



MATHURANCE

English Version

Introduction

Dans le secteur de l'assurance des risques industriels, l'estimation des réserves techniques est essentielle pour garantir la capacité des assureurs à indemniser les sinistres tout en maintenant leur stabilité financière. Ces réserves couvrent les sinistres déclarés mais non encore réglés ainsi que les sinistres tardifs, rendant leur évaluation complexe en raison de la variabilité des montants et des délais de règlement.

L'utilisation des modèles statistiques classiques, comme les triangles de liquidité (Run-Off Triangles), reste une approche de référence, mais l'évolution des risques et l'augmentation des données nécessitent des méthodes plus avancées. Ce projet vise à développer un modèle prédictif performant combinant mathématiques, processus stochastiques et intelligence artificielle (IA), tout en garantissant une interprétabilité et une applicabilité concrète pour les compagnies d'assurance.

Problem Statement

You are required to design a predictive model for technical reserves based on the provided dataset.

The model should be developed by leveraging advanced Mathematics, Statistics, and Stochastic Processes, while also integrating cutting-edge Artificial Intelligence (AI) techniques, including Machine Learning (ML) and Deep Learning (DL) models.

The goal of the model is to provide a reliable estimation of the liquidity triangle (Run-Off Triangle), enabling the anticipation of settled claims as well as outstanding late claims.

However, beyond predictive performance, the model must be interpretable and explainable to ensure that actuaries and reserve managers can understand the factors influencing the estimates. It is crucial that the results can be analyzed and justified using clear indicators, intuitive visualizations, and explanations of the model's decision-making process.

Moreover, the proposed solution should be practically applicable for insurance companies. It should be compatible with existing systems and seamlessly integrate into claims management and reserving processes. Special attention should be given to the model's robustness against data variations and its ability to adapt to changes in the insurance market.

Data Description

The database represents a sample of data from the Algerian insurance market. It contains **90,229 observations** (rows: claims) spanning a 10-year period from **2014 to 2023**. It records claim settlements for industrial risks, categorized by year of occurrence.

The columns in the dataset are as follows:

- **Exercice:** (Year of Settlement): The year in which claims were settled.
- **Branche:** (Insurance Branch): The branch corresponding to industrial risks.
- **Code Produit:** (Product Code): The codes of insurance products related to industrial risks.
- **Désignation Produit:** (Product Name): The name of the insurance products.
- **Sous-Branche:** (Sub-Branch): The subcategories within industrial risks, where each sub-branch includes multiple insurance products.
- **Date Survenance:** (Date of Occurrence): The date when the claim event occurred.
- **Règlement:** (Settlement Amount): The amount paid for each claim settlement.

Note: Claim patterns vary depending on the specific insurance product and sub-branch.



French Version

Introduction

In the industrial risk insurance sector, estimating technical reserves is crucial to ensuring that insurers can compensate claims while maintaining financial stability. These reserves account for reported but unpaid claims as well as late-emerging claims, making their assessment complex due to the variability in amounts and settlement delays.

Traditional statistical models, such as liquidity triangles (Run-Off Triangles), remain standard practice, but the evolving nature of risks and increasing data volumes call for more advanced methods. This project aims to develop an efficient predictive model that integrates mathematics, stochastic processes, and artificial intelligence (AI) while ensuring interpretability and practical applicability for insurance companies.

Problème

Vous devez concevoir un modèle prédictif des réserves techniques en exploitant les données fournies.

Le modèle devra être développé en combinant des méthodes avancées de Mathématiques, Statistiques et Processus Stochastiques, tout en intégrant des approches innovantes basées sur l'Intelligence Artificielle (IA), telles que les modèles de Machine Learning (ML) et Deep Learning (DL).

L'objectif du modèle est de fournir une estimation fiable du triangle de liquidité (Run-Off Triangle), permettant d'anticiper l'évolution des sinistres réglés ainsi que des sinistres tardifs non encore remboursés.

Cependant, au-delà de la performance prédictive, le modèle doit être interprétable et explicable afin de permettre aux actuaires et aux gestionnaires de réserves de comprendre les facteurs influençant les estimations. Il est essentiel que les résultats puissent être analysés et justifiés à l'aide d'indicateurs clairs, de visualisations compréhensibles et d'explications sur les décisions du modèle.

De plus, la solution développée doit être applicable dans un contexte opérationnel pour les compagnies d'assurance. Elle devra être compatible avec les systèmes existants et s'intégrer dans les processus de gestion des sinistres et de provisionnement. Une attention particulière devra être portée à la robustesse du modèle face aux variations des données et à sa capacité à s'adapter aux évolutions du marché assurantiel.

Description des Données

La Base de Données représente un échantillon de données du marché algérien des assurances. Elle comporte **90 229 observations** (lignes : sinistres) sur une période de 10 ans allant de **2014 à 2023**. Elle représente les règlements des sinistres des risques industriels par année de survenance

Les colonnes représentent:

- **Exercice:** représente le années de règlements des sinistres
- **Branche:** correspond à la branche des risques industriels
- **Code Produit:** les codes des produits d'assurance afférents à la branche des risques industriels
- **Désignation Produit:** le nom des produits
- **Sous-Branche:** relative aux sous branches des risques industriels, chaque sous-branche comporte plusieurs produits d'assurance
- **Date Survenance:** correspond à la date de survenance du sinistre
- **Règlement:** le montant de règlement de chaque sinistre

NB : Le comportement de la sinistralité varie en fonction du produit et de la sous-branche concernée.