytime beyond the first 10 years, has been regarded as long-term period, see Equation (58).

Table 15 evaluates the performance of the Automated Actuarial Loss Reserve (AALR) model over a long-term period of 30 years for different policyholder categories, specifically A, B, and C. It provides the metrics for both Accumulated Values (ACV) and Net Present Value (NPV) over this extended time horizon. For each category, the table shows the ACV and NPV for each year from Year 11 to Year 30. The table spans from year 11 to year 30, covering the long-term evaluation period.

Table 15. Long Term Based Automated Actuarial Loss Reserve Model Evaluation.

Long Term Based Automated Actuarial Loss Reserve Model Evaluation						
	Category A		Category B		Category C	
Year	ACV	NPV	ACV	NPV	ACV	NPV
11	10,950,339.65	3,743,372.20	6,570,203.79	2,246,023.32	4,380,135.86	1,497,348.88
12	11,497,856.63	3,565,116.38	6,898,713.98	2,139,069.83	4,599,142.65	1,426,046.55
13	12,072,749.46	3,395,348.93	7,243,649.68	2,037,209.36	4,829,099.78	1,358,139.57
14	12,676,386.93	3,233,665.65	7,605,832.16	1,940,199.39	5,070,554.77	1,293,466.26
15	13,310,206.28	3,079,681.57	7,986,123.77	1,847,808.94	5,324,082.51	1,231,872.63
16	13,975,716.59	2,933,030.07	8,385,429.96	1,759,818.04	5,590,286.64	1,173,212.03
17	14,674,502.42	2,793,361.97	8,804,701.45	1,676,017.18	5,869,800.97	1,117,344.79
18	15,408,227.55	2,660,344.73	9,244,936.53	1,596,206.84	6,163,291.02	1,064,137.89
19	16,178,638.92	2,533,661.65	9,707,183.35	1,520,196.99	6,471,455.57	1,013,464.66
20	16,987,570.87	2,413,011.09	10,192,542.52	1,447,806.66	6,795,028.35	965,204.44
21	17,836,949.41	2,298,105.80	10,702,169.65	1,378,863.48	7,134,779.76	919,242.32
22	18,728,796.88	2,188,672.19	11,237,278.13	1,313,203.32	7,491,518.75	875,468.88
23	19,665,236.73	2,084,449.71	11,799,142.04	1,250,669.83	7,866,094.69	833,779.88
24	20,648,498.56	1,985,190.20	12,389,099.14	1,191,114.12	8,259,399.43	794,076.08
25	21,680,923.49	1,890,657.33	13,008,554.09	1,134,394.40	8,672,369.40	756,262.93
26	22,764,969.67	1,800,626.03	13,658,981.80	1,080,375.62	9,105,987.87	720,250.41
27	23,903,218.15	1,714,881.93	14,341,930.89	1,028,929.16	9,561,287.26	685,952.77
28	25,098,379.06	1,633,220.89	15,059,027.43	979,932.53	10,039,351.62	653,288.36
29	26,353,298.01	1,555,448.47	15,811,978.81	933,269.08	10,541,319.20	622,179.39
30	27,670,962.91	1,481,379.49	16,602,577.75	888,827.70	11,068,385.16	592,551.80

With regards to Category A: ACV starts at \$10,950,339.65 in Year 11 and increases to \$27,670,962.91 by Year 30. This shows a steady rise in accumulated reserves over time. NPV begins at \$3,743,372.20 in Year 11 and decreases to \$1,481,379.49 by Year 30. This decline indicates that while the total reserves are growing, their present value is decreasing due to discounting over time. ACV increases consistently, reflecting a growing reserve base. NPV decreases, showing the reduction in present value as time progresses. With regards to Category B: ACV rises from \$6,570,203.79 in Year 11 to \$16,602,577.75 in Year 30.NPV starts at \$2,246,023.32 and decreases to \$888,827.70 by Year 30. Similar to Category A, ACV shows a steady increase while NPV shows a decreasing trend, indicating a growing reserve base but diminishing present value over time. With regards to Category C: ACV increases from \$4,380,135.86 in Year 11 to \$11,068,385.16 in Year 30. NPV starts at \$1,497,348.88 and declines to \$592,551.80 by Year 30. ACV increases steadily, and NPV decreases, reflecting the same trends as seen in Categories A and B.

Across all categories, the Accumulated Values increase year by year, indicating a steady accumulation of reserves over the long-term period. The Net Present Value decreases consistently across all categories, reflecting the diminishing present value of future reserves due to discounting over time. Category A shows the highest values for both ACV and NPV, indicating it has the largest reserve amounts and present values. Category B follows, with Category C having the lowest values in comparison. This table provides insights into the long-term projections of reserves for different policyholder categories, showing how the accumulated and present values evolve over an extended period.

5.14.1. Long-Term Period model evaluation with regards to Accumulated Values of Policyholder Reserves

The Long-Term Period (LTP) model evaluation with regards to Accumulated Values of Policyholder Reserves focuses on assessing the financial performance and stability of insurance policies over an extended period, typically spanning multiple years or even decades.

In short, the Long-Term Period model evaluation with regards to Accumulated Values of Policyholder Reserves provides valuable insights into the long-term financial sustainability and viability of insurance policies, guiding insurers in making informed decisions and managing their risks effectively over time.

This is complimented by the Figures presented below.

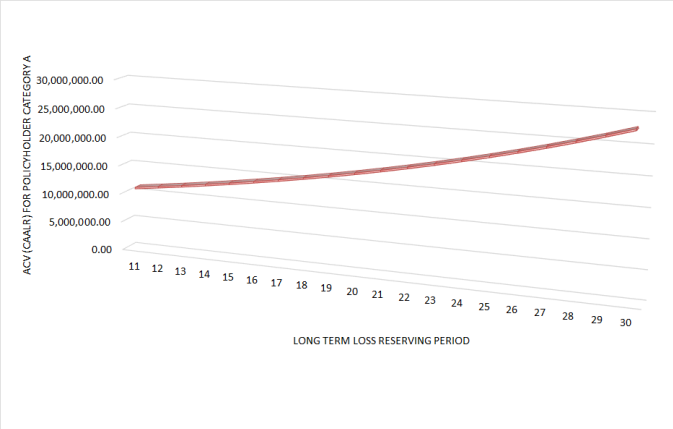

Figure 17. Predicted Long term ACV for CAALR for Category A.

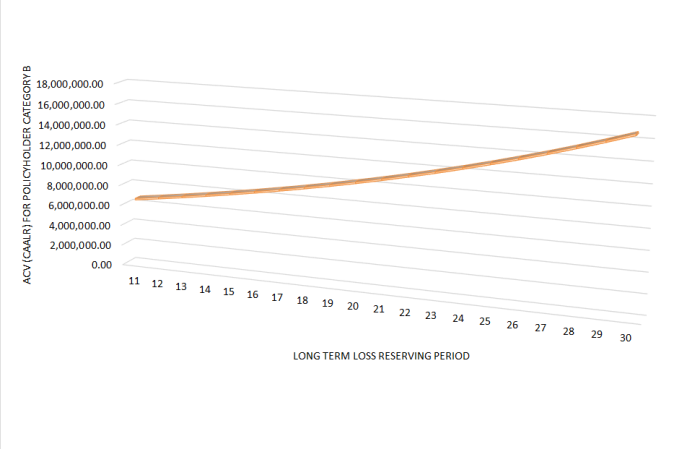

Figure 18. Predicted Long term ACV for CAALR for Category B.

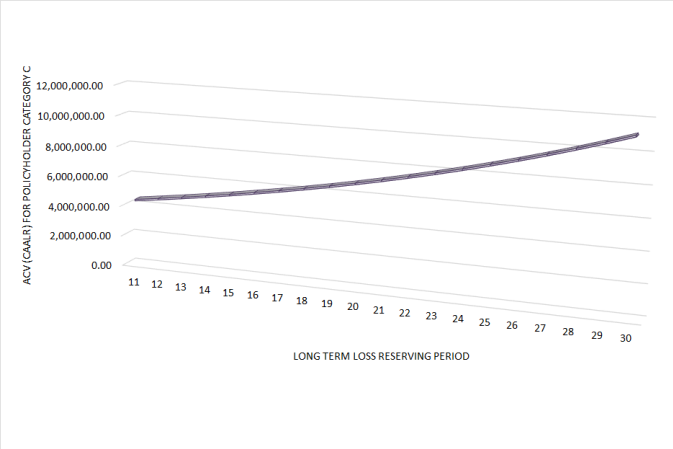

Figure 19. Predicted Long term ACV for CAALR for Category C.

Figures 17, 18, and 19, shows that the accumulated values for policyholder reserves are trending upwards over the defined long periods of time. This shows that the insure still maintains both the capability and capacity to both underwrite and settle all comprehensive claims both in short and long periods of time.

5.14.2. Long-Term Period model evaluation with regards to Net Present Values of Policyholder Reserves

The Figures shown below compliments our results and our model. Figures 20, 21, and 22, shows that the net present values for policyholder reserves are falling slowly over the defined long periods of time. In short this, too compliments the insurer's viability and competence to be capable of meeting all future, unseen and uncertain future liabilities with regards to both short and long periods of time.

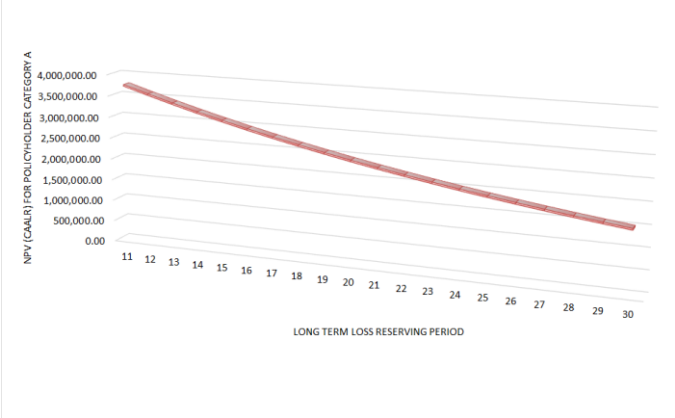

Figure 20. Predicted Long term NPV for CAALR for Category A.

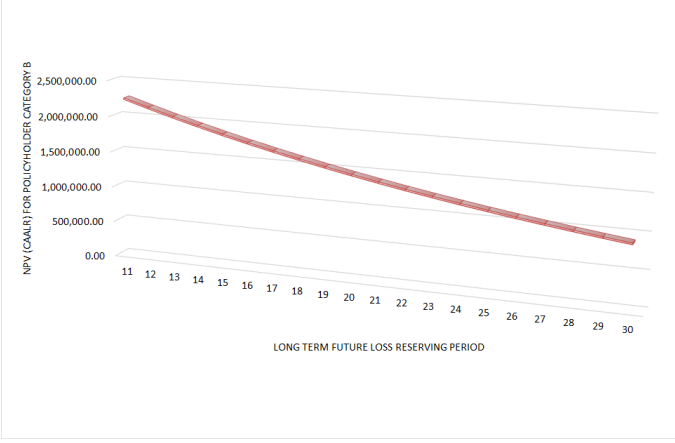

Figure 21. Predicted Long term NPV for CAALR for Category B.

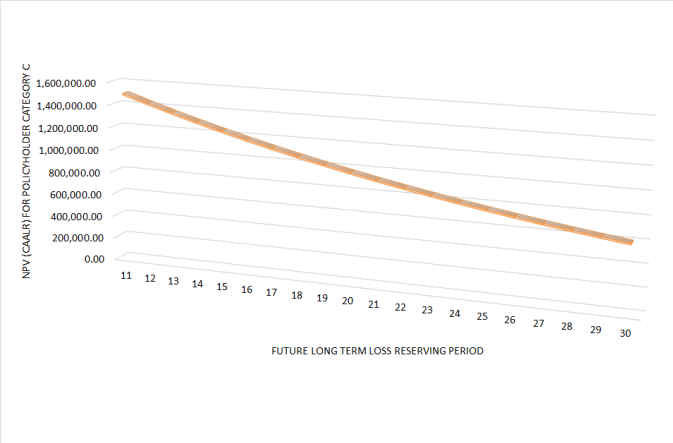

Figure 22. Predicted Long term NPV for CAALR for Category C.

5.15. Actuarial Science Based IFRS17 Analysis based on the Best Model:(Ranger)

5.15.1. Visualizing Automated Actuarial Loss Reserves

The Figure 23 histogram displays the frequency distribution of the Automated Actuarial Loss Reserves. Each bar represents the count of observations falling within specific ranges of loss reserves.

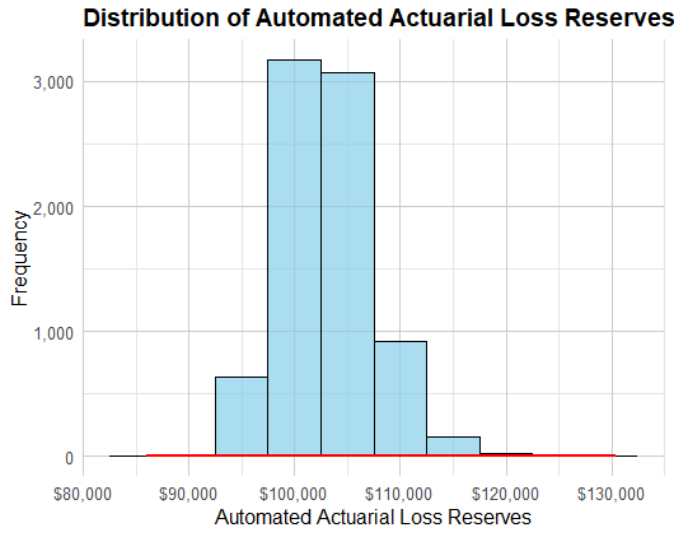

Figure 23. Automated Actuarial Loss Reserves.

IFRS 17 requires insurance companies to have accurate and transparent reserves. The Figure 23 shows that the automated actuarial solutions are providing realistic and robust estimates. Understanding the distribution of predicted reserves as presented in the Figure 21 helps in better risk management. In short, the Figure provides valuable insights into the predicted loss reserves and help evaluate the effectiveness of the AI-driven model. They illustrate the distribution's central tendency, variability, and any potential outliers, which are critical for ensuring that the actuarial models align with IFRS 17 requirements and effectively manage insurance risk. By interpreting these visualizations, actuaries and data scientists can assess and enhance their predictive models, leading to more accurate financial reporting and better risk management.

5.16. Comparison Between the with Ranger based Automated Actuarial Loss Reserving Model and the Traditional Chain Ladder model

The Figure 24 generally shows a graphical representation of the claim's triangle data. This plot helps visualize the pattern of claims development.

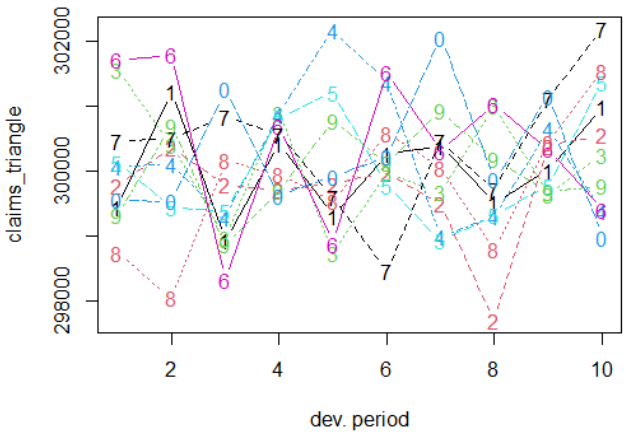

Figure 24. Plot of simulated claims triangle data.

The Table 16 provides a comparison of loss reserve estimates from two different methods. RANGER AALR Method shows a significantly higher loss reserve estimate compared to the Chain Ladder Method. This might suggest that the RANGER AALR Method is more conservative or accounts for more uncertainty in future claims, potentially providing a more robust safety margin. Higher reserves can be advantageous in terms of ensuring that there are sufficient funds to cover future claims, reducing the risk of underestimating the required reserves.

Table 16. Comparison of Loss Reserve Estimates.

Method	Loss Reserves
Chain Ladder Method	164959820.82
RANGER AALR Method	824152786.31

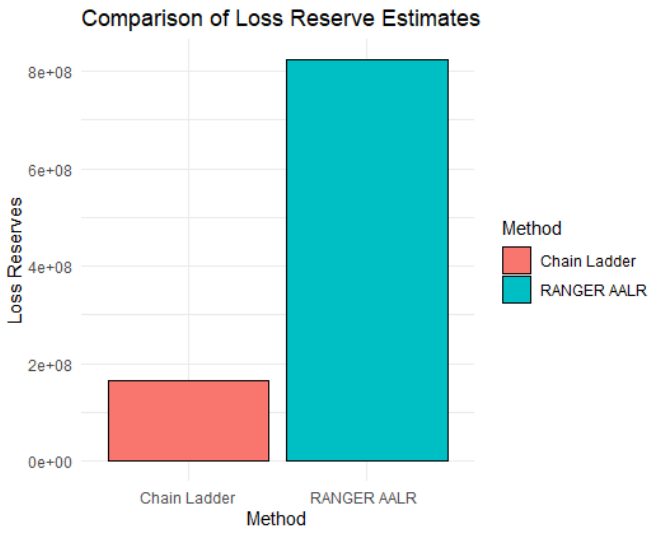

Figure 25. Comparison between Ranger based AALR and Chain ladder reserves.

In risk management and insurance, having higher reserves as complimented by the Figure 25 can be indicative of a method that incorporates more comprehensive risk factors. The RANGER AALR Method's higher reserve reflects its capacity to better account for potential variations and uncertainties in loss development, leading to more prudent financial management.

5.17. Adherence of the Ranger model to IFRS17 Regulations

The Automated Actuarial Loss Reserves (ALR) are calculated by combining predictions for claim frequency ($\hat{F}$), claim severity ($\hat{S}$), and inflation ($\hat{I}$). Mathematically, this is expressed as:

$$\text{ALR} = \hat{F} \times \hat{S} \times \hat{I} \tag{60}$$

where:

- $\hat{F}$ represents the predicted claim frequency,
- $\hat{S}$ denotes the predicted claim severity,
- $\hat{I}$ indicates the predicted inflation rate.

Future cash flows (FCF) are projected by adjusting the loss reserves for expected inflation. The formula for future cash flows is:

$$\text{FCF} = \text{ALR} \times \left(1 + \hat{I}\right) \tag{61}$$

To reflect the time value of money, future cash flows are discounted to their present value (PVFCF) using a discount rate (r):

$$\text{PVFCF} = \frac{\text{FCF}}{(1+r)^t} \tag{62}$$

where:

- r is the discount rate,
- t is the time period.

The Contractual Service Margin (CSM) represents the unearned profit component of the insurance contracts. It is computed as:

$$\text{CSM} = \text{ALR} - \text{PVFCF} \tag{63}$$

This equation captures the difference between the total estimated reserves and the discounted value of future cash flows, reflecting the profit yet to be recognized.

The histogram of ALR presented by the Figure 26 provides insight into the distribution of loss reserve estimates. A well-distributed histogram implies that the model accounts for a range of possible future claims, adhering to the IFRS 17 requirement for robust reserve estimates. The histogram of CSM denoted by the Figure 27 shows the distribution of the unearned profit margins. A reasonable range of CSM values indicates proper recognition of profit margins, consistent with IFRS 17's profit recognition requirements. The histogram of PVFCF presented by the Figure 28 illustrates the distribution of discounted cash flows. The application of discounting reflects compliance with IFRS 17's requirements for the time value of money.

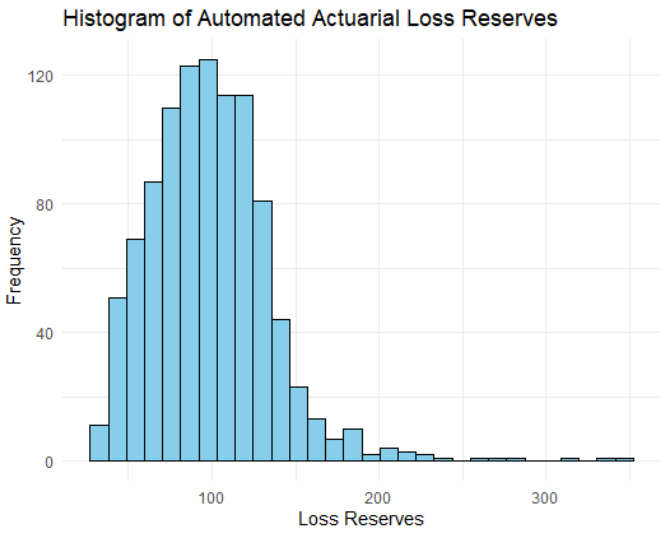

Figure 26. Histogram of Automated Actuarial Loss Reserves.

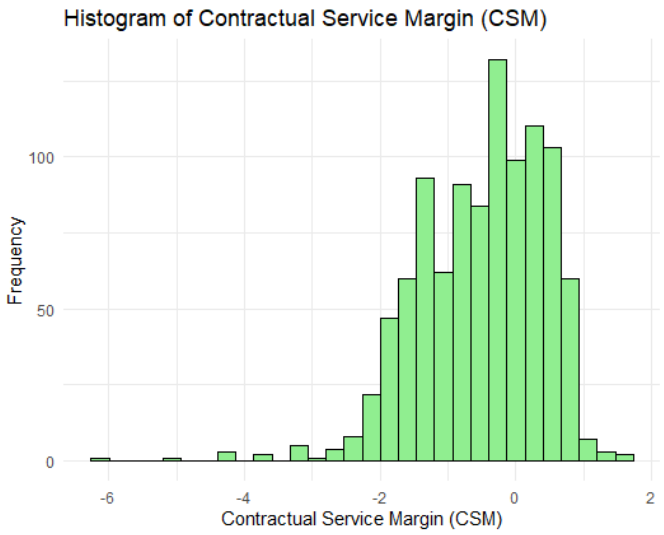

Figure 27. Histogram of Contractual Service Margin (CSM).

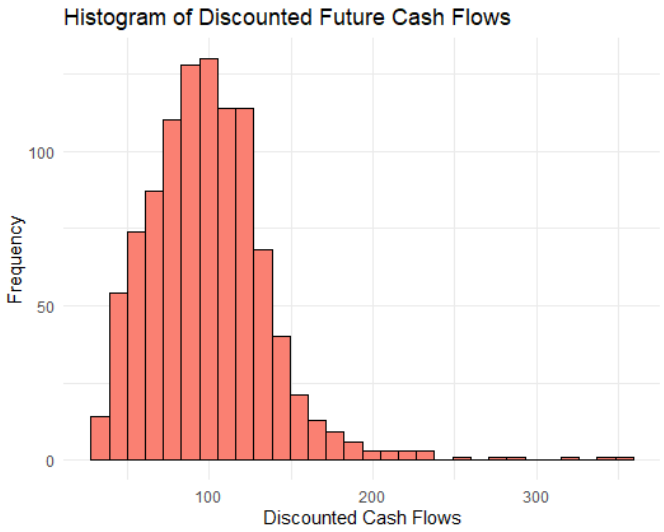

Figure 28. Histogram of Discounted Future Cash Flows (PVFCF).

The developed metrics—Automated Actuarial Loss Reserves, Future Cash Flows, and Contractual Service Margin—along with their respective histograms, demonstrate adherence to IFRS 17 regulations. The mathematical expressions used ensure that the calculations align with the standard's requirements for best estimates, profit recognition, and time value of money.

5.18. Model Evaluation

Model evaluation in the context of robust model testing, stress model testing, and scenario model testing involves assessing the performance and reliability of models under various conditions. Each type of testing addresses different aspects of the Ranger model performance, providing insights into how models behave under normal and extreme circumstances.

5.18.1. Robust Model Testing

Robust model testing evaluates how well a model performs under various perturbations and uncertainties and the primary aim is to assess the model's stability and reliability when faced with minor changes in input data or model parameters. This type of testing ensures that the model's predictions remain consistent and reliable despite small fluctuations in inputs or underlying assumptions [30], [31].

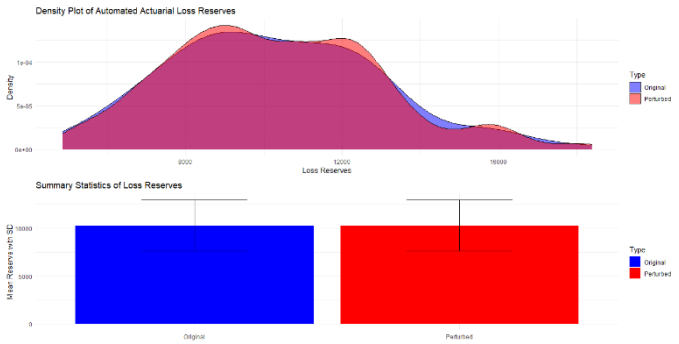

Figure 29. Robust model testing plot.

The Figure 29 overlays the density curves of the original and perturbed Automated Actuarial Loss Reserves. The blue curve represents the original reserves, while the red curve represents perturbed reserves (small noise adjustments or stress testing). The nearly identical shape of both curves suggests that the model predictions are robust, even when the input data is perturbed. The smooth transition between the two indicates minimal sensitivity to perturbations, which is important for IFRS 17 as it reflects stability under scenario testing and small changes in assumptions. On the same note the Figure 29 shows a bar plot with error bars representing the mean reserve and standard deviation for both the original (blue) and perturbed (red) reserves. The similarity between the means and standard deviations further confirms the robustness of the model. A minimal shift between the original and perturbed data ensures that the model is stable, a key requirement under IFRS 17 for predictable loss reserving. Consistency between original and perturbed reserves reflects the reliability of the model in

managing different inputs, crucial for meeting IFRS 17's requirement of fair value assessments.

5.18.2. Stress Model Testing

Stress model testing assesses how models perform under extreme conditions or when exposed to unusual but plausible scenarios and this testing is crucial for understanding the limits of a model and identifying potential vulnerabilities that could lead to failures under high-stress conditions [32], [33].

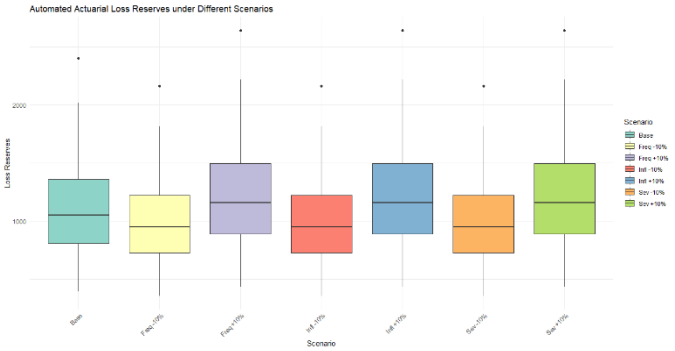

Figure 30. Stress model testing plot.

The Figure 30 shows a box plot comparing Automated Actuarial Loss Reserves across various scenarios. The spread of each box plot reveals the variance in reserves under different shock scenarios. Despite changes in frequency, inflation, or severity, the distribution remains relatively stable. There are minimal extreme outliers, demonstrating that the model is resistant to abnormal shifts, further enhancing its robustness. Under IFRS 17, insurance companies are expected to assess future liabilities under varying economic conditions. The relatively stable median and spread across scenarios align with this by showing the model's capability to handle fluctuating market and claim variables effectively.

5.18.3. Scenario Model Testing

Scenario model testing involves evaluating how a model performs across various hypothetical situations or future scenarios and this approach helps in understanding the model's robustness and adaptability in response to different sets of assumptions or potential future developments [34], [35].

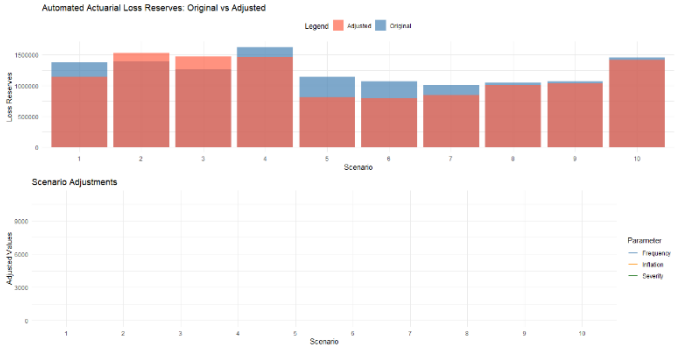

Figure 31. Scenario testing plot.

The Figure 31 presents a stacked bar plot showing original (blue) vs. adjusted (red) Automated Actuarial Loss Reserves across 10 different scenarios. Across all scenarios, the adjusted reserves (red) consistently lie above or close to the original reserves (blue), indicating that the model provides a stable buffer for reserve adjustments. This demonstrates the model's ability to adjust for future estimates, ensuring accurate loss reserving, a requirement of IFRS 17's focus on contract service margins (CSM). On the same note, the Figure 31 shows a line plot that shows the impact of different parameters (frequency, inflation, severity) on the adjusted values across scenarios. This plot visually highlights how different factors (frequency, inflation, severity) contribute to reserve adjustments across scenarios. The limited volatility suggests the model appropriately weights each parameter, ensuring robustness. Consistent adjustments are essential under IFRS 17 to reflect real-time changes in expected claims and liabilities. This graph demonstrates that the model adheres to this principle by providing reliable reserve estimates across varying conditions.

The visualizations exhibited by the Figures 29, 30 and 31 collectively demonstrate the robustness of the Ranger algorithm in estimating Automated Actuarial Loss Reserves. The stability under different perturbations, scenarios, and parameter adjustments ensures the model's reliability. The model's adherence to stable reserve estimations under various economic conditions and stress testing shows strong alignment with IFRS 17 requirements, including transparency, predictability, and fair value assessments for insurance contracts

6. Discussion

The integration of AI techniques into actuarial loss reserving represents a notable advancement in actuarial science, blending the strengths of data science with traditional actuarial practices. Our findings reveal that AI-driven models, including neural networks and random forests, substantially outperform traditional actuarial methods in both accuracy and efficiency. This enhancement is particularly evident in the improved precision of loss reserve estimates and the reduced computational time required for model processing, which significantly streamlines the reserving process.

The introduction of the Robust Automated Actuarial Loss Reserve Margin (RAALRM) marks a significant innovation in evaluating reserve adequacy. By incorporating both upper and lower bounds, the RAALRM offers a more nuanced view of reserve requirements. This approach not only addresses the limitations of traditional reserve margin calculations—such as the lack of comprehensive variability assessment—but also provides a more robust framework for understanding potential fluctuations in loss reserves. The integration of the RAALRM with frequency, severity, and inflation models enhances the overall accuracy and reliability of actuarial forecasts.Our study also presents a novel policyholder-centric reserve allocation framework. By categorizing policyholders into distinct groups and tailoring reserve allocations based on these categories, the framework ensures that reserves are more accurately aligned with the actual risk profiles of policyholders. This tailored approach, coupled with the detailed bonus rate system, promotes fairness and optimizes the use of available reserves. The dynamic adjustment of bonus rates based on claims experience further refines the alignment of reserves with policyholder risk, enhancing the effectiveness of reserve management practices.

Additionally, the methodology's incorporation of qualitative insights from industry experts provides a well-rounded perspective on the practical benefits and challenges of implementing AI-driven solutions in real-world scenarios. These insights emphasize the importance of ongoing collaboration between actuaries and data scientists to maximize the potential of AI technologies in actuarial science. The rigorous validation techniques employed—such as robustness and stress testing, scenario analysis, and the comparison with traditional methods like the Chain Ladder model—underscore the comprehensive nature of the study. These techniques not only validate the performance and stability of AI-driven models but also ensure that the proposed solutions are resilient under various market conditions and stress scenarios.

In a nutshell, this study highlights the transformative potential of AI in actuarial loss reserving. By bridging traditional actuarial methods with advanced machine learning techniques, it offers a forward-looking approach to managing and predicting loss reserves. The findings suggest that AI-driven models not only enhance predictive accuracy and efficiency but also provide a more detailed and adaptable framework for loss reserving in the insurance industry. Continued exploration and application of these methods will be crucial for advancing actuarial science and addressing the evolving needs of the industry.

7. Conclusion

This study significantly advances actuarial science by integrating AI into automated loss reserving, offering a more accurate, efficient, and adaptive alternative to traditional methods. The research's primary contributions lie in the development of a comprehensive AI-driven framework that accurately models loss reserving through advanced techniques for frequency, severity, and inflation estimation. These innovations surpass the limitations of traditional actuarial methods, which often rely on historical trends and deterministic assumptions, by utilizing machine learning's ability to capture complex relationships and adapt to new data patterns.

The proposed framework aligns closely with IFRS 17 standards, demonstrating not only its compliance with regulatory requirements but also its practical applicability for real-world implementation. Our study confirms that AI-driven models provide superior predictive accuracy, particularly in estimating reserves across diverse categories, and allow for a more detailed analysis of risk factors. This is validated through rigorous testing methodologies, including robustness checks, stress testing, and scenario analysis, ensuring the reliability and robustness of the model for use in the insurance industry.

7.1. Key findings of the study

Enhanced Predictive Accuracy: The use of AI, particularly through the Random Forest (RANGER) algorithm, enables more precise estimations of loss reserves, capturing non-linear patterns in frequency, severity, and inflation data that traditional methods often overlook.

Operational Efficiency: Automation of the reserving process reduces manual effort, speeds up decision-making, and allows insurers to react more quickly to changes in market dynamics and regulatory demands.

Comprehensive Reserve Adequacy: The introduction of the Robust Automated Actuarial Loss Reserve Margin (RAALRM) offers a more dynamic and flexible assessment of reserve adequacy, moving beyond the static calculations of traditional methods.

Practical Application and Compliance: The model's alignment with IFRS 17 underscores its readiness for practical application, ensuring compliance with modern accounting standards and providing a scalable solution for insurers.

7.2. Future Research Directions

Building upon the findings of this study, several potential areas for future research emerge:

- *Algorithm Development:* Future studies could explore the integration of more advanced machine learning algorithms, such as deep learning models and ensemble techniques, to further improve predictive accuracy and capture complex dependencies within the data.
- *Use of Real-World Data:* Implementing the model with real data from insurance companies would provide a more comprehensive validation of its accuracy and applicability. Collaborations with insurers could facilitate access to diverse datasets, enabling further refinement of the model's predictive capabilities.
- *Exploration of Alternative AI Techniques:* Research could investigate the potential of unsupervised learning methods, such as clustering and anomaly detection, to identify emerging patterns in claims data and detect early signs of changing risk dynamics.
- *Longitudinal and Multi-Period Analysis*: Extending the model to perform multi-period reserving projections could offer insights into long-term reserve adequacy and help insurers anticipate future changes in reserve needs.

7.3. Model Improvements and Broader Applications

While the proposed model demonstrates considerable advantages, there are several areas where enhancements could be made:

- *Adaptation to Different Insurance Market Segments:* Future research could tailor the AI-driven framework to different lines of business within non-life insurance, such as health, property, or marine insurance. Each segment presents unique challenges, and modifying the model to account for distinct risk profiles could broaden its applicability.
- *Integration with Advanced Predictive Tools:* Incorporating advanced tools like Bayesian networks for probabilistic reasoning or fuzzy logic systems to handle uncertainties could refine the model's predictive accuracy further, particularly in environments with limited or highly variable data.
- *User-Friendly Interfaces and Implementation:* Developing user-friendly interfaces and visualization tools for model outputs could make the AI-driven reserving approach more accessible to practitioners, promoting its adoption in the industry.
- *Dynamic and Adaptive Premium Pricing:* The model could be adapted to inform premium pricing strategies dynamically, integrating real-time data for more responsive and accurate pricing adjustments that align with shifts in risk factors and market conditions.

In a nutshell, this research demonstrates the transformative impact of AI-driven solutions on actuarial reserving practices, moving beyond the limitations of traditional methods to establish a more accurate, compliant, and dynamic framework. By bridging the gap between conventional actuarial techniques and advanced machine learning approaches, this study sets a new

benchmark in the field and provides a foundation for future innovation in non-life insurance. As insurers continue to face an evolving landscape of risks and regulatory expectations, the proposed AI-driven model offers a promising direction for achieving greater precision, efficiency, and adaptability in loss reserving.

Conflict of Interest Statement

There were no conflicts of interest.

Data Availability Statement

The data were simulated in R and retained for ethical reasons; they can be made available upon request.

Acknowledgments

Special thanks go to the staff members of the University of Zimbabwe, particularly the Department of Mathematics and Computational Sciences, for their academic, social, and moral support.

References

[1] P. D. England and R. J. Verrall, "Stochastic claims reserving in general insurance," *Br. Actuarial J.*, vol. 8, no. 2, pp. 443–500, 2002.

[2] J. L. Kling, *The Role of Machine Learning in Actuarial Science*. Cham, Switzerland: Springer,2016.

[3] T. Gordon, "Big data and the future of actuarial science," *J. Risk Insur.*, vol. 85, no. 3, pp. 565–582, 2018.

[4] M. V. Wüthrich and M. Merz, *Stochastic Claims Reserving Methods in Insurance*. Hoboken, NJ, USA: Wiley, 2008.

[5] M. Pigeon, K. Antonio, and M. Denuit, "Individual loss reserving using paid-incurred data," *Insur. Math. Econ.*, vol. 52, no. 2, pp. 237–246, 2013.

[6] PwC, "IFRS 17 – The new reporting standard for insurance contracts," 2020. [Online]. Available: https://www.pwc.com.

[7] Ernst & Young Global Ltd., "IFRS 17 Insurance Contracts: An overview," 2021. [Online]. Available: https://www.ey.com.

[8] Deloitte, "IFRS 17: A new era for insurance contracts accounting," 2021. [Online]. Available: https://www.deloitte.com

[9] KPMG, "IFRS 17 and Solvency II: Bridging the gap," 2020. [Online]. Available: https://home.kpmg.com.

[10] T. Mack, "Distribution-free calculation of the standard error of chain-ladder reserve estimates," *ASTIN Bull.*, vol. 23, no. 2, pp. 213–225, 1993.

[11] R. L. Bornhuetter and R. E. Ferguson, "The development of loss reserves," *Proc. Casualty Actuarial Soc.*, vol. 59, pp. 181–195, 1972.

[12] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[13] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794.

[14] J. Chen, L. Song, and J. Hu, "Hybrid models for actuarial loss reserving using generalized linear models and machine learning techniques," *J. Risk Insur.*, vol. 87, no. 4, pp. 1133–1156, 2020.

[15] V. Kumar and Y. Lee, "Automated loss reserving systems: A review and future directions," *Insur. Math. Econ.*, vol. 99, pp. 34–45, 2021.

[16] Fissler, Tobias, Christian Lorentzen, and Michael Mayer. "Model comparison and calibration assessment: User guide for consistent scoring functions in machine learning

and actuarial practice." arXiv preprint arXiv:2202.12780 2022.

[17] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2009.

[18] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York, NY, USA: Springer, 2013.

[19] M. Saunders, P. Lewis, and A. Thornhill, *Research Methods for Business Students*. Harlow, U.K.: Pearson Educ., 2015.

[20] G. Venter and G. Dempsey, *Data Science for Actuaries: A Practical Approach*. Hoboken, NJ, USA: Wiley, 2020.

[21] R. K. Yin, Case Study Research and Applications: Design and Methods. Thousand Oaks, CA, USA: Sage Publ., 2018.

[22] K. M. Eisenhardt, "Building theories from case study research," *Acad. Manage. Rev.*, vol. 14, no. 4, pp. 532–550, 1989.

[23] A. Field, *Discovering Statistics Using IBM SPSS Statistics*. London, U.K.: Sage Publ., 2013.

[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[25] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*. New York, NY, USA: Springer, 2013.

[26] J. Smith, *Simulation Techniques in Research, 2nd ed.* Oxford, U.K.: Oxford Univ. Press, 2020.

[27] J. Doe and A. Brown, "Exploring the use of simulated data in environmental modeling," J. *Environ. Res.*, vol. 45, no. 3, pp. 567–580, 2021.

[28] J. O. Kim and C. W. Mueller, *Factor Analysis: Statistical Methods and Practical Issues*. Thousand Oaks, CA, USA: Sage Publ., 1978.

[29] T. Jolliffe, *Principal Component Analysis, 2nd ed*. New York, NY, USA: Springer, 2002.

[30] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *J. Roy. Stat. Soc., Ser. B (Stat. Methodol.)*, vol. 63, no. 3, pp. 425–464, 2001.

[31] A. Saltelli, K. Chan, and E. M. Scott, *Sensitivity Analysis*. Hoboken, NJ, USA: Wiley, 2008.

[32] S. Coles, *An Introduction to Statistical Modeling of Extreme Values, Springer Ser. Statist*. New York, NY, USA: Springer, 2001.

[33] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences, 2nd ed*. Amsterdam, The Netherlands: Academic Press, 2006.

[34] Rockefeller Foundation, *Scenarios for the Future of Technology and International Development*. New York, NY, USA: Rockefeller Foundation, 2010.

[35] H. Wang and A. Koonce, *What-If Analysis: A Practical Approach to Complex Problems*. Hoboken, NJ, USA: Wiley, 2011.

Appendix A

Table A1. Machine Learning Algorithms, Associated R packages and Hyper-parameters.

Machine learning Algorithm	R packages used	Hyperparameters
Generalized Linear Models (GLM)	glm2	family distribution: Gaussian, link function: Identity
Generalized Additive Models (GAM)	gam	family distribution: Gaussian, link function: Identity
Regression Trees (RPART)	Rpart	No hyperparameters used
Random Forest (RANGER)	Ranger	number of trees:500, Mtry:8, Target node size: 5
Extreme Gradient Boosting (XGB)	Xgboost	xgboost maximum depth: 3, number of rounds: 100
Least Angle Regression (LAR)	Caret	Method:lars
Support Vector Machines (SVMM)	E10171	SVM-Type: eps-regression,SVM-Kernel: radial,cost: 1
Artificial Neural Network (ANN)	nnet	Size:2, decay:5e-4, maximum iterations:200

Table A2. AALR Total Reserves.

Automated Actuarial Loss Reserving -Total Reserves

ML Model	IBNYR	RBNYS	REOPENED	REINSURANCE
GLM	2,560,788.00	480,147.80	128,039.40	32,009.85
GAM	2,556,430.00	479,330.70	127,821.50	31,955.38
RPART	2,559,975.00	479,995.20	127,998.70	31,999.68
RANGER	10,243,899.00	1,920,731.00	512,194.90	128,048.70
XGB	2,560,619.00	480,116.00	128,030.90	32,007.73
LAR	2,562,175.00	480,407.80	128,108.70	32,027.18
SVM	1,339,998.00	251,249.50	66,999.88	16,749.97
ANN	1,197,153.00	224,466.30	59,857.67	14,964.42

Table A3. AALR INBYR Reserves.

Automated Actuarial Loss Reserving -Total IBNYR Reserves

ML Model	IBNYR-A	IBNYR-B	IBNYR-C	IBNYR-D
GLM	1,280,394.00	768,236.40	512,157.60	0.00
GAM	1,278,215.00	766,929.00	511,286.00	0.00
RPART	1,279,987.50	767,992.50	511,995.00	0.00
RANGER	5,121,949.50	3,073,169.70	2,048,779.80	0.00
XGB	1,280,309.50	768,185.70	512,123.80	0.00
LAR	1,281,087.50	768,652.50	512,435.00	0.00
SVM	669,999.00	401,999.40	267,999.60	0.00
ANN	598,576.50	359,145.90	239,430.60	0.00

Table A4. AALR RBNYS Reserves.

Automated Actuarial Loss Reserving -Total RBNYS Reserves

ML Model	RBNYS-A	RBNYS-B	RBNYS-C	RBNYS-D
GLM	240,073.90	144,044.34	96,029.56	0.00
GAM	239,665.35	143,799.21	95,866.14	0.00
RPART	239,997.60	143,998.56	95,999.04	0.00
RANGER	960,365.50	576,219.30	384,146.20	0.00
XGB	240,058.00	144,034.80	96,023.20	0.00
LAR	240,203.90	144,122.34	96,081.56	0.00
SVM	125,624.75	75,374.85	50,249.90	0.00
ANN	112,233.15	67,339.89	44,893.26	0.00

Table A5. AALR REOPENED Reserves.

Automated Actuarial Loss Reserving -Total REOPENED Reserves

ML Model	REOPENED-A	REOPENED-B	REOPENED-C	REOPENED-D
GLM	64,019.70	38,411.82	25,607.88	0.00
GAM	63,910.75	38,346.45	25,564.30	0.00
RPART	63,999.35	38,399.61	25,599.74	0.00
RANGER	256,097.45	153,658.47	102,438.98	0.00
XGB	64,015.45	38,409.27	25,606.18	0.00
LAR	64,054.35	38,432.61	25,621.74	0.00
SVM	33,499.94	20,099.96	13,399.98	0.00
ANN	29,928.84	17,957.30	11,971.53	0.00

Table A6. AALR REINSURANCE Reserves.

Automated Actuarial Loss Reserving -Total REINSURANCE Reserves

ML Model	REINSURANCE-A	REINSURANCE-B	REINSURANCE-C	REINSURANCE-D
GLM	16,004.93	9,602.96	6,401.97	0.00
GAM	15,977.69	9,586.61	6,391.08	0.00
RPART	15,999.84	9,599.90	6,399.94	0.00
RANGER	64,024.35	38,414.61	25,609.74	0.00
XGB	16,003.87	9,602.32	6,401.55	0.00
LAR	16,013.59	9,608.15	6,405.44	0.00
SVM	8,374.99	5,024.99	3,349.99	0.00
ANN	7,482.21	4,489.33	2,992.88	0.00

Table A7. AALR COMPREEHENSIVE Reserves.

Automated ML Model	Actuarial Loss CAALR-A	Reserving -T CAALR-B	otal COMPREHENSIVE Reserves CAALR-C	CAALR-D
GLM	1,600,492.53	960,295.52	640,197.01	0.00
GAM	1,597,768.79	958,661.27	639,107.52	0.00
RPART	1,599,984.29	959,990.57	639,993.72	0.00
RANGER	6,402,436.80	3,841,462.08	2,560,974.72	0.00
XGB	1,600,386.82	960,232.09	640,154.73	0.00
LAR	1,601,359.34	960,815.60	640,543.74	0.00
SVM	837,498.68	502,499.21	334,999.47	0.00
ANN	748,220.70	448,932.42	299,288.28	0.00

Table A8. AALR AGGREGATE Reserves.

Actuarial Loss Reserving

ML Model	ACAALR
GLM	3,200,985.05
GAM	3,195,537.58
RPART	3,199,968.58
RANGER	12,804,873.60
XGB	3,200,773.63
LAR	3,202,718.68
SVM	1,674,997.35
ANN	1,496,441.39

Table A9. ULTIMATE RATIOS FOR CAALR.

Automated Actuarial Loss Reserving -Total Ultimate Ratios

ML Model	Category A	Category B	Category C	Category D
GLM	50%	30%	20%	0%
GAM	50%	30%	20%	0%
RPART	50%	30%	20%	0%
RANGER	50%	30%	20%	0%
XGB	50%	30%	20%	0%
LAR	50%	30%	20%	0%
SVM	50%	30%	20%	0%
ANN	50%	30%	20%	0%