



المؤتمر العربي للإكتواريين ٢٠٢٥ ARAB ACTUARIAL CONFERENCE 2025



Panel « Delayed Claims and Risk Pricing »

Intervenants : Kamel Debiche (BST), Najed ksouri (ERM Partners).



Co Organizers



المكتب المتخصص بالتعريف في مجال التأمينات
Bureau Spécialisé de Tarification en Assurances

IBNR et Machine Learning : défis et perspectives



Alger, Algérie – Juillet 2025

Kamel Debiche
Actuaire, Chef d'études (BST)



- Le Machine Learning intéresse de plus en plus les assureurs pour mieux prévoir les provisions et réduire les erreurs.
- Il permet d'analyser les données plus en détail qu'avec les méthodes classiques.

Problématique :

- Les méthodes traditionnelles perdent des infos importantes.
- Elles ne captent pas bien la complexité des sinistres.
- Il faut choisir entre simplicité et précision

Pourquoi le ML ?

- Détecter plus tôt les tendances cachées et anomalies.
- Réduire l'erreur moyenne (MAE).

Machine Learning in P&C Insurance: A Review for Pricing and Reserving (Blier-Wong et al., 2020)

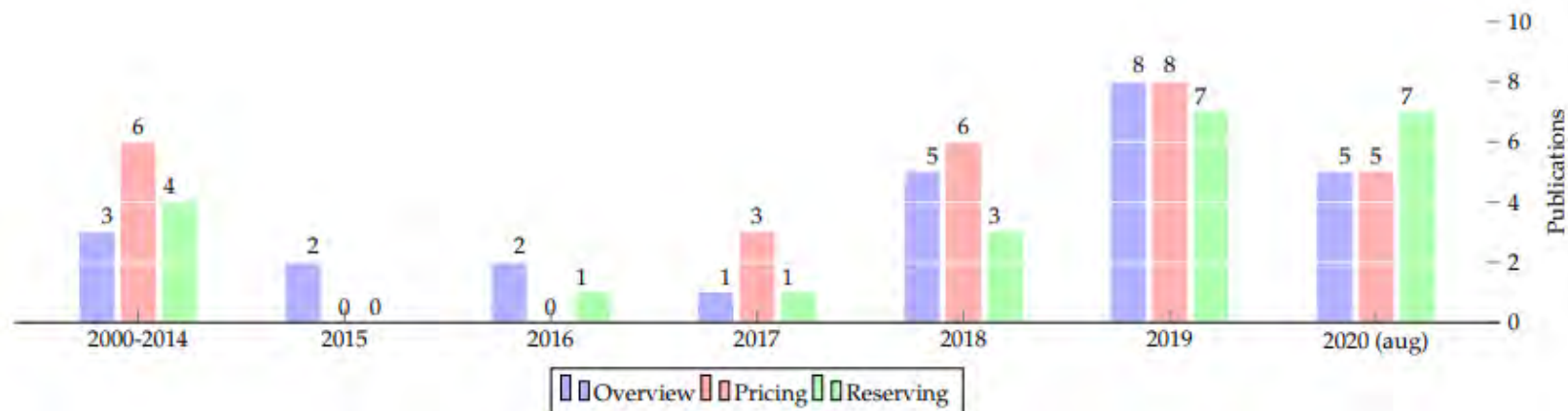


Figure 1. Number of publications per year.

Étude comparative : Méthode traditionnelle vs Machine Learning



المكتب المتخصص في الترخيف في مجال التأمينات
Bureau Spécialisé de Tarification en Assurances

Objectif : Comparer l'approche traditionnelle Mack/Chain-Ladder à des méthodes Machine Learning pour estimer les provisions IBNR.

Méthode traditionnelle (classique) : Modèle Mack / Chain-Ladder

- Méthode stochastique linéaire, fondée sur la continuité des facteurs de développement.
- Hypothèses fortes : homogénéité temporelle, absence d'effets calendaires ou non-linéaires.

Approche Machine Learning : Modèles supervisés adaptés aux triangles de sinistres.

GLM régularisé : (*Penalized GLM*)

- LASSO (pénalisation L1)
- Ridge (pénalisation L2)

XGBoost : (Gradient Boosted Decision Trees)

- Régression Tweedie

En million de DA

Année d'origine	Année de développement												
	1	2	3	4	5	6	7	8	9	10	11	12	13
2010	1 749	4 159	6 542	7 857	8 609	9 544	9 901	10 020	10 101	10 160	10 187	10 227	10 256
2011	2 122	6 238	8 613	9 980	10 892	12 628	12 959	13 089	13 205	13 263	13 330	13 385	
2012	2 319	6 604	9 505	11 383	13 311	14 934	15 203	15 561	15 696	15 819	15 887		
2013	2 586	7 048	10 391	12 313	15 658	16 731	17 045	17 199	17 384	17 499			
2014	2 934	8 057	11 406	13 339	17 252	18 319	18 561	18 820	18 984				
2015	3 065	8 006	11 290	13 155	17 557	18 313	18 755	19 038					
2016	2 779	7 218	10 456	12 448	13 530	14 841	15 333						
2017	2 834	7 470	10 553	11 991	13 298	14 421							
2018	3 133	7 696	9 887	11 867	13 069								
2019	3 328	6 472	9 505	11 506									
2020	2 018	4 932	6 993										
2021	3 076	6 368											
2022	2 766												

- Triangle de règlements cumulés au titre de la garantie RC AUTO-

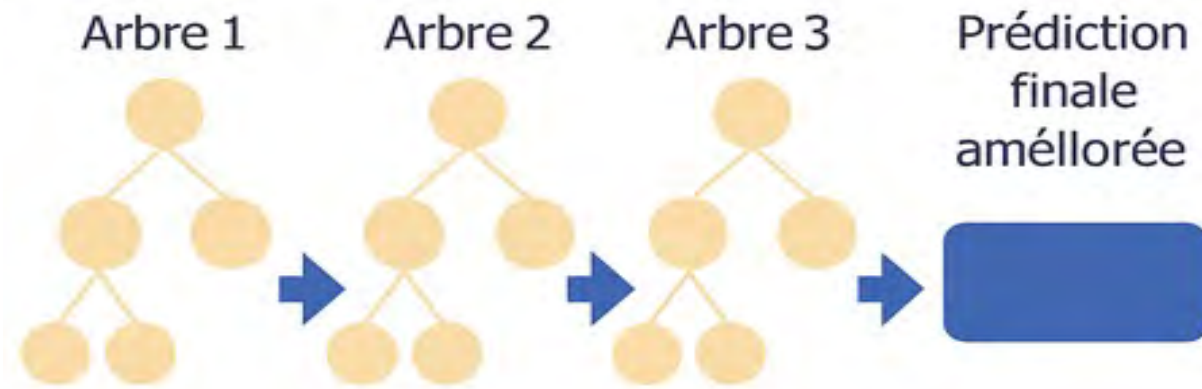
LASSO & Ridge : Extensions du modèle GLM

- **GLM (Generalized Linear Model)** : outil classique pour modéliser les triangles de sinistres.
- **Ridge & LASSO** : mêmes principes que GLM, mais ajoutent une pénalisation :
- **Ridge Regression (L2)** : réduit les grands coefficients mais les conserve tous.
- **LASSO Regression (L1)** : certains coefficients = 0 → sélection automatique de variables.
- **Avantages** : stabilité, robustesse, prévention du surapprentissage.
- **Application IBNR** : ajuster la réserve en contrôlant la complexité du modèle

XGBoost : eXtreme Gradient Boosting



المكتب المتخصص في التعريفة في مجال التأمينات
Bureau Spécialisé de Tarification en Assurances



- **Famille** : Apprentissage supervisé, GBM (Gradient Boosting Machines).
- **Principe** : enchaîner plusieurs arbres pour corriger les erreurs → modèle puissant.
- **Pourquoi** : capte relations non linéaires, interactions complexes.
- **Cas IBNR** : prédire incréments futurs du triangle → somme → provision.
- **Distribution Tweedie** : adaptée aux données de sinistres.

Calibrage et Paramétrage des Modèles

- **Ridge / LASSO :** lambda choisi par cross-validation
 - $\alpha = 0 \rightarrow$ Ridge.
 - $\alpha = 1 \rightarrow$ LASSO.
- **XGBoost :**
 - eta (learning rate) $\rightarrow$ règle la vitesse $\rightarrow$ ex. 0.1.
 - max_depth $\rightarrow$ contrôle la complexité.
 - objective $\rightarrow$ reg:tweedie pour pertes.
- **Validation croisée :**
 - Diviser train/test.
 - Vérifier MAE & RMSE.
 - Éviter le surapprentissage.

Résultats comparatifs :

Méthode	Provision estimée (Millions)	Erreur	Ratio Erreur / Provision	RMSE
Mack (Chain Ladder)	34 814	MSEP = 4 173	12 %	
Ridge	30 406	MAE = 579	1,9 %	769
LASSO	24 331	MAE = 645	2,6 %	833
XGBoost (Tweedie)	35 109	MAE = 126	0,36 %	205

Les modèles de **Machine Learning** (Ridge, LASSO, XGBoost) affichent une erreur plus faible que la méthode traditionnelle **Mack**, confirmant une meilleure capacité de prédiction. **XGBoost** ressort comme le modèle ayant donné le **meilleur résultat**, grâce à sa puissance pour capter des relations non linéaires et interactions complexes.

Conclusion :

- L'intégration du **Machine Learning** ouvre de nouvelles perspectives pour améliorer la précision des provisions IBNR, même à partir de données agrégées.
- Bien que des limites existent (données parfois insuffisamment détaillées, nécessité d'une bonne calibration), le potentiel est considérable pour compléter et affiner les méthodes traditionnelles.
- Des méthodes comme XGBoost montrent que le Machine Learning exploite des relations non linéaires et peut ainsi devenir un véritable levier d'innovation actuarielle, à condition de s'appuyer sur des données fiables et une interprétation maîtrisée.



ARAB ACTUARIAL CONFERENCE 2025

***ENJEUX DU
MACHINE LEARNING
EN PROVISIONNEMENT***

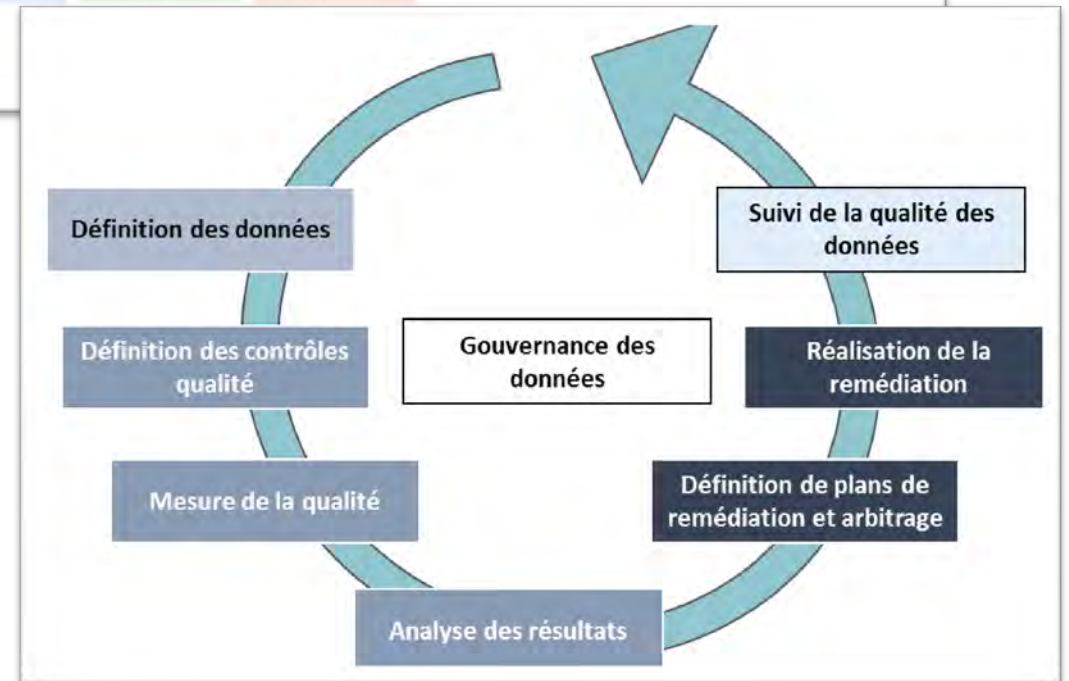
Enjeu 1 : Qualité des données

❖ La qualité de données fait dorénavant l'objet d'une **politique écrite**. Pourquoi ? Comment ?

❖ Le provisionnement avec Machine Learning **résiste il bien à une mauvaise QDD ?**

❖ Généralement **NON** (Un historique riche et complet est nécessaire)

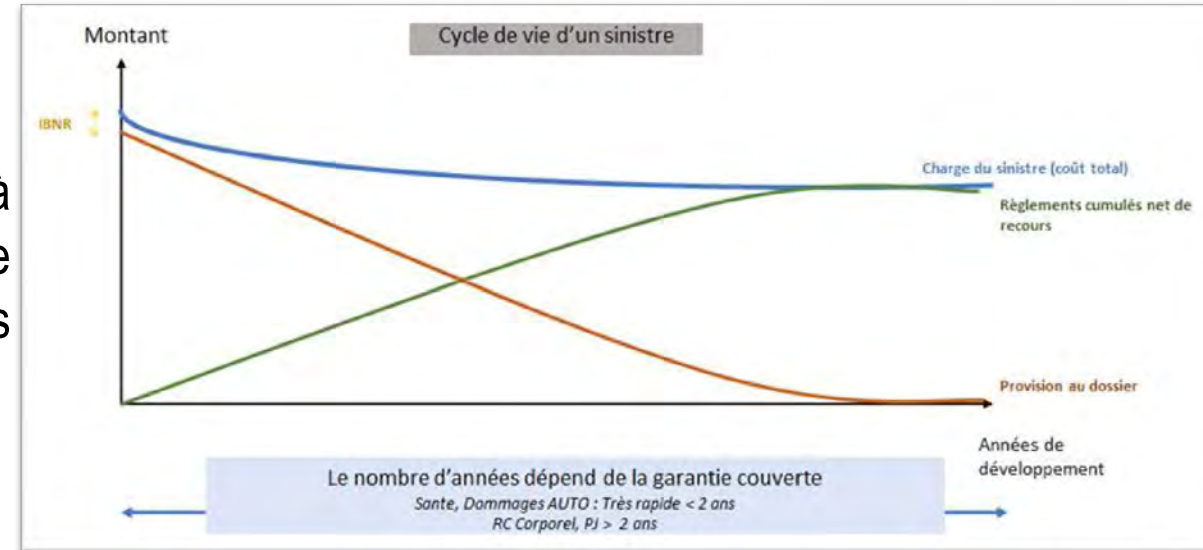
❖ Que faire ? **Effectuer des missions de fiabilisation de données en amont et en aval de la collecte.**



Enjeu 2 : Ruptures de gestion

❖ Qu'est ce qu'une rupture de gestion ?

- Un changement dans le cycle de vie des sinistres dû à des phénomènes internes (nouvelle politique de souscription, nouveaux process ou SI, ...) ou externes (inflation, changement réglementaire, ...)



❖ Le provisionnement avec Machine Learning résistent ils bien aux ruptures de gestion?

- ❖ Souvent **NON** (→ Passage aux méthodes de crédibilité)
- ❖ Parfois **OUI** si le modèle est bien enrichi par un historique de ruptures similaires



Enjeu 3 : Traitement des dérives/extrêmes

☀ Qu'est ce qu'un sinistre **Extrême** ? Grave ? Attritionnel ?

☀ Qu'est ce qu'une **dérive** de sinistralité ? Fréquence ?

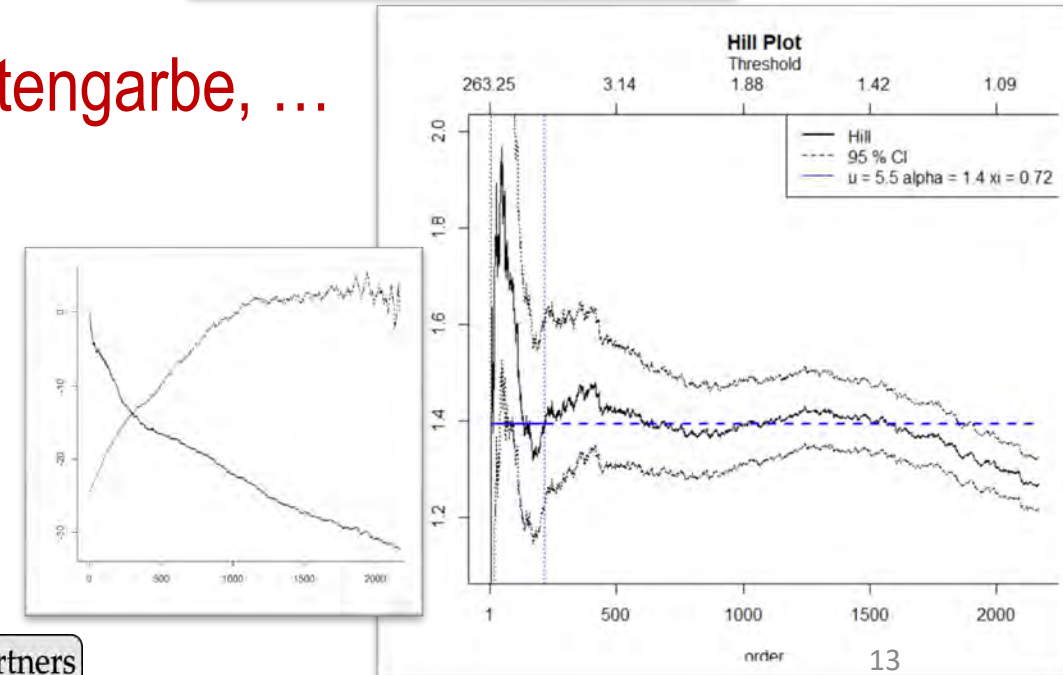
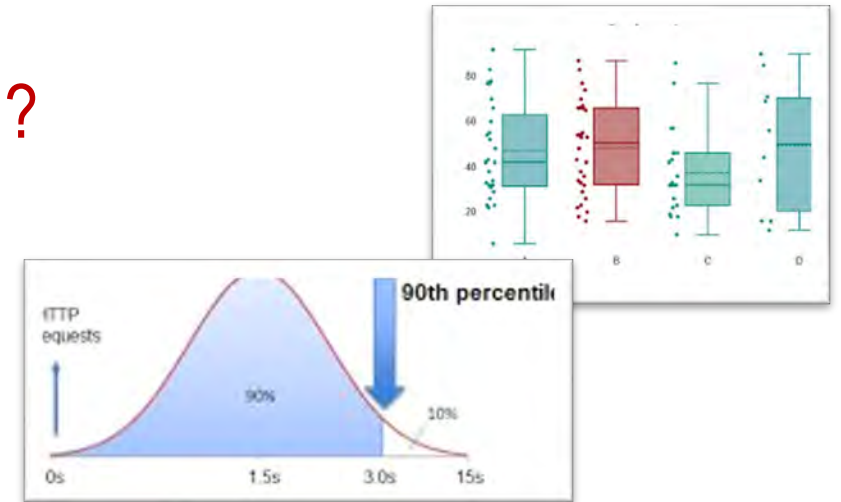
☀ Comment les visualiser ? **Histogramme**, **Boxplot**, ...

☀ Comment les détecter ? **Estimateur de Hill**, **Gerstengarbe**, ...

☀ Comment les provisionner ? Sont ils réassurés ?

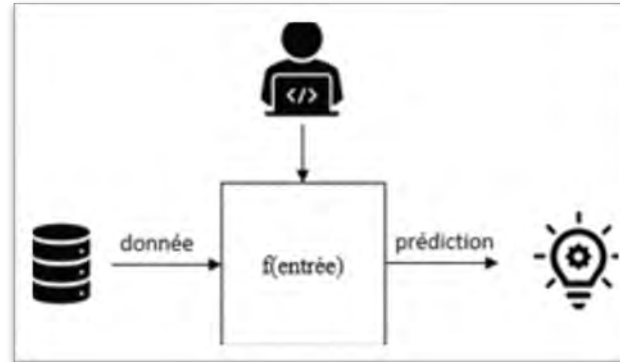
☀ **Non réassurés : Par égalisation**

☀ **Réassurés : Par marge de sécurité**

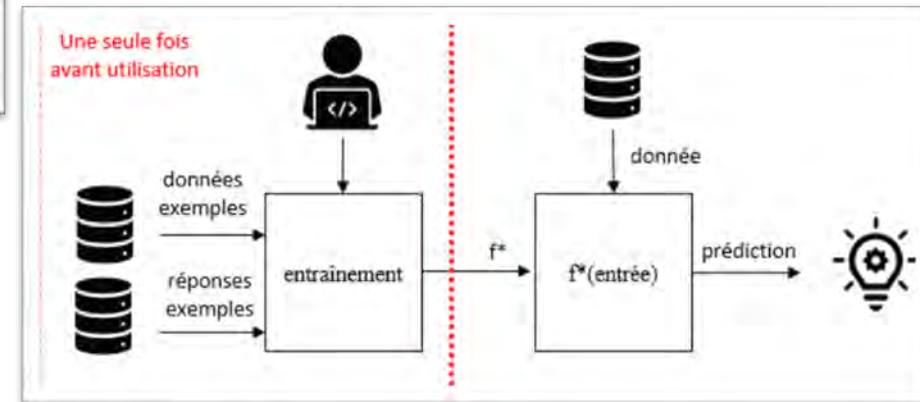


Enjeu 4 : Transparence des algorithmes

❑ Quel **confiance** donner aux algorithmes ?



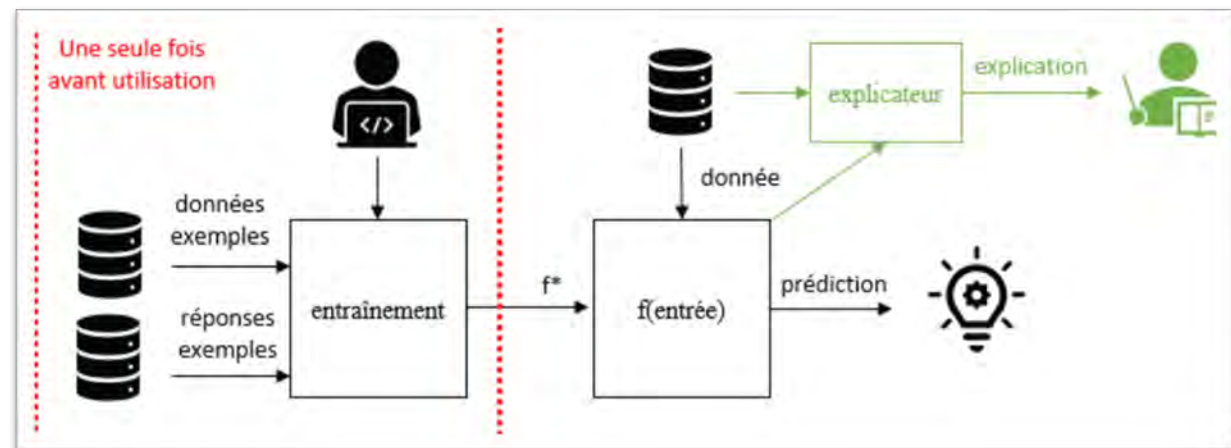
❑ Qui porte la **responsabilité** des résultats ?



❑ Quelle solution ?

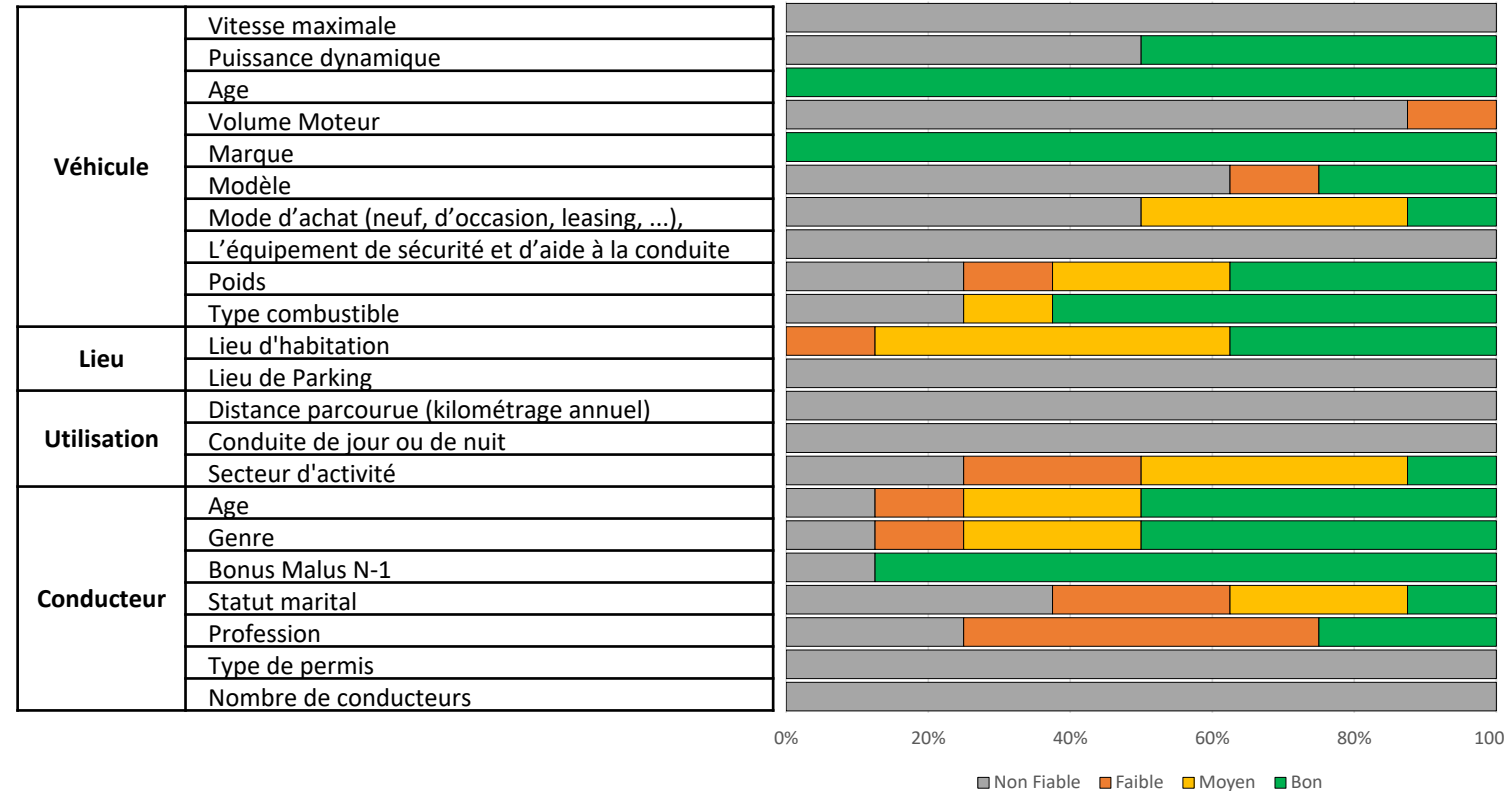
❑ Les explicateurs ?

❑ GLM - Machine Learning driven ?



Enjeu 5 : Cohérence avec la tarification

- ☐ Le modèle de provisionnement doit il coïncider avec celui du coût de la Tarification ?
- ☐ En théorie **OUI** mais pas forcément (un décalage d'un an existe déjà)
- ☐ Un tarif comprend parfois des considérations sociales et politiques. (Exemple : Zonier)

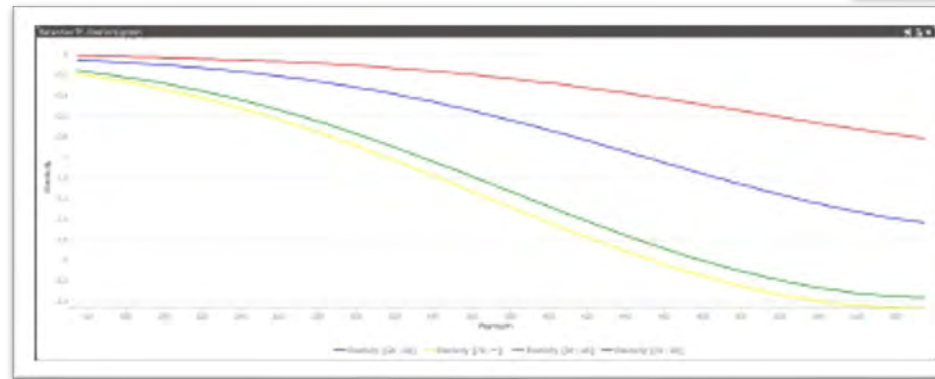


ML et tarification

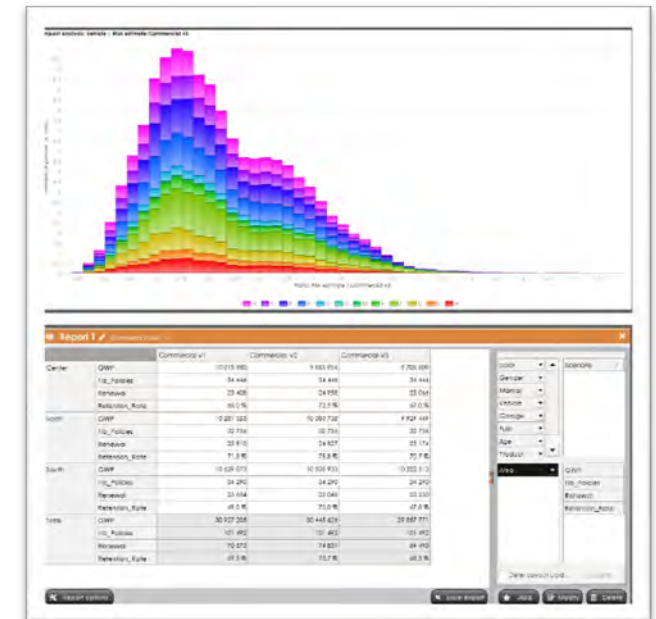
❑ Tarif central et zoniers



❑ Comportement de l'assuré
et son élasticité prix



❑ Profit et stress testing





المؤتمر العربي للإكتواريين ٢٠٢٥ ARAB ACTUARIAL CONFERENCE 2025



Merci pour votre attention !



Co Organizers



المكتب المتخصص بالتعريفية في مجال التأمينات
Bureau Spécialisé de Tarification en Assurances