

Tarification en assurance santé à l'ère du machine learning : une approche actuarielle hybride

Health insurance pricing in the era of machine learning: A hybrid actuarial approach

Ayoub Laaroussi^{1*}, Ali Ouchekkir Ali¹, Fatima zahrae El Mouden¹

¹ Faculty Of Legal, Economic And Social Sciences Agdal, Mohamed V University, Rabat, Morocco

*Corresponding author: ayoub_laaroussi3@um5.ac.ma

Résumé

La tarification en assurance santé constitue un enjeu majeur pour l'équité tarifaire, la soutenabilité financière des portefeuilles et la performance des assureurs. Elle repose traditionnellement sur le cadre actuariel fréquence–sévérité, généralement implémenté à l'aide de modèles linéaires généralisés (GLM). Toutefois, la complexité croissante des données de sinistres met en évidence certaines limites des approches paramétriques classiques. Cet article propose un cadre méthodologique hybride combinant modèles actuariels traditionnels et techniques de machine learning pour estimer le coût individuel attendu des sinistres en assurance santé. Plusieurs approches paramétriques et non paramétriques sont comparées à partir d'un jeu de données de sinistres. Les résultats montrent que les méthodes de machine learning peuvent améliorer la performance prédictive tout en soulignant l'importance de l'interprétabilité et de la gouvernance des modèles dans un environnement assurantiel réglementé.

Mots-clés : Tarification de l'assurance santé ; Modèles fréquence–sévérité ; Apprentissage automatique ; Modèles actuariels ; Intelligence artificielle explicable ; Gouvernance des modèles ; AssurTech.

Abstract

Health insurance pricing plays a key role in ensuring tariff fairness, portfolio sustainability and insurers' financial performance. Traditionally, pricing relies on the actuarial frequency–severity framework implemented through generalized linear models (GLM), valued for their robustness and interpretability. However, the increasing complexity of claims data challenges the ability of purely parametric models to capture nonlinear relationships and interactions between risk factors. This study proposes a hybrid framework combining traditional actuarial models with machine learning techniques to estimate individual expected claim costs in health insurance. Using a claims dataset, several parametric and non-parametric models are compared, including ensemble approaches. The results indicate that machine learning methods can improve predictive performance, particularly in terms of discrimination and probabilistic accuracy, while highlighting the importance of interpretability and model governance in regulated insurance environments.

Keywords: Health insurance pricing; Frequency–severity models; Machine learning; Actuarial models; Explainable AI; Model governance; InsurTech



Copyright © 2025 The Author(s). Published by IRAFEM.

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

La tarification en assurance santé joue un rôle central dans la gestion des risques et la soutenabilité financière des portefeuilles d'assurance. Elle consiste à estimer le coût attendu des sinistres futurs à partir des caractéristiques observables des assurés afin de concilier équité tarifaire, compétitivité commerciale et équilibre technique. Dans la pratique actuarielle, cette estimation repose traditionnellement sur le cadre fréquence–sévérité, généralement mis en œuvre à l'aide de modèles linéaires généralisés (GLM), appréciés pour leur robustesse statistique et leur interprétabilité. Toutefois, l'augmentation du volume et de la complexité des données issues des systèmes d'information met en évidence certaines limites des approches paramétriques classiques, notamment pour modéliser des relations non linéaires et des interactions complexes entre facteurs de risque. Dans ce contexte, les techniques d'apprentissage automatique (Machine Learning) suscitent un intérêt croissant dans la littérature actuarielle en raison de leur capacité à exploiter des jeux de données riches et à améliorer la précision des prédictions de sinistralité et de coûts (Kapse et al., 2025). Par ailleurs, des approches récentes cherchent à concilier performance prédictive et interprétabilité, notamment à travers des modèles neuronaux additifs adaptés au pricing assurantiel (Laub, Pho & Wong, 2025).

Malgré ces avancées, l'adoption opérationnelle du machine learning dans le secteur de l'assurance demeure limitée en raison d'exigences fortes en matière d'interprétabilité, de gouvernance des modèles et de conformité réglementaire. Dans les environnements régulés, les modèles doivent en effet être explicables et auditables afin de justifier les décisions tarifaires auprès des fonctions de contrôle interne et des autorités de supervision (EIOPA, 2025). L'essor des approches d'Explainable AI (XAI) répond précisément à cet enjeu en visant à rendre les décisions des modèles complexes plus transparentes (Wikipedia, 2026). Dans ce contexte, cet article propose un cadre méthodologique hybride combinant le modèle actuariel fréquence–sévérité et des techniques de machine learning afin d'estimer le coût individuel attendu des sinistres en assurance santé. L'étude vise à analyser les performances prédictives de différentes approches et à discuter les arbitrages entre performance, interprétabilité et gouvernance des modèles dans la pratique actuarielle.

2. Revue de littérature

2.1. Tarification actuarielle en assurance santé

La tarification actuarielle traditionnelle repose sur la modélisation statistique des sinistres selon une décomposition fréquence–sévérité : la fréquence représente le nombre attendu d'événements de sinistre pour un assuré sur une période donnée, tandis que la sévérité correspond au coût moyen (ou à la distribution des montants) conditionnellement à la survenance d'un sinistre. Ce cadre structure les approches de *ratemaking* en assurance, en permettant d'estimer un coût technique moyen par unité d'exposition, puis de construire un tarif cohérent avec les objectifs de mutualisation, de segmentation et d'équilibre technique. Dans la pratique, cette décomposition est fréquemment mise en œuvre via des modèles linéaires généralisés (GLM), qui offrent une base statistique solide et une interprétabilité appréciée des fonctions actuarielle, audit et conformité (Goldburd et al., 2020). Les GLM se sont imposés comme référence opérationnelle en tarification notamment grâce à leur structure multiplicative et à la lisibilité des effets des facteurs de risque, facilitant la justification technique des décisions tarifaires auprès des parties prenantes internes et externes. Cette robustesse explique leur diffusion durable dans les pratiques actuarielles, y compris dans des dispositifs de tarification modernisés (Goldburd et al., 2020). Cependant, la littérature

souligne que les modèles paramétriques classiques peuvent se révéler moins performants lorsque les données deviennent riches, hétérogènes et fortement interactionnelles, situation fréquente en assurance santé (effets combinés de variables sociodémographiques, patterns de recours aux soins, et hétérogénéité médicale). Les limites apparaissent notamment lorsque les relations sous-jacentes sont non linéaires ou de grande dimension, ce qui ouvre la voie à des approches plus flexibles (Holvoet al., 2023).

2.2. Machine learning en tarification

L'usage des techniques de machine learning en tarification s'est intensifié ces dernières années, avec pour objectif d'améliorer la précision prédictive au-delà des modèles linéaires, en capturant des non-linéarités, interactions et effets de seuil difficiles à représenter dans un cadre GLM standard. Les approches d'ensemble, notamment Random Forest et Gradient Boosting (XGBoost/LightGBM/CatBoost), ainsi que certaines architectures de deep learning, figurent parmi les méthodes les plus étudiées et les plus utilisées dans les travaux appliqués (Holvoet al., 2023). Plusieurs études comparatives récentes confirment que dans différents contextes assurantiels, ces méthodes peuvent dépasser les modèles paramétriques sur des critères de performance prédictive (erreurs, discrimination, stabilité), tout en montrant que les résultats dépendent fortement de la qualité des données, de la sélection des variables, du protocole de validation et de la gouvernance du modèle (Holvoet al., 2023). Néanmoins, la littérature met également en évidence une tension structurante : l'augmentation de performance prédictive s'accompagne souvent d'une perte d'interprétabilité intrinsèque, rendant la justification des décisions tarifaires plus complexe pour les décideurs non techniques et, surtout, plus sensible dans des environnements réglementés. Cette tension explique que la question ne se limite plus au "meilleur modèle", mais s'élargit vers des arbitrages entre performance, transparence et contrôle du risque de modèle (Aas et al., 2024).

2.3. Gouvernance, explicabilité et exigences réglementaires

L'adoption de modèles d'apprentissage automatique en assurance s'inscrit dans un cadre où la gouvernance des modèles devient une dimension aussi critique que la performance prédictive. Les exigences portent sur l'explicabilité, l'auditabilité, la robustesse, la gestion des biais, la documentation, et la capacité à surveiller les modèles dans le temps (dérive, recalibrage, ré-entraînement). Dans la littérature actuarielle récente, l'assurance est explicitement décrite comme un domaine où la prédiction doit être conciliée avec l'explicabilité et l'équité, compte tenu des impacts économiques et sociétaux des décisions (segmentation, accessibilité, discrimination) (Aas et al., 2024). Sur le plan institutionnel, les superviseurs européens ont récemment renforcé les attentes relatives à l'encadrement des usages de l'IA dans l'assurance. L'Opinion d'EIOPA (2025) clarifie les principes de gouvernance et de gestion des risques applicables aux systèmes d'IA dans le secteur assurantiel, en insistant sur une approche proportionnée et fondée sur les risques, et sur l'alignement avec les exigences existantes (Solvabilité II, distribution, protection des données, etc.) (EIOPA, 2025). Plus largement, le contexte réglementaire européen évolue avec l'entrée en vigueur du Règlement (UE) 2024/1689 (AI Act), qui établit des obligations spécifiques pour certains systèmes d'IA, notamment en matière de gestion des risques, de qualité des données, de traçabilité et de transparence (European Parliament & Council of the European Union, 2024). En parallèle, des cadres de référence internationaux, comme le NIST AI Risk Management Framework (AI RMF 1.0), structurent la gestion des risques IA sur tout le cycle de vie (conception, déploiement, suivi), en mettant l'accent sur la fiabilité, la transparence, la gouvernance et la responsabilité (NIST, 2023). Dans ce contexte, la littérature converge vers l'idée que l'intégration du ML en tarification nécessite des dispositifs d'Explainable AI (XAI) et des

mécanismes de gouvernance continus : la valeur des modèles ne se réduit pas à la précision, mais inclut leur capacité à être expliqués, audités et maîtrisés dans le temps, condition essentielle de leur acceptabilité opérationnelle et réglementaire (Aas et al., 2024).

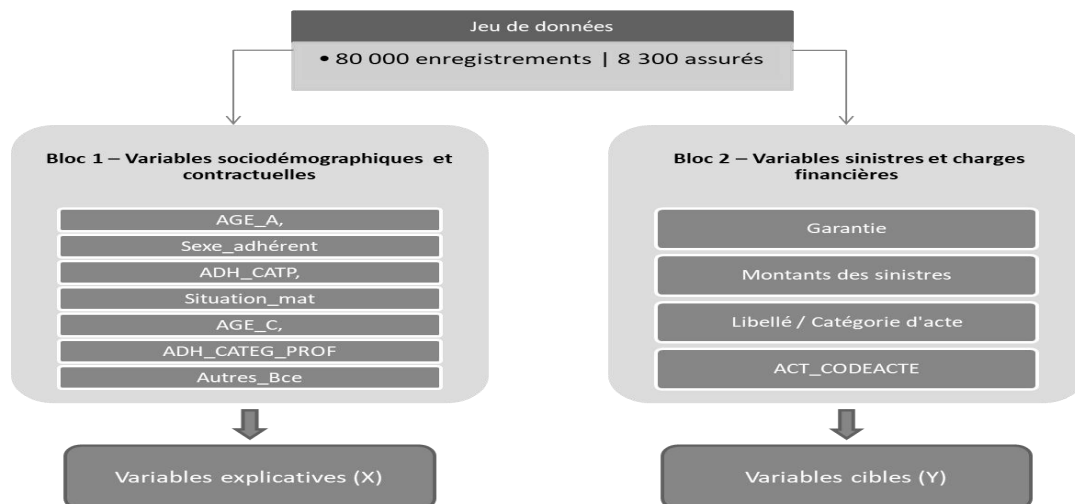
3. Méthodologie

La présente section s'inscrit dans une démarche exploratoire visant à analyser la faisabilité, la pertinence managériale et les enjeux de gouvernance liés à l'intégration de différentes familles de modèles actuariels et de machine learning dans un processus de tarification en assurance santé. Elle présente successivement (i) la structure des jeux de données mobilisés, (ii) la modélisation des probabilités de sinistre, (iii) la modélisation des charges, et (iv) l'intégration des deux composantes pour le calcul du coût individuel estimé. (Sinha, Desai & Mahal, 2014)

3.1. Description des jeux de données utilisés

Dans cette étude, nous exploitons un jeu de données comprenant plus de 80 000 enregistrements relatifs à l'historique des sinistres de 8 300 assurés. La structure des données s'organise autour de deux blocs complémentaires (Figure 1);

Figure 1 : Structure des données individuelles utilisées pour la modélisation



3.1.1. Données sociodémographiques et contractuelles

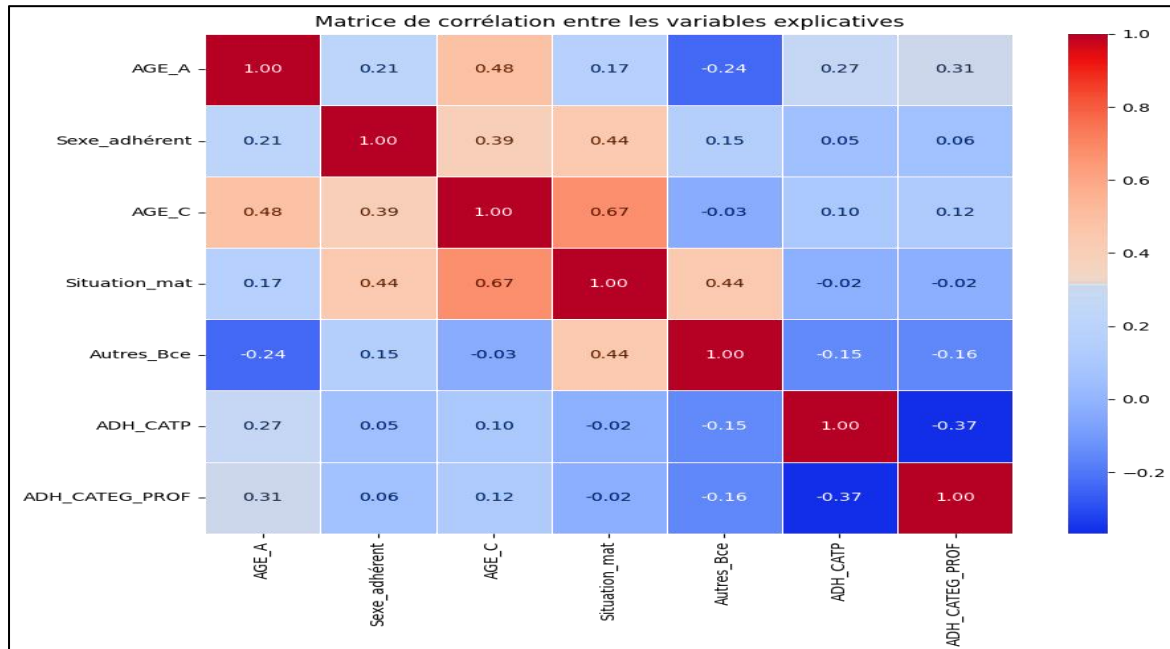
Le premier bloc de variables regroupe les caractéristiques sociodémographiques et contractuelles des assurés. Ces variables constituent les principaux déterminants observables mobilisés en tarification pour capturer l'hétérogénéité des comportements de recours aux soins et des niveaux de risque individuels. Elles comprennent notamment :

- AGE_A : âge de l'adhérent
- Sexe_adhérent : sexe de l'adhérent
- AGE_C : âge du conjoint
- Situation_mat : situation matrimoniale
- Autres_Bce : présence de bénéficiaires complémentaires
- ADH_CATP, ADH_CATEG_PROF : catégories de police et catégorie professionnelle

Ces variables sont utilisées comme variables explicatives dans les modèles afin d'estimer leur influence sur la probabilité de sinistre et le niveau des dépenses de santé, conformément aux pratiques classiques de tarification actuarielle basées sur la segmentation des risques. Afin d'examiner les relations entre ces variables, une matrice de corrélation a été construite (Figure

2). Celle-ci permet d'évaluer l'intensité des relations linéaires entre les variables explicatives et de détecter d'éventuels problèmes de multicolinéarité.

Figure 2 : Matrice de corrélation entre les variables explicatives



L'analyse de la matrice de corrélation (Figure 2) met en évidence des corrélations globalement faibles à modérées entre les variables explicatives. Les relations les plus marquées concernent AGE_A et AGE_C (0,48), reflétant la proximité d'âge au sein des couples, ainsi que AGE_C et Situation_mat (0,67), liée à la présence d'un conjoint. Globalement, aucune corrélation très élevée n'est observée, ce qui suggère l'absence de multicolinéarité forte et confirme la pertinence de l'utilisation conjointe de ces variables dans les modèles.

3.1.2. Données sinistres et charges financières

Le second bloc décrit l'activité de sinistre et les charges associées. Il comprend notamment :

- ACT_CODEACTE : identifiant (code) de l'acte sinistré
- Libellé / catégorie d'acte : information descriptive permettant de regrouper les actes par nature médicale
- Montants des sinistres : charges financières observées (dépenses)

Ce niveau de granularité (acte + montant) est central en assurance santé, car il permet (i) de relier le coût à un contenu médical, (ii) de produire des indicateurs par type d'acte, et (iii) d'améliorer la cohérence métier des résultats (pilotage médical, contrôle, prévention) (Buckle, Biscaglia & Silveiro, 2024).

3.1.3. Table de correspondance des actes et construction des variables de charge

Afin d'améliorer l'interprétation et d'assurer la cohérence du traitement des actes, nous mobilisons une table de correspondance associant, pour chaque code acte :

- le libellé de l'acte,
- la catégorie médicale correspondante,

Ce dictionnaire joue un rôle méthodologique clé : il permet (i) de standardiser les regroupements d’actes, (ii) de garantir la traçabilité des transformations de données (de l’acte brut à la variable modélisée), et (iii) de renforcer l’auditabilité du pipeline, ce qui devient particulièrement important lorsque les modèles prédictifs s’inscrivent dans un dispositif de tarification numérique. (EIOPA, 2025)

Tableau 1 : Table de correspondance des codes actes et catégorie médicale

Code Acte	Libellé de l'acte	Catégorie médicale	ACT_CODEACTE
PH-001	Consultation médecin généraliste	Consultation	PH
OPT-001	Monture + verres correcteurs	Optique	OPTIQ
ORT-001	Appareil orthodontique	Orthodontie	ORTHOP
CON-001	Consultation spécialiste	Consultation spc.	CONS
INF-001	Hospitalisation médicale	Hospitalisation	HOSP
OPH-001	Consultation ophtalmologique	Ophtalmologie	OPHTALMO
NEU-001	IRM cérébrale	Neurologie	NEURO

3.2. Modélisation des probabilités de sinistre

La première composante du cadre fréquence–sévérité consiste à estimer, pour chaque assuré, la probabilité d’occurrence d’un sinistre sur la période d’observation. Compte tenu de la structure des données (indicateurs binaires par catégorie d’acte/poste), cette étape est formulée comme un problème de classification supervisée.

3.2.1 Définition de la variable cible

Deux niveaux de variables cibles sont définis afin de répondre à des objectifs complémentaires de tarification et de pilotage médical :

- Cible globale (binaire) : ($Y=1$) si l’assuré présente au moins un sinistre (tous postes confondus) sur la période d’observation, et ($Y=0$) sinon.
- Cibles par poste (multi-label) : pour chaque poste de soins (k) (p. ex. biologie, dentaire, hospitalisation), on définit ($Y_k=1$) si au moins un acte relevant du poste (k) est observé sur la période, et ($Y_k=0$) sinon.

Cette formulation permet d’estimer des probabilités spécifiques par poste, utiles pour la tarification technique et l’analyse des déterminants du recours aux soins. Dans la pratique, l’estimation peut être réalisée par apprentissage un-contre-tous (un modèle par poste) ou via des modèles multi-output, selon le choix d’implémentation.

Tableau 2: Structure des variables cibles pour la modélisation des probabilités

ID Assuré	Actes observés sur la période	BIO	DEN	HOSP	KIN	Sinistre global (Y)
A001	Consultation + analyse sanguine	1	0	0	0	1
A002	Aucun acte	0	0	0	0	0
A003	Soins dentaires + kinésithérapie	0	1	0	1	1
A004	Hospitalisation	0	0	1	0	1
...
A8300	Hospitalisation	0	0	1	0	1

3.2.2. Modèles individuels

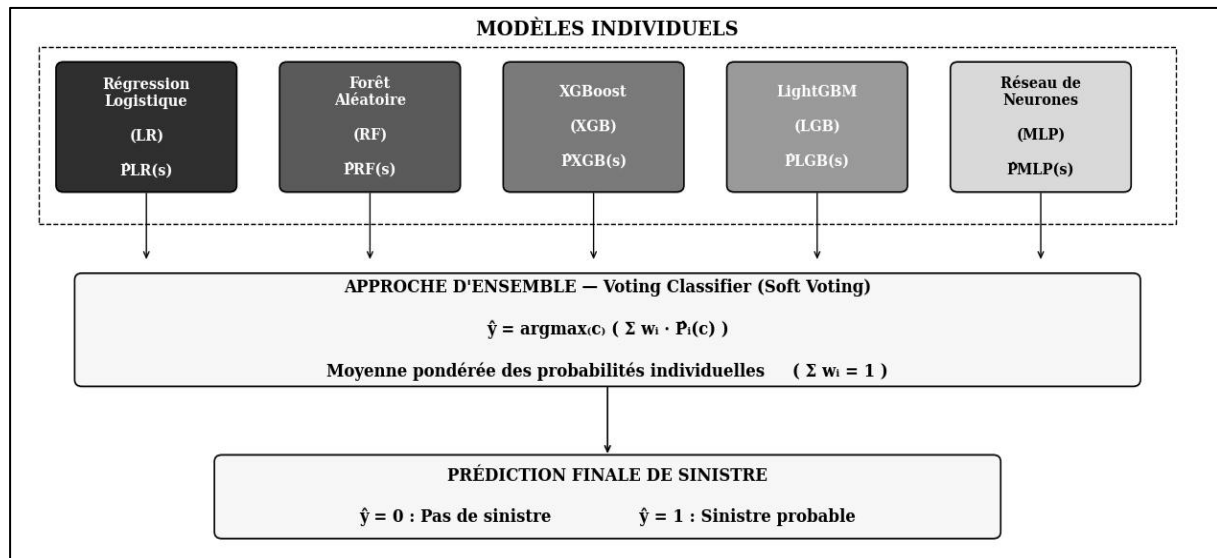
Plusieurs familles de classifieurs sont mobilisées afin de comparer (i) un modèle de référence interprétable, et (ii) des modèles non paramétriques capables de capturer non-linéarités et interactions :

- Régression logistique : référence paramétrique, interprétable, adaptée aux problématiques de scoring.
- Forêt aléatoire (Random Forest) : modèle d'ensemble par bagging, performant et robuste aux interactions.
- Gradient Boosting (XGBoost / LightGBM) : boosting séquentiel optimisant une fonction de perte, généralement performant sur données tabulaires.
- Réseau de neurones (MLP) : modèle flexible pour représenter des relations non linéaires complexes.

3.2.3. Stratégie d'apprentissage ensembliste (soft voting)

Afin d'améliorer la robustesse et la stabilité des probabilités prédites, une approche d'ensemble de type soft voting est mise en œuvre. Cette méthode consiste à combiner les probabilités estimées par plusieurs modèles de base afin d'obtenir une prédiction finale plus stable que celle produite par un modèle unique. Le principe repose sur l'agrégation des probabilités prédites (moyenne simple ou pondérée), permettant d'exploiter la complémentarité entre modèles paramétriques interprétables et modèles non paramétriques capables de capturer des non-linéarités et interactions complexes (Figure 3).

Figure 3 : Architecture de l'approche d'ensemble par Soft Voting



L'efficacité de cette approche dépend notamment de la diversité des modèles de base, de la calibration des probabilités et de la validation hors-échantillon. Dans cette étude, la performance du modèle d'ensemble est évaluée à l'aide d'un découpage apprentissage/test complété par une validation croisée. En pratique, cette stratégie est implémentée via le VotingClassifier de scikit-learn en mode voting = "soft", qui agrège les probabilités issues des différents modèles pour produire la prédiction finale.

3.2.4. Validation et métriques d'évaluation

La validation des modèles est réalisée hors-échantillon à l'aide d'un schéma apprentissage/test, complété par une validation croisée sur l'échantillon d'apprentissage,

conformément aux bonnes pratiques d'évaluation des modèles prédictifs (scikit-learn developers, 2025). Les performances sont évaluées à l'aide de métriques adaptées aux prédictions probabilistes et au potentiel déséquilibre des classes :

- AUC-ROC : mesure la capacité de discrimination globale du modèle ;
- Brier Score : évalue la qualité probabiliste et la calibration des prédictions, des valeurs plus faibles indiquant une meilleure performance (Cartus et al., 2023) ;
- AUC-PR / Average Precision : utilisée en complément lorsque la classe positive est rare pour certaines garanties.

Dans une perspective de déploiement en environnement AssurTech, l'évaluation intègre également des considérations de gouvernance des modèles, notamment la traçabilité des transformations de données, la maîtrise du risque de modèle et le suivi de dérive, en cohérence avec les cadres récents de gestion des risques liés à l'IA (NIST AI RMF) et les exigences du secteur assurantiel (NIST, 2023).

3.3 Modélisation des charges des sinistres

Conformément au cadre fréquence–sévérité, les charges de sinistres sont modélisées séparément de la composante de fréquence. Pour chaque garantie k , la variable cible correspond au montant total des dépenses observées par assuré sur la période d'étude. L'estimation repose sur un modèle linéaire généralisé (GLM) de Poisson avec fonction de lien logarithmique, permettant de modéliser l'espérance conditionnelle des charges en fonction des caractéristiques sociodémographiques et contractuelles des assurés, tout en conservant une interprétation multiplicative des effets :

$$E(C_{ik} | X_i) = \exp(\beta_0^{(k)} + X_i^T \beta^k)$$

où C_{ik} désigne la charge associée à la garantie k pour l'assuré i , et X_i le vecteur des variables explicatives.

Le choix du GLM de Poisson s'explique par les propriétés des données de dépenses de santé, généralement asymétriques, positives et fortement hétérogènes, avec une proportion importante de valeurs nulles. Ce cadre permet ainsi de garantir la positivité des prédictions et de modéliser efficacement la variabilité des charges. Une analyse préliminaire met toutefois en évidence une surdispersion pour certaines garanties (notamment HOSP et D_MAX), caractérisées par une forte variabilité des montants. Afin d'en tenir compte, un GLM quasi-Poisson est utilisé, ce qui permet d'ajuster les erreurs standards des coefficients tout en conservant les estimations des paramètres. Une alternative aurait consisté à utiliser un modèle binomial négatif, explicitement conçu pour traiter la surdispersion. Toutefois, dans le cadre exploratoire de cette étude et afin de conserver un modèle facilement interprétable, l'approche quasi-Poisson offre un compromis approprié entre robustesse statistique et lisibilité actuarielle (Goldburd et al., 2020 ; Buckle et al., 2024).

3.4. Calcul des coûts totaux estimés

Le calcul des coûts totaux estimés pour chaque assuré est réalisé en intégrant les probabilités de sinistre prédites et les charges estimées selon la formule :

$$\text{Coût total}_i = \sum_k (P_{ik} \times C_{ik})$$

Equation (3) : Coût total individuel estimé

La méthode suit trois étapes : (1) calcul des probabilités de sinistre par garantie via le modèle ensembliste ; (2) estimation des charges associées à chaque garantie via le GLM de Poisson ;

(3) intégration par multiplication des probabilités et des charges pour obtenir la prime technique individuelle estimée.

4. Résultats

4.1 Prédiction des probabilités de sinistre

La prédiction des probabilités de survenance des sinistres repose sur une approche ensembliste combinant plusieurs modèles de classification. Cette stratégie vise à exploiter la complémentarité entre un modèle paramétrique interprétable, la régression logistique, et des modèles non paramétriques capables de capturer des relations non linéaires et des interactions complexes, notamment Random Forest, XGBoost, LightGBM et un réseau de neurones multicouche. Les probabilités finales sont obtenues à partir d'un mécanisme de soft voting, qui agrège les probabilités issues des différents modèles afin de produire une estimation plus robuste et moins sensible aux biais propres à une méthode particulière.

La performance du modèle d'ensemble est évaluée à l'aide de plusieurs indicateurs complémentaires. L'AUC (Area Under the Curve) constitue la principale mesure de la capacité de discrimination entre les assurés présentant un sinistre et ceux n'en présentant pas. La validation croisée permet d'évaluer la stabilité des performances hors échantillon, tandis que les matrices de confusion et les rapports de classification fournissent une analyse détaillée de la structure des erreurs de prédiction. Dans l'ensemble, les résultats indiquent que l'approche ensembliste permet d'obtenir des probabilités à la fois bien calibrées et discriminantes, améliorant ainsi la capacité de segmentation du risque et la qualité potentielle des décisions de tarification en assurance santé.

4.1.1. Évaluation des modèles de probabilités

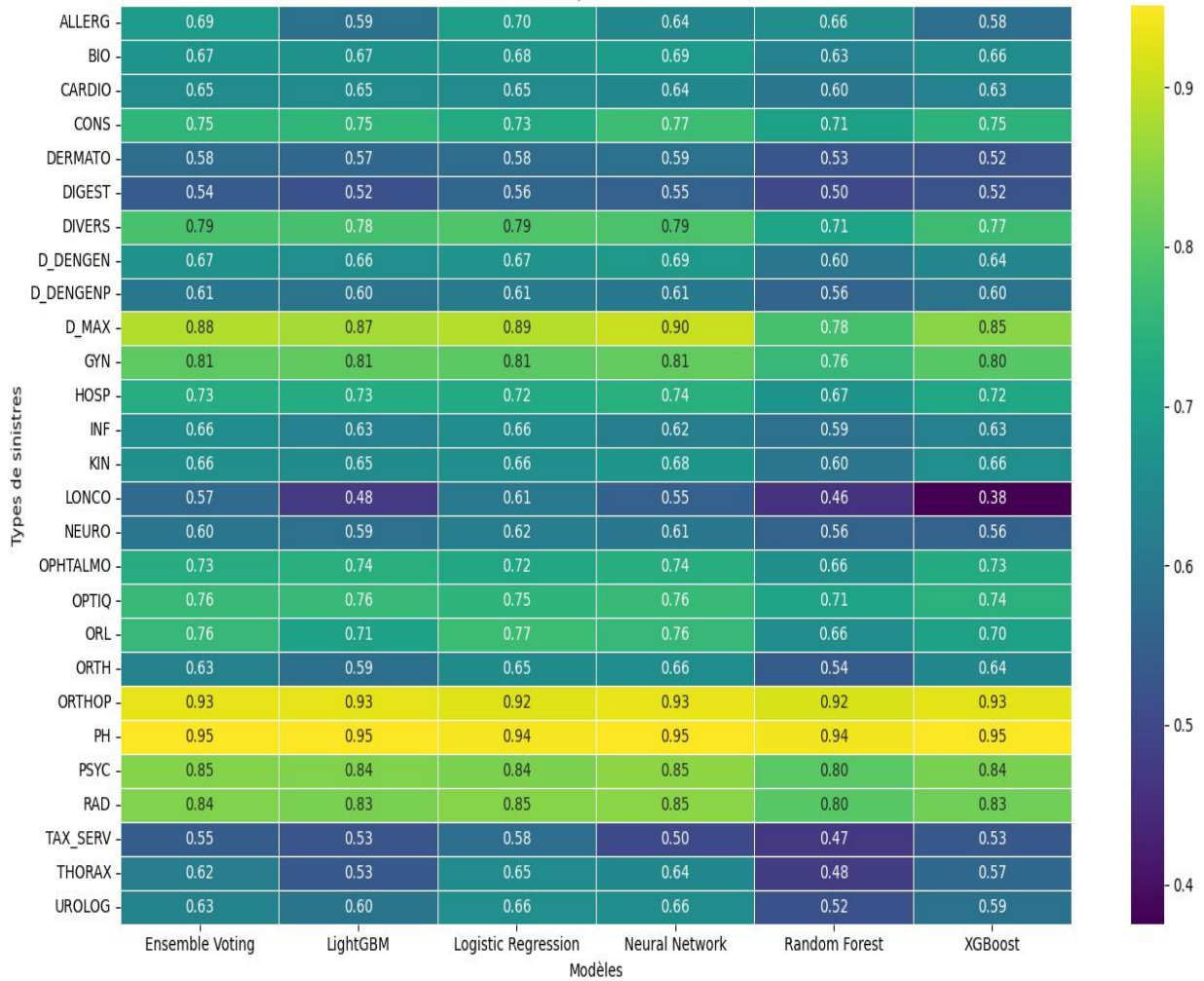
Afin de comparer les performances des différents modèles de classification, une synthèse des principaux indicateurs d'évaluation est présentée dans le tableau 3. Celui-ci regroupe les valeurs moyennes de l'AUC, du Brier Score ainsi que les résultats obtenus en validation croisée pour chaque modèle considéré.

Tableau 3 : Comparaison des performances des modèles de classification

Modèle	AUC moyen	Brier Score moyen	AUC validation croisée
Régression logistique	0.68	0.22	0.67
Random Forest	0.62	0.10	0.63
XGBoost	0.66	0.07	0.65
LightGBM	0.67	0.07	0.66
Réseau de neurones (MLP)	0.69	0.06	0.68
Ensemble Soft Voting	0.67	0.07	0.67

Les résultats montrent que les modèles d'apprentissage automatique présentent globalement de meilleures performances probabilistes que la régression logistique, notamment en termes de calibration des probabilités. Le réseau de neurones multicouche obtient les meilleurs résultats moyens, tandis que les modèles de gradient boosting (XGBoost et LightGBM) affichent également des performances robustes. Le modèle d'ensemble basé sur une stratégie de soft voting permet quant à lui de maintenir un bon compromis entre performance et stabilité des prédictions. La figure 4 présente une analyse plus détaillée des performances AUC obtenues pour chaque type de sinistre et pour l'ensemble des modèles.

Figure 4: Performance AUC par sinistre et modèle



4.1.2 Évaluation des modèles hybrides : analyse des performances prédictives

L'évaluation des performances des modèles repose sur plusieurs indicateurs complémentaires permettant d'apprécier à la fois la capacité de discrimination et la qualité de calibration des probabilités prédites. Deux analyses principales ont été réalisées : l'examen des courbes ROC et l'analyse du Brier Score. Les courbes ROC permettent d'évaluer la capacité des modèles à discriminer les assurés susceptibles de générer un sinistre de ceux n'en générant pas. La Figure 5 présente les courbes ROC obtenues pour l'ensemble des garanties à partir du modèle d'ensemble.

Les résultats montrent une performance globalement satisfaisante, avec des valeurs d'AUC majoritairement comprises entre 0,60 et 0,90, traduisant une capacité de discrimination correcte à élevée selon les garanties. Certaines prestations, telles que PH et ORTHOP, présentent des performances particulièrement élevées, tandis que d'autres garanties affichent une capacité discriminante plus modérée. Afin d'évaluer la qualité probabiliste des prédictions, la distribution du Brier Score a également été analysée pour les différents modèles. La Figure 6 présente la comparaison des scores obtenus.

Figure 5 : Courbes ROC de l'ensemble learning pour l'ensemble des garanties

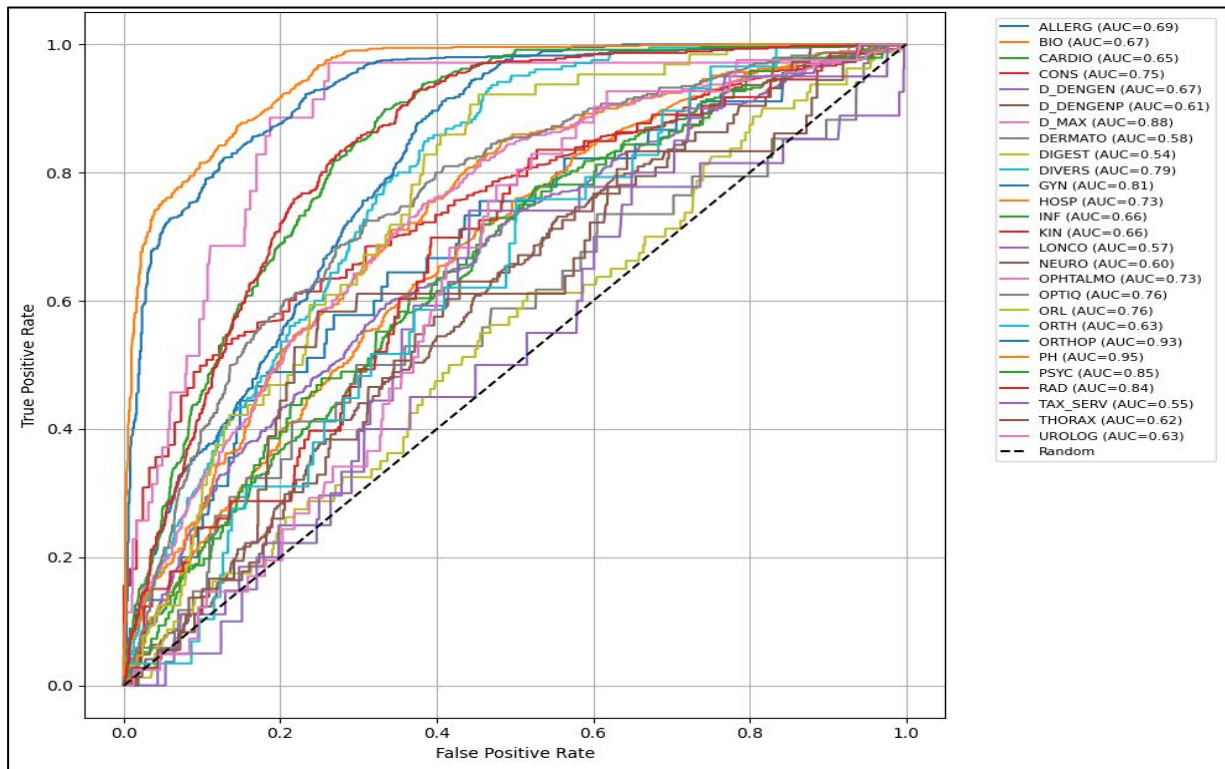
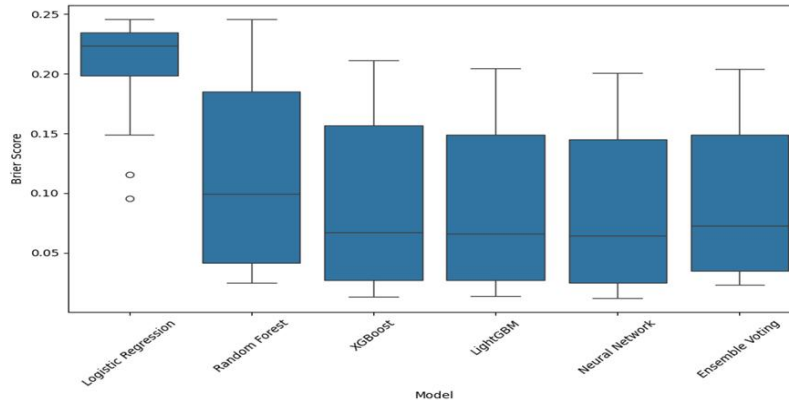


Figure 6 : Distribution du Brier Score par modèle



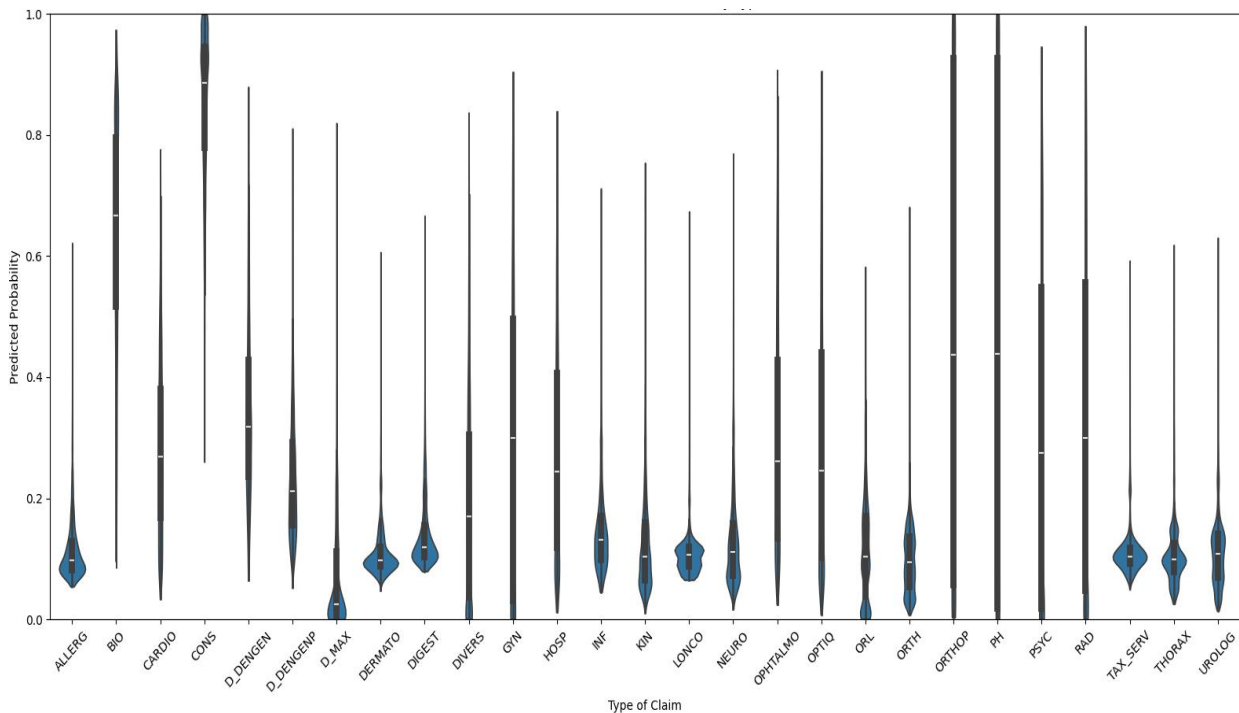
Les résultats indiquent que les modèles d'apprentissage automatique, notamment XGBoost, LightGBM et les réseaux de neurones, présentent généralement des Brier Scores plus faibles, traduisant une meilleure calibration des probabilités. À l'inverse, la régression logistique affiche des scores plus élevés et une variabilité plus importante. Le modèle Ensemble Voting se situe dans une position intermédiaire, en combinant les prédictions de plusieurs algorithmes afin d'améliorer la robustesse globale des estimations.

4.1.3. Analyse des distributions des probabilités prédites

Au-delà des indicateurs globaux tels que l'AUC ou le Brier Score, l'analyse de la distribution des probabilités prédites permet d'examiner la manière dont le modèle différencie les niveaux de risque entre les garanties. La Figure 7 présente la distribution des probabilités de sinistre estimées pour chaque type de garantie sous forme de diagrammes de violon, permettant de

visualiser simultanément la dispersion, la densité et la médiane des probabilités. L'analyse de cette figure met en évidence plusieurs éléments intéressants. Premièrement, on observe que certaines garanties présentent des probabilités prédites relativement faibles et fortement concentrées, notamment pour les postes tels que D_MAX, ORL ou DIGEST. Cela reflète une faible fréquence d'occurrence des sinistres, conduisant le modèle à attribuer des probabilités globalement basses.

Figure 7 : Distribution des probabilités de sinistre prédites par type de garantie



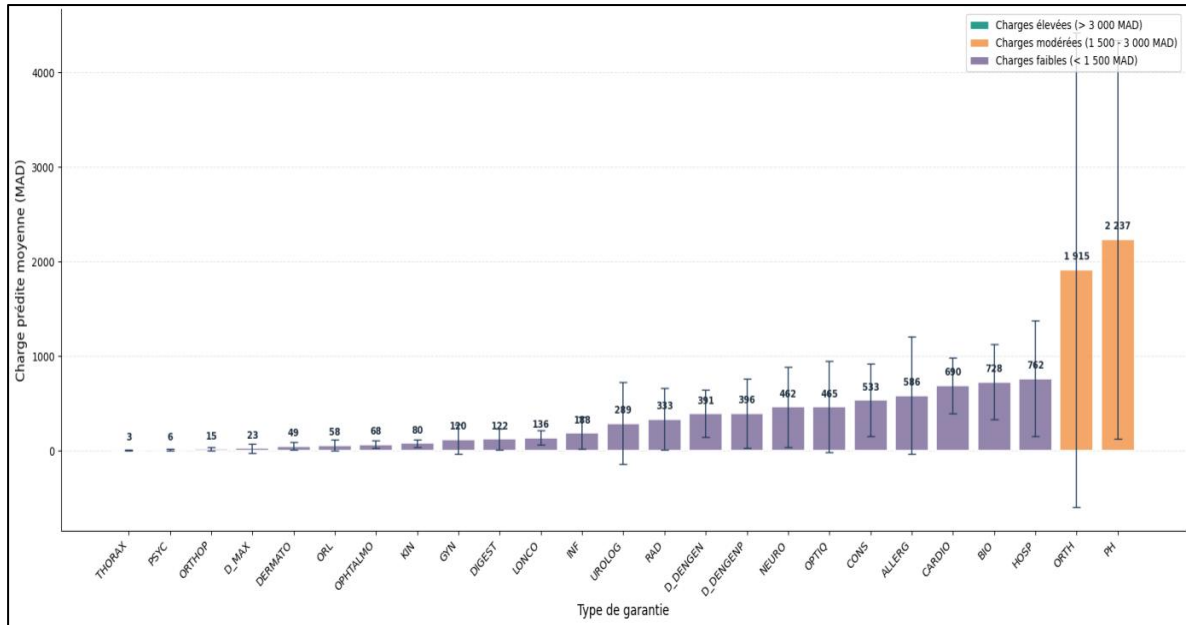
À l'inverse, d'autres garanties présentent des distributions plus étalées, traduisant une plus grande variabilité du risque entre les assurés. C'est notamment le cas pour des postes comme PH, PSYC ou ORTH, où le modèle identifie une hétérogénéité plus importante des profils de risque. Par ailleurs, certaines garanties comme CONS ou BIO présentent des distributions relativement élevées, ce qui est cohérent avec la fréquence plus importante des actes médicaux associés à ces postes de soins dans les données observées. Dans l'ensemble, cette analyse confirme que le modèle parvient à différencier les niveaux de risque selon les catégories de soins, tout en capturant l'hétérogénéité des comportements de consommation médicale. La distribution des probabilités prédites constitue ainsi un outil utile pour évaluer la cohérence actuarielle du modèle et son potentiel d'utilisation dans un cadre de tarification ou de pilotage du risque.

4.2 Modélisation des charges des sinistres

Après l'estimation des probabilités d'occurrence des sinistres, la seconde composante du cadre fréquence-sévérité consiste à modéliser les charges associées aux sinistres. L'objectif est d'estimer le montant attendu des dépenses de santé conditionnellement à la survenance d'un sinistre, afin d'obtenir une évaluation plus complète du coût du risque assurantiel. Pour cela, un modèle linéaire généralisé (GLM) de Poisson avec fonction de lien logarithmique a été estimé pour chaque type de garantie. La figure 8 présente la distribution des charges prédites par garantie, accompagnée d'intervalles d'incertitude. Les résultats mettent en évidence une hétérogénéité importante des niveaux de dépenses selon les postes de soins.

Certaines garanties, notamment PH, ORTH et HOSP, présentent des charges moyennes plus élevées ainsi qu’une variabilité plus importante, ce qui reflète le caractère potentiellement coûteux de ces prestations médicales. À l’inverse, d’autres garanties telles que D_MAX, ORL ou THORAX affichent des niveaux de dépenses moyens plus faibles et une dispersion plus limitée. Ces résultats confirment l’existence de profils de coûts différenciés selon les catégories de soins, ce qui souligne l’intérêt de modéliser séparément les composantes de fréquence et de sévérité dans le cadre du modèle fréquence–sévérité utilisé pour la tarification en assurance santé.

Figure 8 : Charges de sinistres prédites par garantie avec intervalles d’incertitude.



Afin d’examiner l’influence des variables explicatives, la Figure 9 présente une carte de chaleur des coefficients estimés du GLM par garantie. Cette représentation permet d’identifier le sens et l’intensité des effets des variables. Les coefficients positifs indiquent une augmentation attendue des charges, tandis que les coefficients négatifs suggèrent un effet modérateur. Les résultats montrent que certaines variables contractuelles et socio-démographiques exercent des effets différenciés selon les garanties, traduisant l’hétérogénéité des comportements de consommation de soins.

La Figure 10 présente les erreurs standards associées aux coefficients estimés. Les valeurs observées restent globalement faibles, ce qui indique une précision satisfaisante des estimations et suggère une relative stabilité des coefficients du modèle. La Figure 11 illustre la significativité statistique des variables explicatives à travers les p-values associées aux coefficients. La majorité des valeurs observées sont très faibles, confirmant la contribution statistiquement significative des variables retenues dans l’explication des charges de sinistres. Dans l’ensemble, ces résultats montrent que la modélisation par GLM permet de capturer efficacement les déterminants des charges de sinistres et leur variabilité selon les garanties. L’intégration de cette composante de sévérité avec la modélisation de la fréquence permet ainsi d’obtenir une évaluation plus complète du coût attendu du risque, constituant une étape essentielle dans les processus actuariels de tarification et de gestion du portefeuille en assurance santé.

Figure 9 : Heatmap des coefficients estimés du GLM par garantie

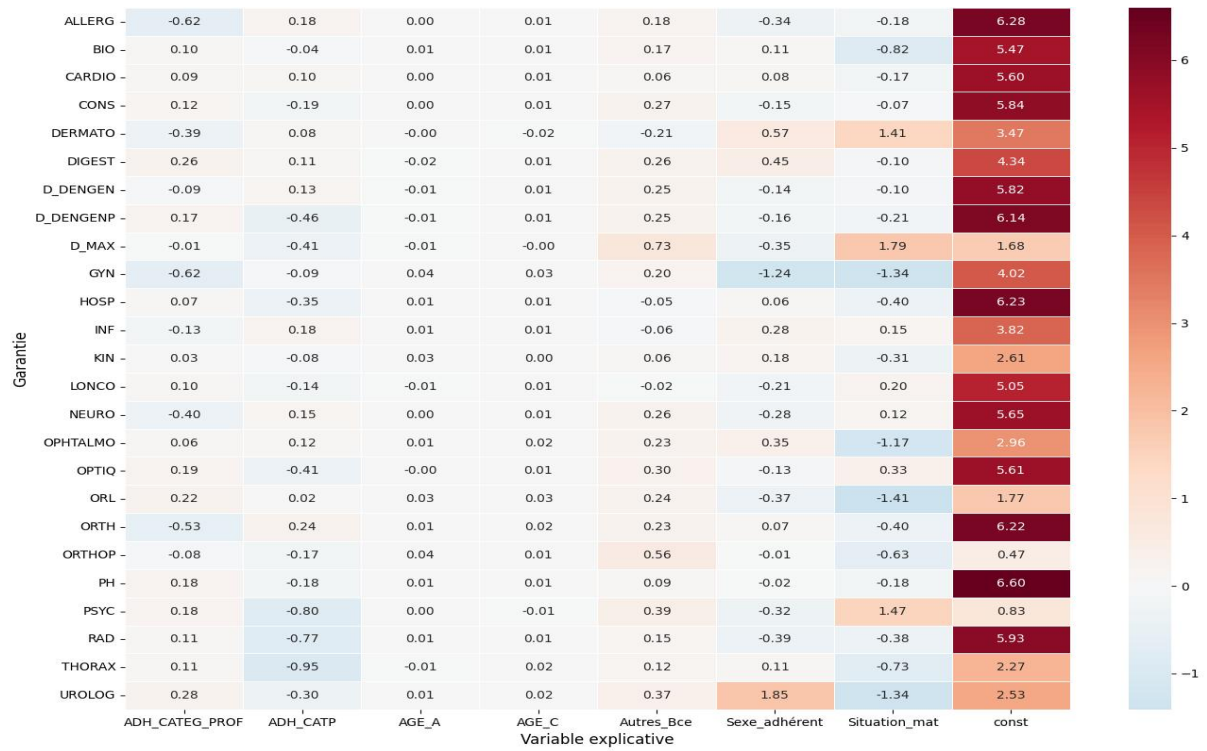


Figure 10 : Heatmap des erreurs standards du GLM

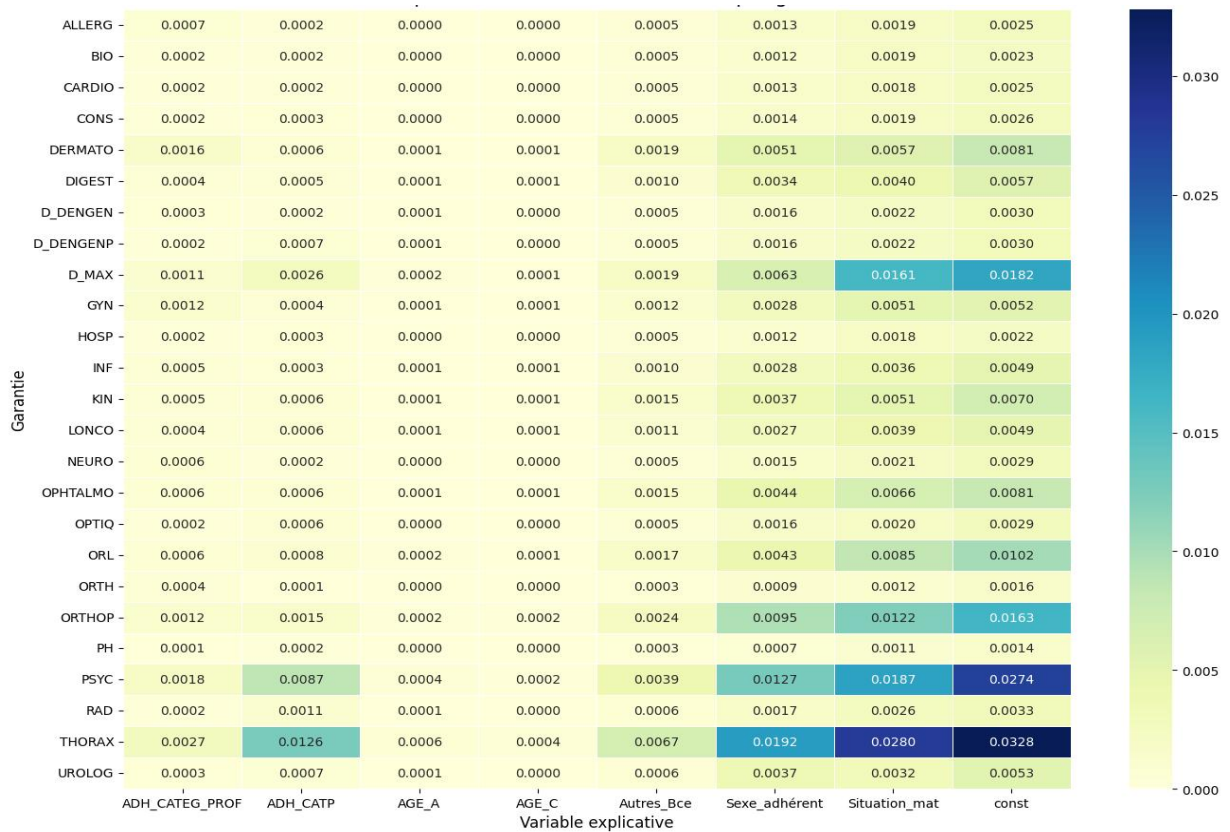


Figure 11 : Heatmap des p-values du GLM par garantie



4.3 Calcul des coûts totaux estimés

Dans un premier temps, les variables individuelles (Bloc 1) des assurés sont utilisées comme variables explicatives du modèle. À partir de ces informations, un modèle d'ensemble permet d'estimer, pour chaque assuré i et chaque garantie k , la probabilité de survenance du sinistre, notée :

$$P_{ik} = P(\text{Sinistre}_k | X_i)$$

Dans un second temps, les charges attendues conditionnelles sont estimées à l'aide d'un GLM de Poisson, permettant de modéliser l'espérance des dépenses associées à chaque garantie :

$$C_{ik} = E[\text{Charge}_k | X_i]$$

Le coût attendu par garantie est ensuite obtenu en combinant les deux composantes du modèle fréquence-sévérité :

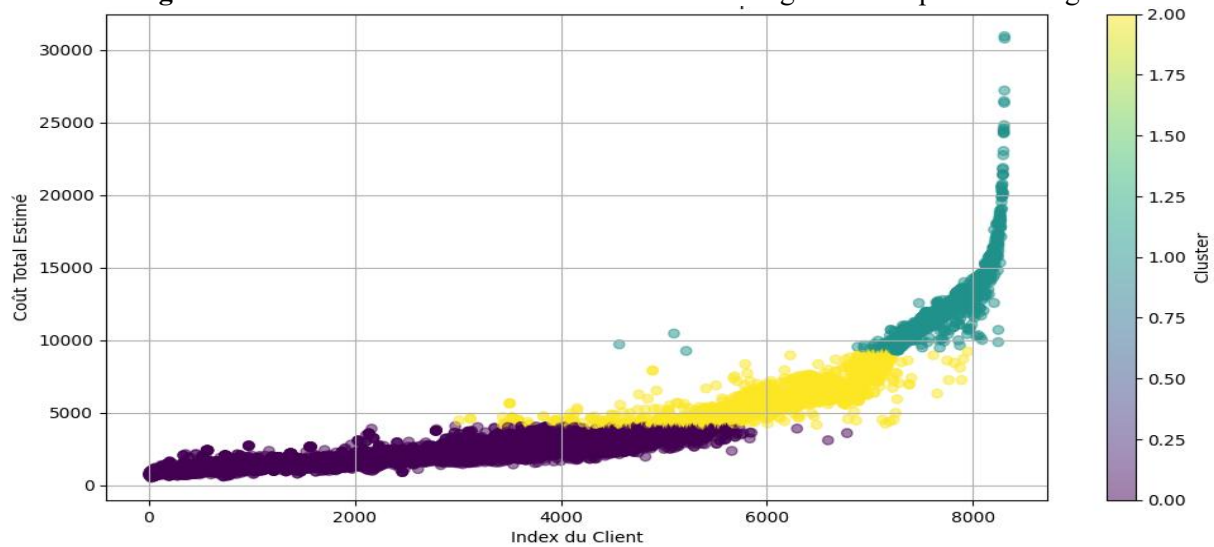
$$\text{Coût estimé}_{ik} = P_{ik} \times C_{ik}$$

Le coût total individuel estimé correspond alors à la somme des coûts attendus sur l'ensemble des garanties :

$$\text{Coût total}_i = \sum_k (P_{ik} \times C_{ik})$$

Afin d'analyser l'hétérogénéité des profils de risque, les coûts totaux estimés sont utilisés pour réaliser une segmentation des assurés à l'aide d'un algorithme de clustering k-means ($k = 3$). Cette segmentation permet d'identifier trois profils principaux :

- un groupe à faible risque, regroupant la majorité des assurés ;
- un groupe à risque intermédiaire ;
- un groupe à haut risque, caractérisé par une sinistralité attendue plus élevée.

Figure 12 : Distribution des coûts totaux attendus et segmentation par clustering

5. Discussion

Cette étude contribue à la littérature en sciences de gestion et en management des systèmes d'information en proposant un cadre opérationnel d'aide à la décision pour l'intégration des techniques de machine learning dans les processus de tarification en assurance santé. Elle met en évidence les conditions organisationnelles et méthodologiques permettant de concilier performance prédictive, explicabilité des modèles et exigences de gouvernance dans un environnement fortement réglementé. Le cadre proposé éclaire les choix managériaux relatifs à l'architecture des modèles, aux dispositifs de validation, ainsi qu'à l'organisation des fonctions actuarielles, data et conformité autour du cycle de vie des modèles. Les résultats confirment l'intérêt d'intégrer des techniques de machine learning dans le cadre actuariel classique fréquence-sévérité pour la tarification en assurance santé. Les modèles non paramétriques présentent globalement de meilleures performances prédictives que le modèle de référence paramétrique, en particulier en termes de discrimination et de qualité probabiliste. Ces gains s'expliquent par leur capacité à capter des non-linéarités et des interactions complexes entre variables explicatives, fréquemment observées dans les données de sinistres santé.

Toutefois, l'amélioration de la performance n'est ni uniforme ni systématique selon les postes de soins et les segments d'assurés. Pour certains risques plus stables ou faiblement fréquents, les modèles traditionnels conservent des performances comparables, tout en offrant une meilleure interprétabilité. Ces résultats soulignent le caractère complémentaire, plutôt que substitutif, des approches de machine learning par rapport aux modèles actuariels classiques. L'approche ensembliste par soft voting apparaît comme un compromis méthodologique pertinent, permettant d'améliorer la robustesse et la stabilité des probabilités prédites. En agrégeant plusieurs modèles aux propriétés différentes, cette stratégie limite la dépendance à une spécification unique et renforce la fiabilité des résultats hors-échantillon, un point essentiel en contexte de tarification opérationnelle.

Sur le plan managérial, les résultats suggèrent que l'adoption du machine learning en tarification doit s'inscrire dans une logique d'évolution progressive des pratiques actuarielles. Les modèles avancés peuvent constituer des outils efficaces d'aide à la décision, en améliorant la segmentation du risque et l'anticipation des charges, à condition de rester

articulés au cadre économique et technique du métier. L'approche hybride proposée facilite l'appropriation des résultats par les équipes actuarielles et de pilotage, en maintenant un lien explicite entre les déterminants du risque et les décisions tarifaires. Elle encourage également une organisation plus transversale du processus de tarification, impliquant les fonctions actuarielles, data, systèmes d'information et contrôle des risques, autour de la qualité des données, de la traçabilité et de la documentation des modèles. Au-delà de la tarification, les probabilités et charges estimées peuvent être mobilisées pour le pilotage médical, la prévention ciblée et la gestion proactive des dépenses de santé, renforçant ainsi la valeur stratégique des modèles prédictifs pour les assureurs.

Dans un environnement réglementé comme l'assurance santé, la performance prédictive des modèles ne constitue qu'un des critères de leur acceptabilité. Les exigences d'explicabilité, d'auditabilité et de maîtrise du risque de modèle jouent un rôle central dans l'encadrement des usages du machine learning. L'approche hybride retenue contribue à répondre à ces exigences en combinant des modèles interprétables et des techniques plus flexibles, tout en s'appuyant sur des dispositifs de traçabilité et de documentation du pipeline de modélisation. Elle s'inscrit ainsi dans une logique de gouvernance des modèles fondée sur le cycle de vie, intégrant validation initiale, suivi des performances et recalibrage dans le temps. Enfin, les enjeux d'équité et de responsabilité sociale propres à l'assurance santé appellent à une utilisation prudente et proportionnée des modèles avancés. Les résultats plaident pour une gouvernance renforcée, permettant de concilier innovation technologique, conformité réglementaire et objectifs d'équité tarifaire.

6. Conclusion

Cette étude a exploré l'utilisation d'approches hybrides combinant des techniques de régression et d'apprentissage automatique pour modéliser les coûts des sinistres en assurance santé. En intégrant des variables explicatives pertinentes liées aux caractéristiques des assurés et à leurs historiques de sinistres, nous avons développé des modèles capables de prédire les probabilités de survenance des sinistres ainsi que les coûts associés. Les résultats montrent que les modèles d'apprentissage automatique, tels que les forêts aléatoires et les réseaux de neurones, surpassent souvent les méthodes classiques de régression en termes de précision prédictive. Toutefois, ils présentent également des défis, notamment en termes de complexité, de suradaptation et de difficulté d'interprétation. Ces résultats soulignent l'importance d'une sélection rigoureuse des modèles et d'une validation croisée minutieuse pour assurer la robustesse des prédictions.

Les implications pratiques de cette recherche sont significatives pour les assureurs, qui peuvent utiliser ces modèles pour améliorer la gestion des risques, optimiser la tarification des primes, et mettre en œuvre des stratégies de prévention plus ciblées. Néanmoins, il est essentiel de reconnaître les limitations de cette étude, notamment la dépendance à la qualité des données et la nécessité d'une infrastructure technique adaptée pour déployer ces modèles en pratique. Enfin, cette recherche ouvre la voie à de futures investigations, notamment sur l'intégration de nouvelles variables explicatives et l'application de techniques d'apprentissage profond. Ces perspectives offrent un potentiel considérable pour affiner encore les estimations des coûts des sinistres et pour étendre l'application de ces modèles à d'autres branches de l'assurance. En conclusion, l'innovation technologique, soutenue par des modèles hybrides avancés, constitue un levier stratégique pour les compagnies d'assurance souhaitant rester compétitives dans un environnement en perpétuelle évolution. L'adoption de ces approches permettra non seulement d'améliorer la précision des prévisions des coûts des sinistres, mais aussi de renforcer la résilience et la performance des entreprises d'assurance.

Références

- Kapse, M., Sharma, V., Vidhale, R., & Vellanki, V. (2025). *Customization of health insurance premiums using machine learning and explainable AI. International Journal of Information Management Data Insights*, 5(1), 100328. <https://doi.org/10.1016/j.jjime.2025.100328>
- Laub, P. J., Pho, T., & Wong, B. (2025). *An interpretable deep learning model for general insurance pricing* (arXiv:2509.08467). arXiv. <https://doi.org/10.48550/arXiv.2509.08467>
- Aas, K., Charpentier, A., Huang, F., & Richman, R. (2024). *Insurance analytics: Prediction, explainability and fairness. Annals of Actuarial Science*, 18(3), 535–539. <https://doi.org/10.1017/S1748499524000289>
- Holvoet, F., Antonio, K., & Henckaerts, R. (2023). *Neural networks for insurance pricing with frequency and severity data: A benchmark study from data preprocessing to technical tariff* (RePEc:arx:papers:2310.12671) [Preprint]. arXiv.org. <http://arxiv.org/pdf/2310.12671>
- Lledó, J., Espinosa, P., & Pérez, V. (2025). *A dataset for health insurance analysis: Integrating individual and area-based contextual variables. Scientific Data*, 13(1), 54. <https://doi.org/10.1038/s41597-025-06372-z>
- Cartus, A. R., Samuels, E. A., Cerdá, M., & Marshall, B. D. L. (2023). *Outcome class imbalance and rare events: An underappreciated complication for overdose risk prediction modeling. Addiction*, 118(6), 1167–1176. <https://doi.org/10.1111/add.16133> (PMCID: PMC10175167)
- Goldburd, M., Khare, A., Tevet, D., & Guller, D. (2020). *Generalized linear models for insurance rating* (Monograph No. 5). Casualty Actuarial Society. https://www.casact.org/sites/default/files/2021-03/8_GLM.pdf
- Buckle, J., Biscaglia, N., & Silveiro, J. (2024, 12 novembre). *Modelling a medical expenses portfolio*. Milliman. <https://www.milliman.com/en/insight/modelling-a-medical-expenses-portfolio>
- Sinha, T., Desai, S., & Mahal, A. (2014). *The value of claims analysis in health microinsurance*. International Labour Office. <https://www.ilo.org/media/435621/download>
- National Institute of Standards and Technology (NIST). (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. <https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf>
- European Insurance and Occupational Pensions Authority (EIOPA). (2025, 6 août). *Opinion on Artificial Intelligence governance and risk management*. https://www.eiopa.europa.eu/publications/opinion-artificial-intelligence-governance-and-risk-management_en
- European Parliament & Council of the European Union. (2024, 13 juin). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence and amending certain Acts (Artificial Intelligence Act)* (OJ L 2024/1689). <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

Wikipedia. (2026). *Explainable artificial intelligence*. In *Wikipedia, The Free Encyclopedia*.
https://en.wikipedia.org/wiki/Explainable_artificial_intelligence

scikit-learn developers. (2025). *Cross-validation: evaluating estimator performance*. In *scikit-learn documentation*.

https://scikit-learn.org/stable/modules/cross_validation.html