

CONSTRUCTION D'UN MODÈLE DE MACHINE LEARNING INTERPRÉTABLE POUR LA TARIFICATION EN ASSURANCE NON-VIE

KRUPOVA Marketa

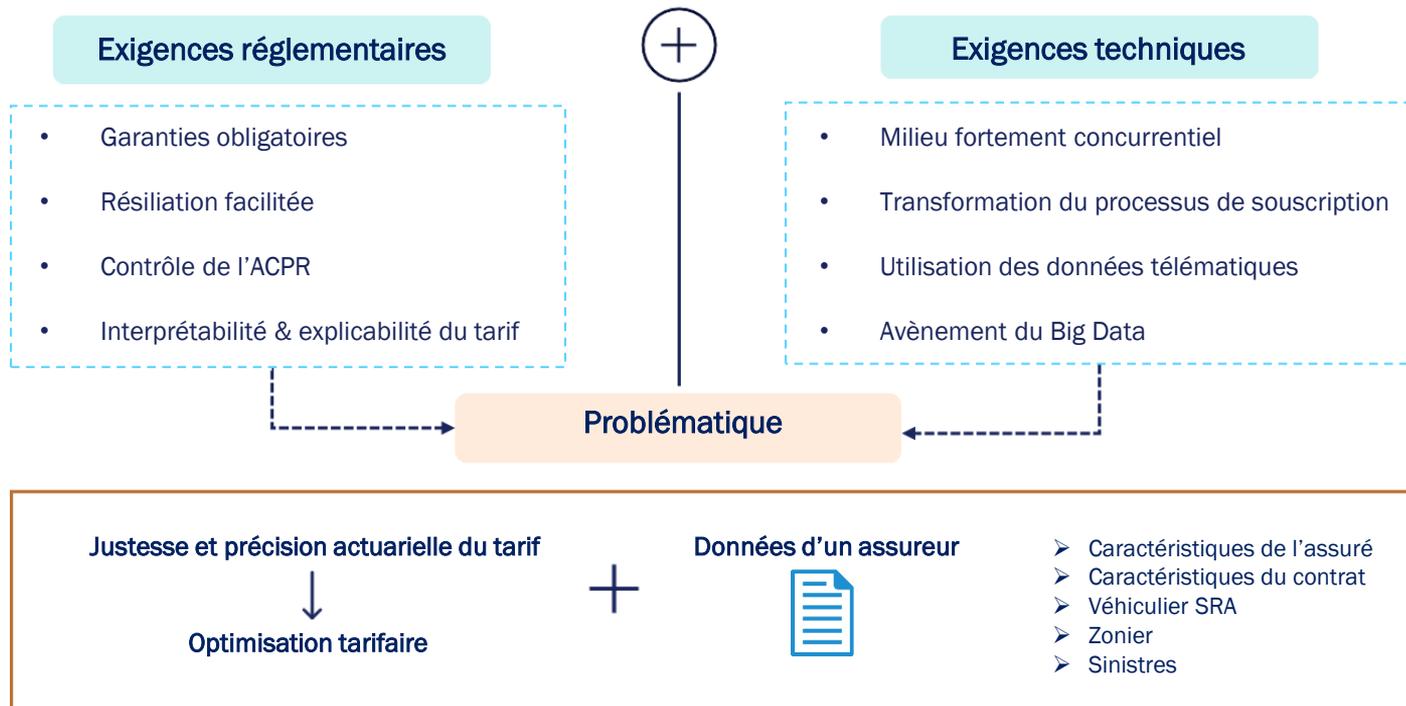
26 septembre 2023

• SOMMAIRE

- 1 • CONTEXTE GÉNÉRAL DE LA TARIFICATION EN ASSURANCE NON-VIE
- 2 • MODÈLE LINÉAIRE GÉNÉRALISÉ ET MODÈLES DE MACHINE LEARNING
- 3 • MODÈLE ADDITIF GÉNÉRALISÉ ET SES DÉCLINAISONS
- 4 • PRÉSENTATION DES RÉSULTATS ET DES LIMITES DE L' ÉTUDE

- 1 • Contexte général de la tarification en assurance non-vie

• CONTEXTE ET PRÉSENTATION GÉNÉRALE



• STRUCTURE PARAMÉTRIQUE DU MODÈLE LINÉAIRE GÉNÉRALISÉ (GLM)

$$g(\mathbb{E}[Y]) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Diagram illustrating the structure of the Generalized Linear Model (GLM) equation:

- variable réponse** (response variable) is associated with $\mathbb{E}[Y]$.
- variables explicatives** (explanatory variables) are associated with X_1, \dots, X_p .
- fonction de lien** (link function) is associated with $g(\mathbb{E}[Y])$.
- effet moyen** (mean effect) is associated with β_0 .
- effet X_1** (effect of X_1) is associated with $\beta_1 X_1$.
- effet X_p** (effect of X_p) is associated with $\beta_p X_p$.

Référence pour la tarification

Fonctions de lien et distributions adaptées

Construction d'une grille tarifaire multiplicative

Coût moyen : loi Gamma, fonction de lien logarithme

Fréquence : loi Poisson, fonction de lien logarithme

• STRUCTURE NON-PARAMÉTRIQUE DES MODÈLES DE MACHINE LEARNING

$$Y = f(X_1, \dots, X_p)$$

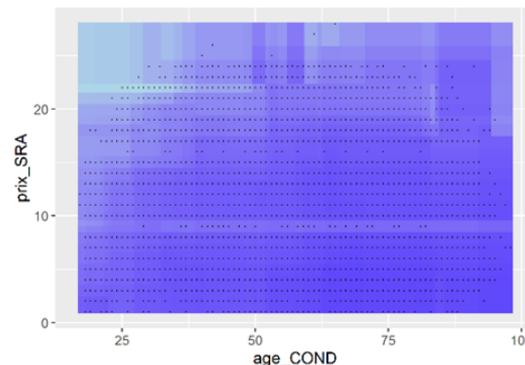
Relation complexe entre Y et X

Intégration des interactions et des non-linéarités

Architectures sophistiquées et agrégation des modèles

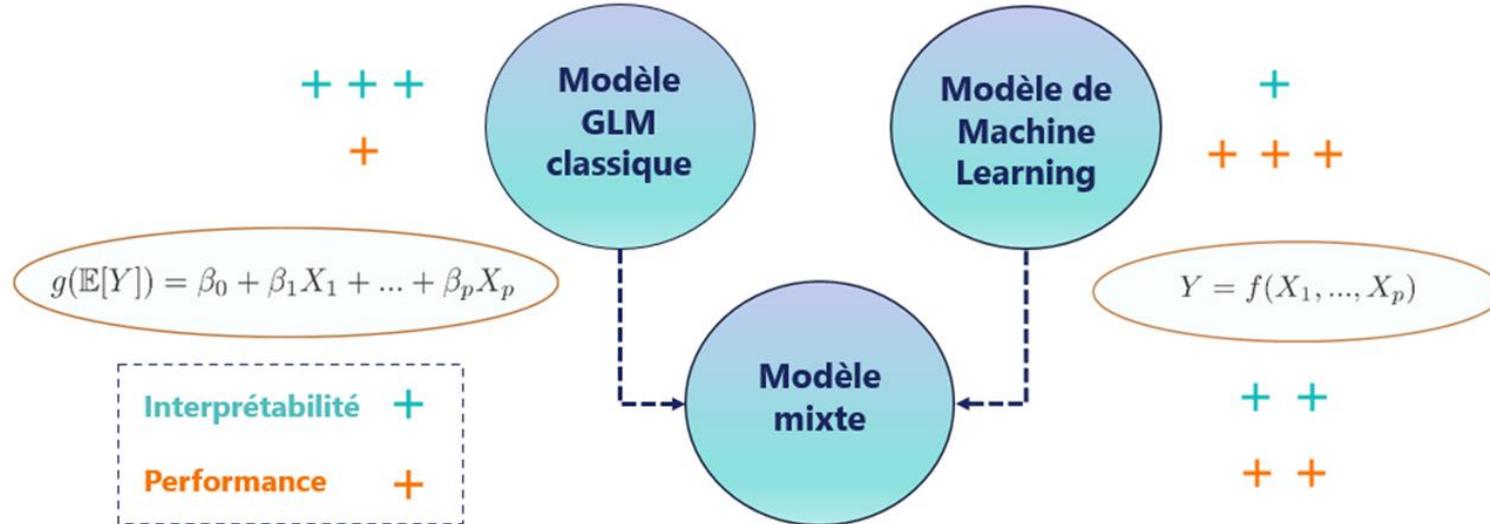
Bagging + Boosting
Réseaux de neurones

manque d'interprétabilité



Exemple de prédiction dans le plan âge conducteur / classe de prix du véhicule pour la modélisation du coût moyen

• SYNTHÈSE ET PRÉSENTATION DE L'OBJECTIF DE L'ÉTUDE



- STRUCTURE SEMI-PARAMÉTRIQUE DU MODÈLE ADDITIF GÉNÉRALISÉ (GAM)

$$g(\mathbb{E}[Y]) = \beta_0 + f_1(X_1) + \dots + f_p(X_p)$$

fonction de lien
 effet moyen
 effet X_1
 effet X_p

Ajout de la non-linéarité

Maintien de la structure additive

Fonctions composantes non-paramétriques

lissage × apprentissage

fonctions splines



structure arborescente

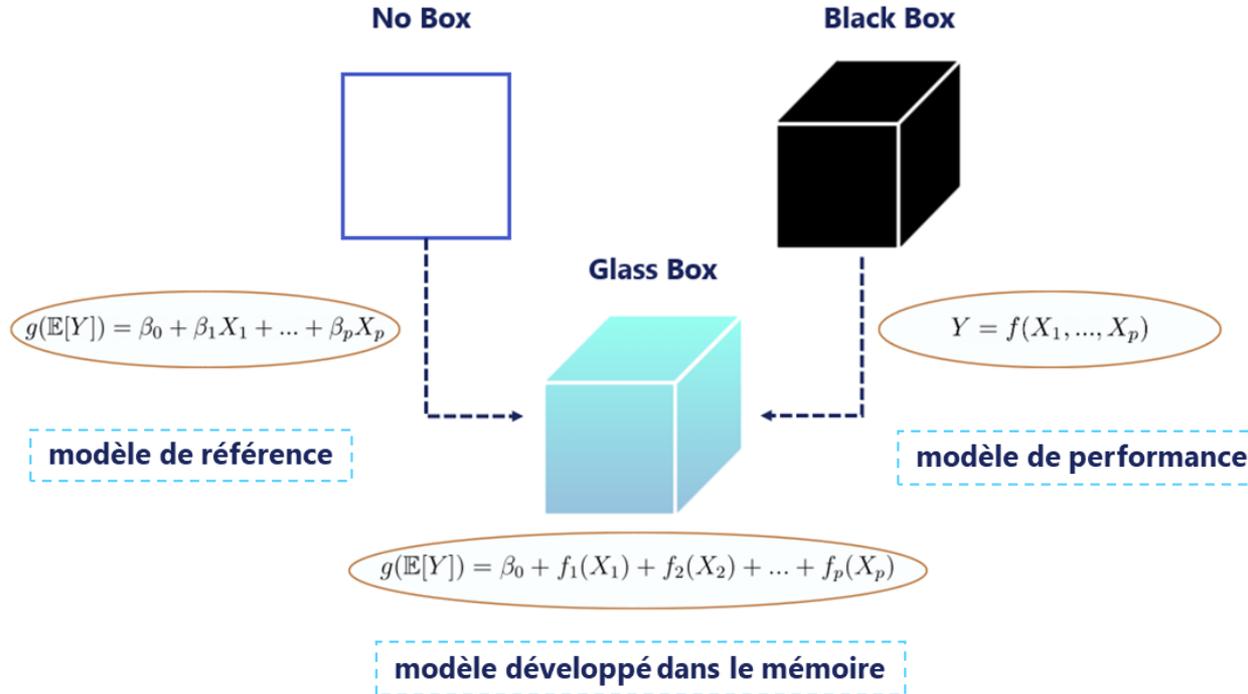


structure neuronale

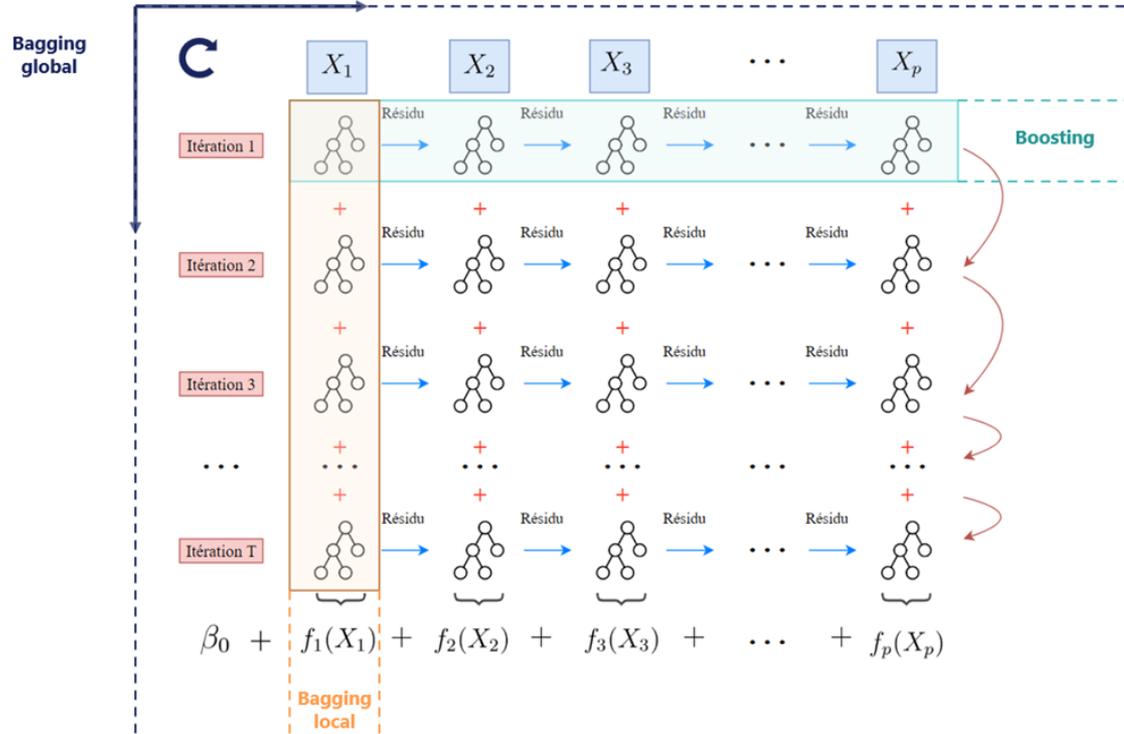


- 3 • Modèle additif généralisé et ses déclinaisons

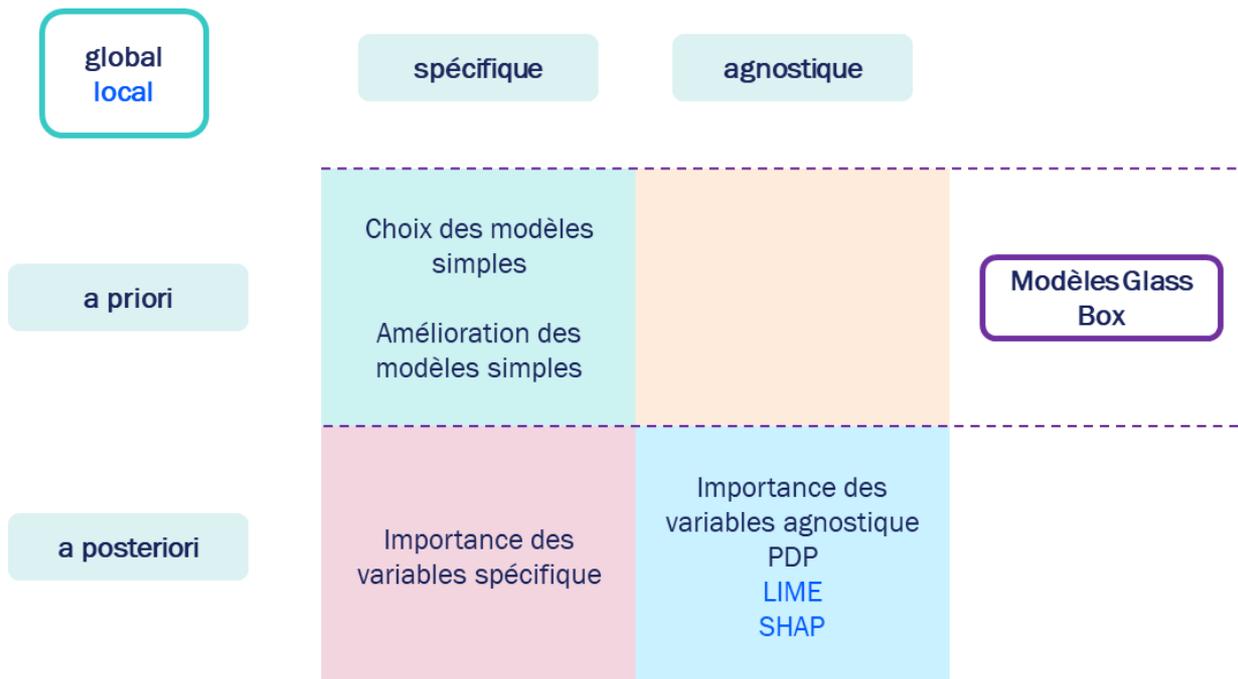
- POSITIONNEMENT DU MODÈLE GLASS-BOX



• MODÈLE ADDITIF GÉNÉRALISÉ AVEC UNE STRUCTURE ARBORESCENTE



• UNE INTERPRÉTABILITÉ INTRINSÈQUE



• INTERPRÉTABILITÉ DES MODÈLES

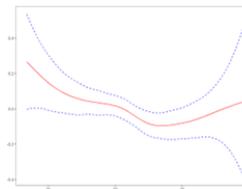
lecture des coefficients

variable	coefficient	multiplier
âge conducteur	-0,003	0,997
ancienneté permis	-0,002	0,998
classe de prix du véhicule	0,032	1,032
zonier	0,008	1,008

Exemple de coefficients obtenus par un modèle GLM



lecture des fonctions



Exemple de fonction composante obtenue par un modèle GAM splines



précision × pertinence actuarielle



Exemple de fonctions composante obtenues par un modèle GAM arbres

• RÉSULTATS DE MODÉLISATION

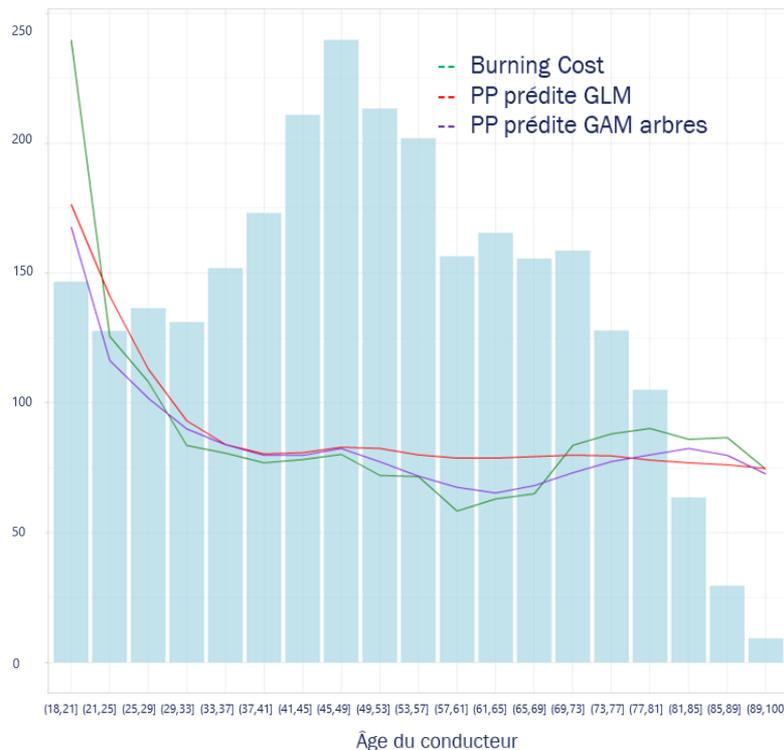
Modélisation du coût moyen

Modélisation de la fréquence

Modélisation de la prime pure

- Métriques de performance
- Temps computationnel
- Analyse univariée des valeurs observées et prédites

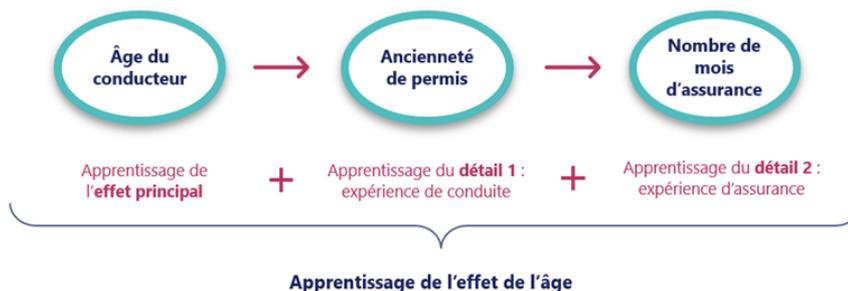
Comparaison du risque observé et prédit avec les modèles GLM et GAM arbres sur la variable âge conducteur



• COMPARAISON AVEC LE MODÈLE GAM SPLINES

Lissage × Apprentissage

- Des **métriques de performance** comparables (au vu des métriques de performance classiques)
- Un gain considérable au niveau du **temps computationnel** (jusqu'à 7 fois plus rapide)
- Un potentiel afférent à la **structure semi-paramétrique** restant à explorer
 - Distributions
 - Fonctions de lien
 - Calibration des paramètres
 - Relations de causalité



Exemple de la prise en compte des relations de causalité dans l'ordre d'apprentissage des variables explicatives

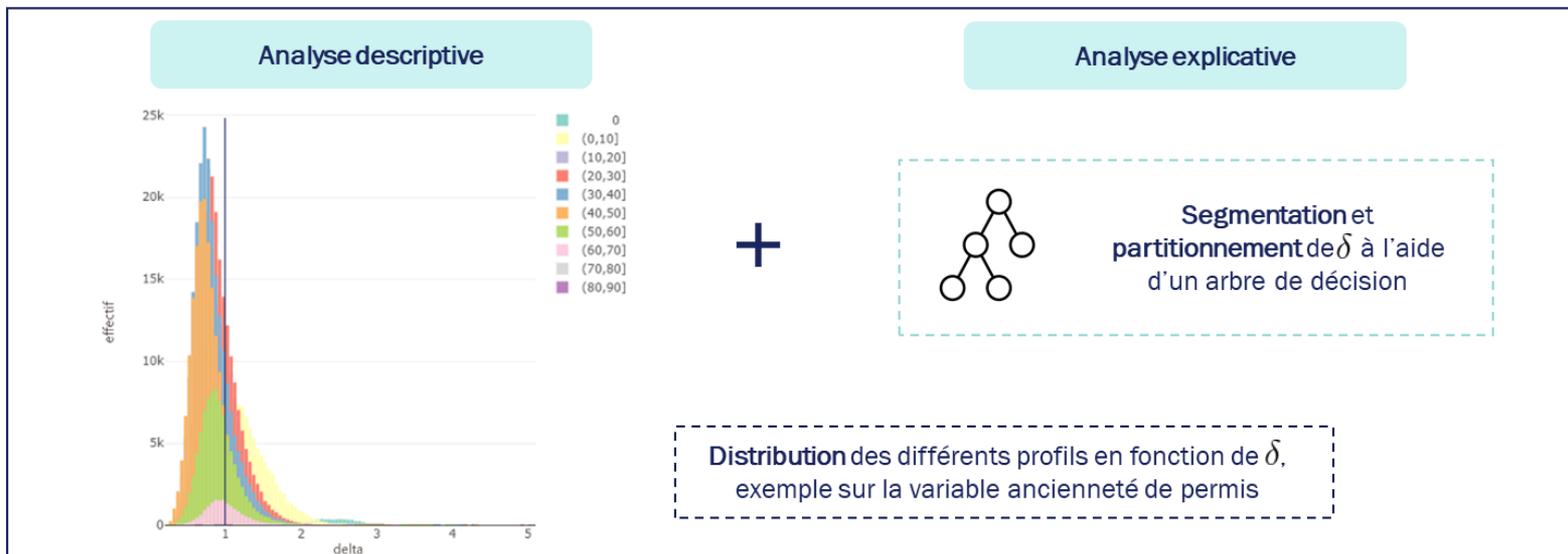
- 4 • Présentation des résultats et des limites de l'étude

• OPTIMISATION TARIFAIRE

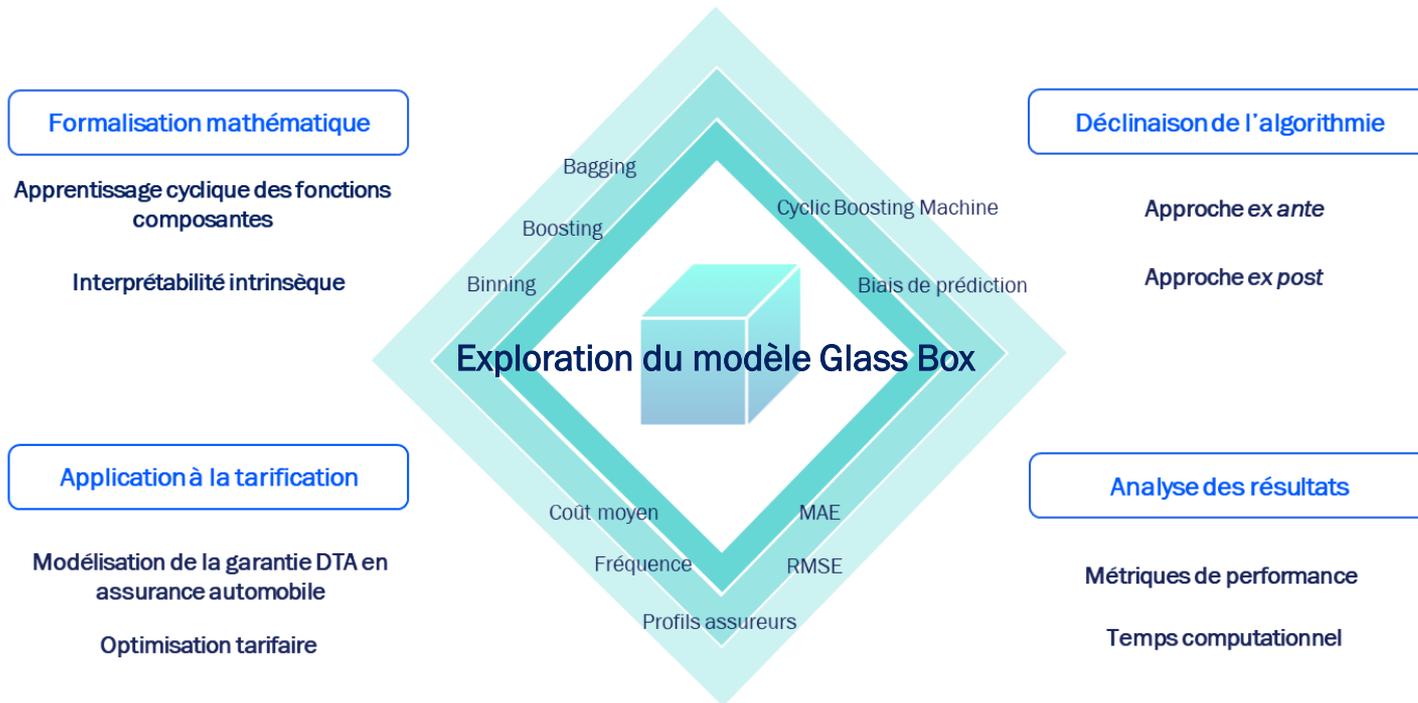
$$\delta = \frac{\text{PP GAM arbres}}{\text{PP GLM}}$$



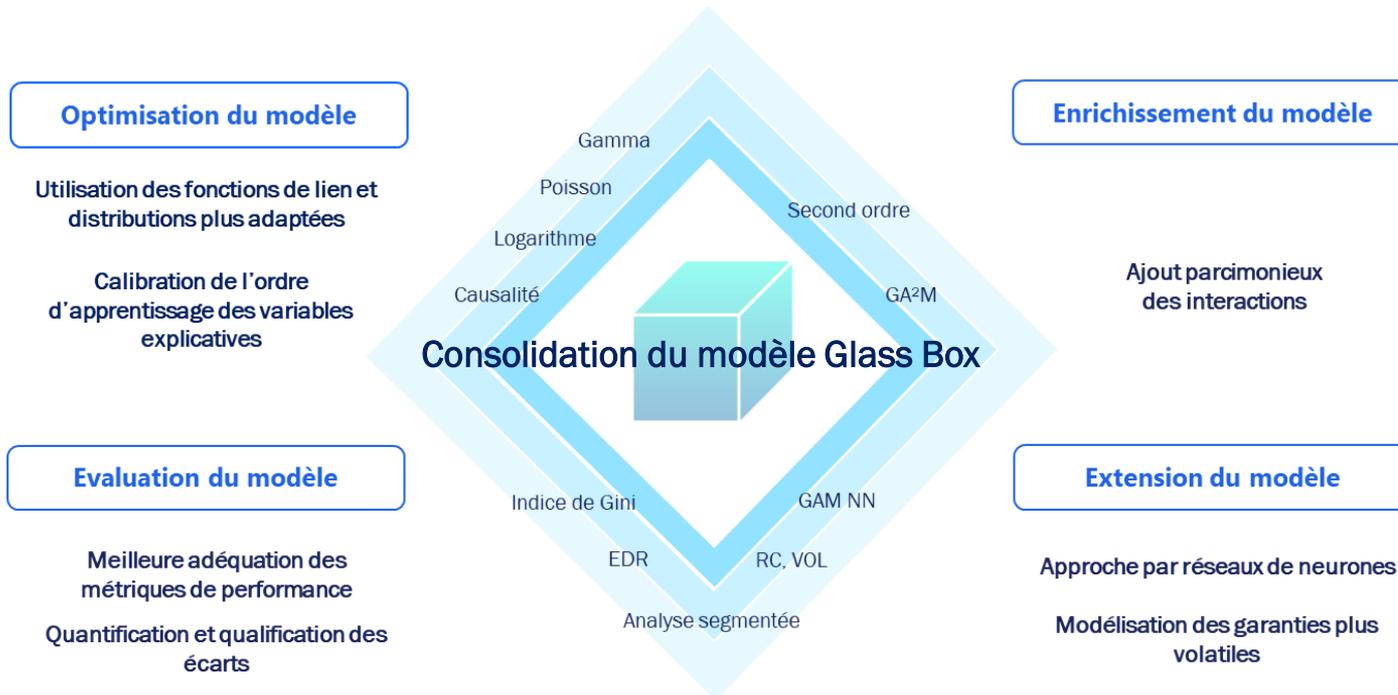
- $\delta > 1$: surtarification
- $\delta < 1$: sous-tarification
- $\delta = 1$: même niveau de tarification



• CONTRIBUTIONS DE L'ÉTUDE



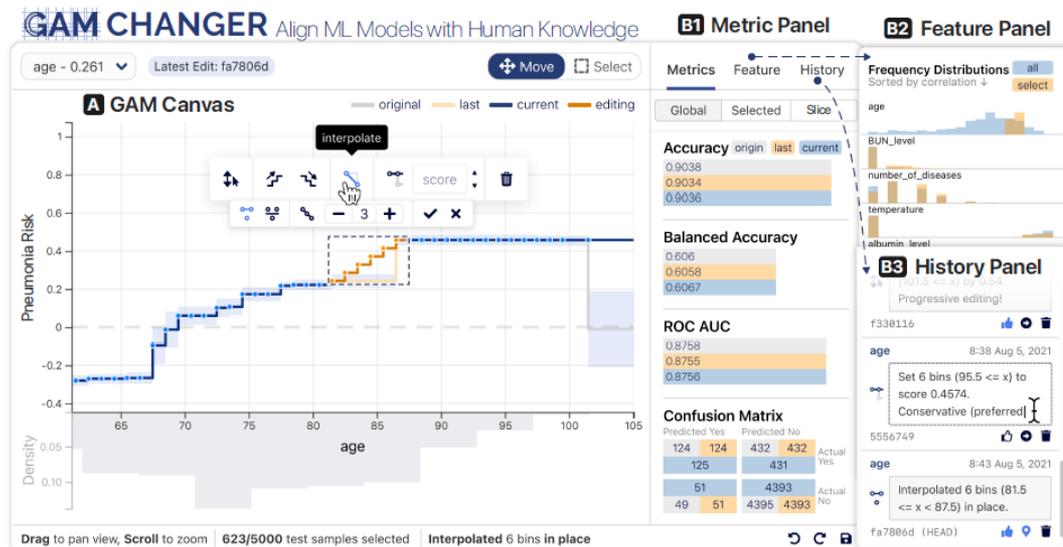
• LIMITES DE L'ÉTUDE ET PERSPECTIVES



- 4 • Présentation des résultats et des limites de l'étude

• EVOLUTIONS DEPUIS LA FIN DU MÉMOIRE

- 1 • Ajout de nouvelles fonctions de lien dans le package InterpretML dont la fonction **logarithmique**
- 2 • Ajout de nouvelles fonctions objectif dont :
 - Déviance **Poisson**
 - Déviance **Gamma**
- 3 • Possibilité de réaliser une **régression isotonique** pour appliquer des contraintes de monotonie
- 4 • Nouvelle **interface de modélisation**



Interface de modélisation GAM Changer