INTERPRÉTABILITÉ DES MODÈLES DE TARIFICATION EN ACTUARIAT

Application à l'assurance automobile

Franklin FEUKAM KOUHOUE

SOMMAIRE

Contexte: **En France, 83%.** C'est la part **des assureurs** qui considèrent que **l'IA va profondément modifier** les processus internes et la relation client. (***Source***, [ACPR], **2022**).

Problématique:

	x_1	x_2	y
Assuré 1	...	...	...
Assuré 2			...
...			...
Assuré n	...	...	...

$\boldsymbol{x_1}$ = *âge de l'assuré*
$\boldsymbol{x_2}$ = *ancienneté du véhicule*
$\boldsymbol{y}$ = *fréquence de sinistres*

Glass-Box

$$\hat{y} = f(x_1, x_2) = \boldsymbol{\beta_0 + \beta_1 x_1 + \beta_2 x_2}$$

Black-Box

$$\hat{y} = g(x_1, x_2)$$
g opaque

Apports des travaux: Dans cette présentation, nous **illustrons** comment les techniques ***d'Explainable AI*** peuvent **améliorer la précision de la tarification** en assurance tout en restant **conformes et transparentes.**

Pour l'Actuaire

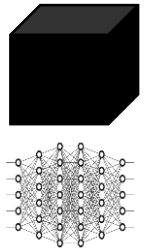
Comprendre le fonctionnement de son modèle peut lui permettre de voir ce qui fonctionne bien ou mal et comprendre pourquoi.

- Identification des **biais d'équité tarifaire** (genre, etc.)
- Identification des **incohérences** dans le modèle.

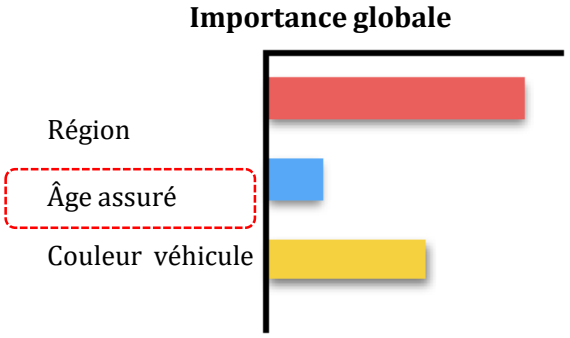
Données

ID	Âge assuré	...	Région	sinistré
1	18	...	Corse	oui
2	45	...	Corse	non
3	35	...	IDF	non
...	...	...	...	...
1000	90	...	PACA	non

Modèle de sinistre



Interprétation

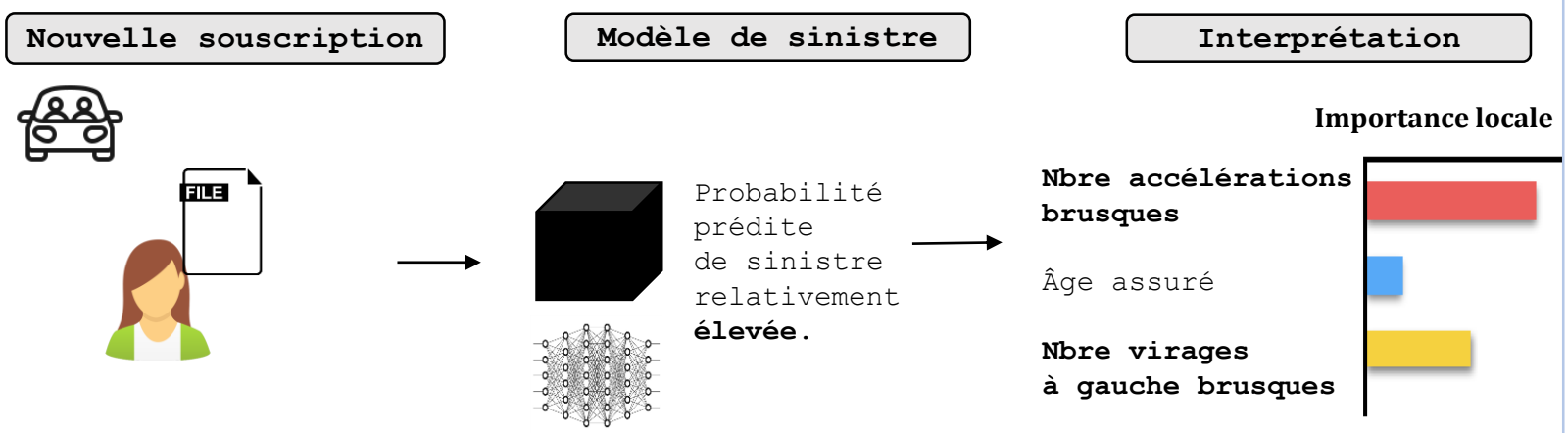


Observations

Pourquoi l'âge de l'assuré est beaucoup moins importante que la couleur du véhicule dans notre modèle de sinistre ?

Pour l'Assureur & l'Assuré

Comprendre les raisons de la hausse de la prime d'assurance.



Observations

En se servant de méthodes d'interprétabilité (par contrefactuelles), l'assureur peut formuler des stratégies d'amélioration du comportement de conduite de l'assuré. Ceci afin de réduire sa fréquence prédite de sinistre par le modèle lors de la prochaine souscription, et par ricochet la prime à payer.

Approches d'interprétabilité

Dans la littérature, nous distinguons de **deux grandes approches d'interprétabilité:**

- Interprétabilité dite **basée sur le modèle (IBM)**
- Interprétabilité dite ***post-hoc***

Approche IBM	Approche post hoc
• Parcimonie • Simulable • Modulaire	• Méthodes Globales vs Méthodes Locales • Méthodes spécifiques au modèle vs Méthodes indépendantes du modèle • Méthodes dépendantes des données vs indépendantes des données

Remarque

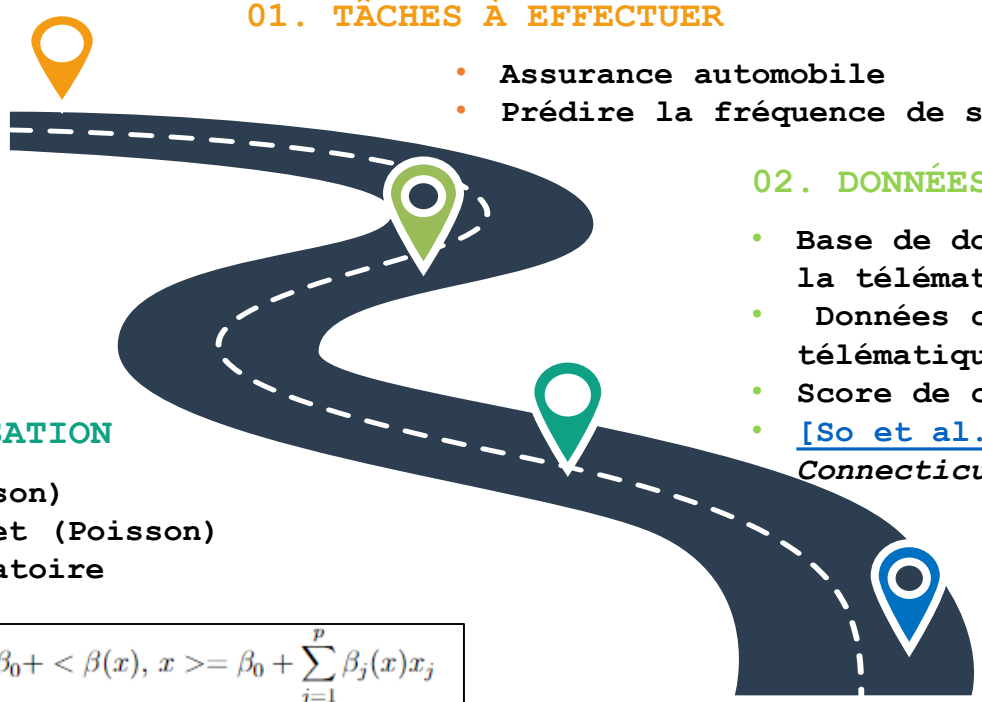
Dans le mémoire nous nous sommes focalisés sur l'approche d'interprétabilité dite *post hoc, et sur les outils* qui sont indépendants des données et indépendants du modèle à expliquer.

Etapes phares de la mise en oeuvre

Nous présentons ici les principales étapes de la mise en œuvre de ces méthodes d'interprétabilité dans notre contexte de tarification automobile.

01. TÂCHES À EFFECTUER

- **Assurance automobile**
- **Prédire la fréquence de sinistre**



02. DONNÉES DISPONIBLES

- **Base de données synthétiques pour la télématique des conducteurs.**
- **Données classiques et télématiques**
- **Score de crédit**
- **[So et al., 2021], *Université du Connecticut***

03. MODÉLISATION

- **GLM (Poisson)**
- **LocalGLMnet (Poisson)**
- **Forêt aléatoire**

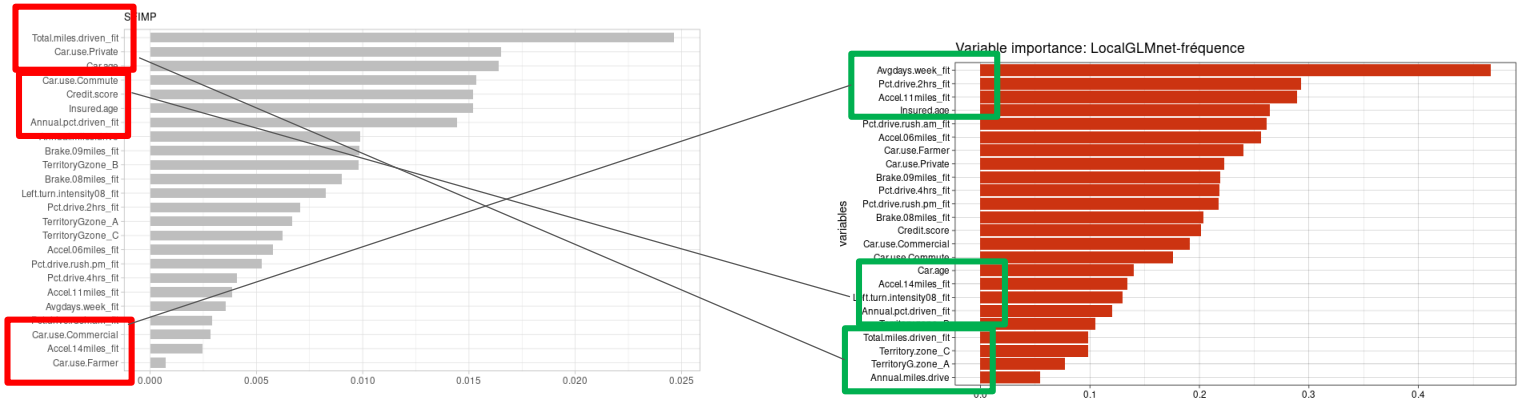
$$x \longmapsto m(\mu(x)) = \beta_0 + < \beta(x), x >= \beta_0 + \sum_{j=1}^{p} \beta_j(x) x_j$$

04. INTERPRÉTATION

- **GLM Poisson**
- **LocalGLMnet Poisson**

Importance des caractéristiques

Un premier élément que nous regardons lorsque nous souhaitons interpréter un modèle est l'importance globale des caractéristiques.



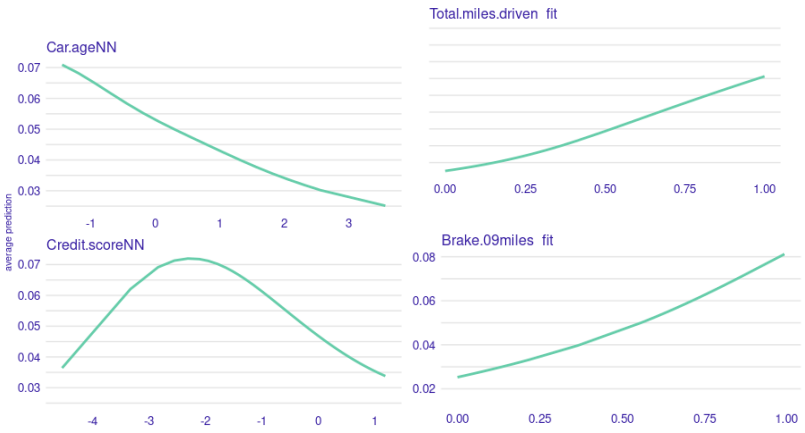
Observation

Le fait que les résultats des deux approches ne coïncident pas nécessairement est du au fait qu'ils ne reposent pas sur le même principe de fonctionnement sous-jacent.

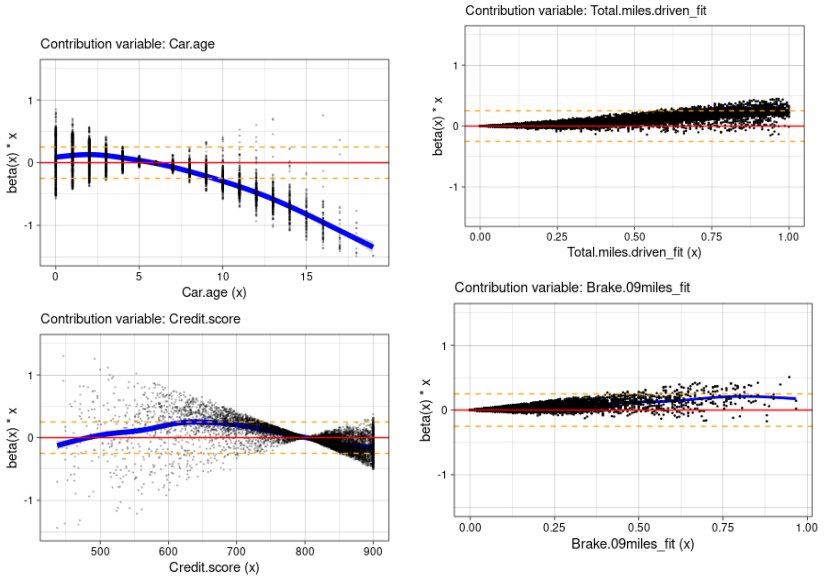
Effets des caractéristiques

Un deuxième élément que nous regardons lorsque nous souhaitons interpréter un modèle est l'effet global des caractéristiques.

ALE (Post hoc)



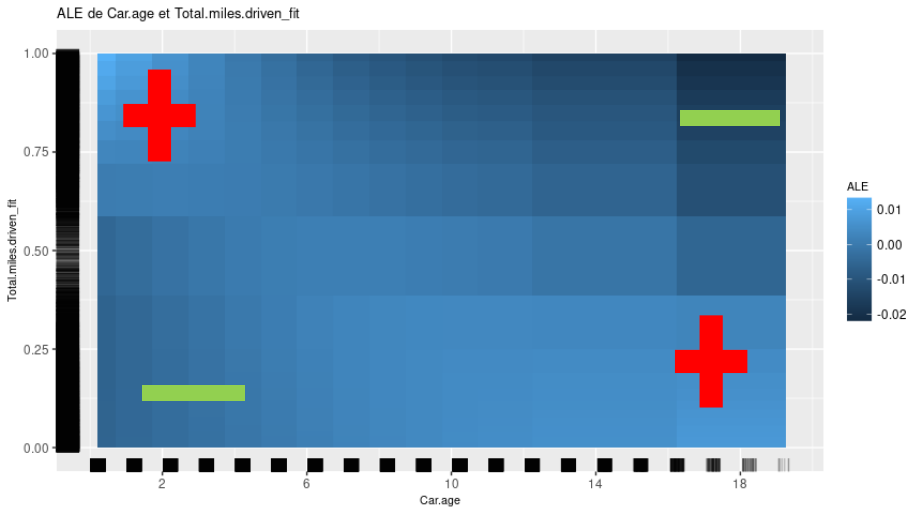
Contribution (IBM)



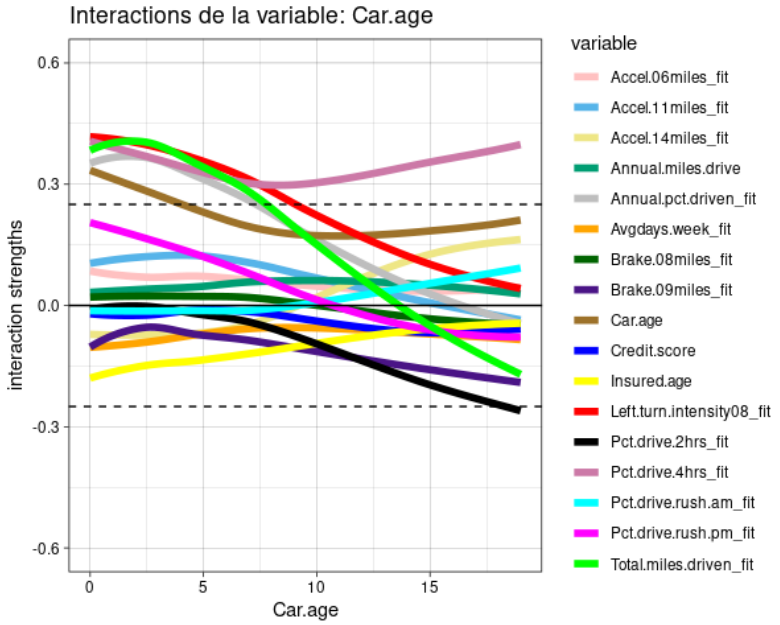
3.CAS D'APPLICATION: RÉSULTATS

Interprétation globale: Interactions entre les caractéristiques (3/3)

ALE-2D *(Post hoc)*



Interactions (IBM)



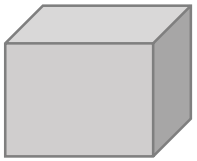
$$\nabla\beta_j(x) = \left(\frac{\partial}{\partial x_1}\beta_j(x), \cdots, \frac{\partial}{\partial x_p}\beta_j(x)\right)^\top \in \mathbb{R}^p$$

3. CAS D'APPLICATION: RÉSULTATS

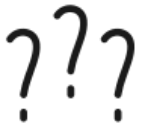
Interprétation locale: illustration (1/5)

Insured.age : **20 ans**	*Car.age* : **0 an**; *Credit.score* : **580**	*Annual.miles.drive* : **12427.2**
Car.use : **Commute**	*TerritoryG* : **zone_C**	*Total.miles.driven_fit* : **0.92**
Annual.pct.driven_fit : **0.63**	*Pct.drive.rush.am_fit* : **0.11**	*Left.turn.intensity08_fit* : **0.96**
Pct.drive.rush.pm_fit : **0.95**	*Brake.08miles_fit* : **0.95**	*Brake.09miles_fit* : **0.83**
Pct.drive.4hrs_fit : **0.90**	*Pct.drive.2hrs_fit* : **0.98**	*Accel.14miles_fit* : **0.85**
Accel.11miles_fit : **0.87**	*Accel.06miles_fit* : **0.96**	*Avgdays.week_fit* : **1**

LocalGLMnet fréquence



Fréquence prédite= **2.77**
(La plus élevée du jeu de données test)

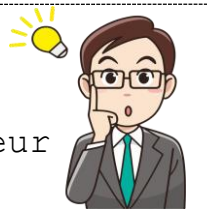


Client(e)

Régulateur

ACPR
BANQUE DE FRANCE

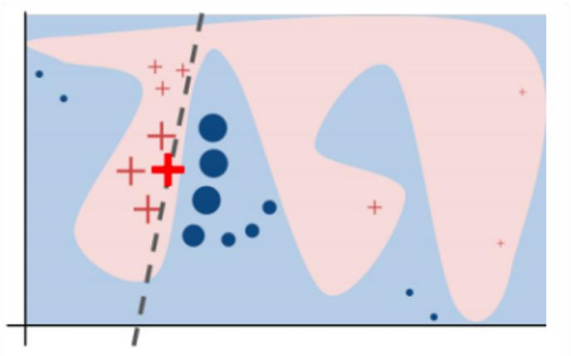
Assureur



3. CAS D'APPLICATION: RÉSULTATS

Interprétation locale: illustration (2/5)

LIME *(Post hoc)*: illustration



Intuition: Construire un modèle de substitution locale interprétable

LIME *(Post hoc)*: définition

explicateur linéaire parcimonieux

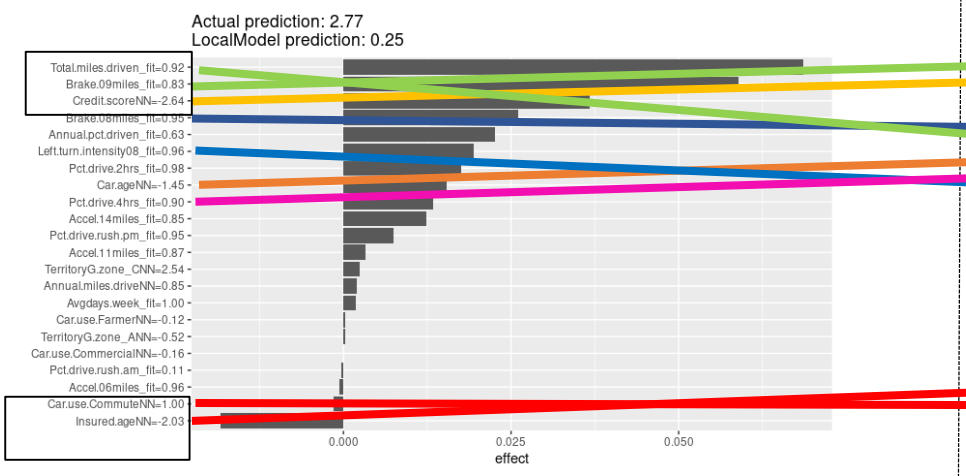
$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \; \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

- $\xi(x)$: une instance
- f : modèle à expliquer
- g : modèle interpretable
- $\mathcal{L}(f, g, \pi_x)$: mesure de l'infidélité
- π_x : mesure de proximité pour définir le voisinage local
- $\Omega(g)$: mesure de complexité

3.CAS D'APPLICATION: RÉSULTATS

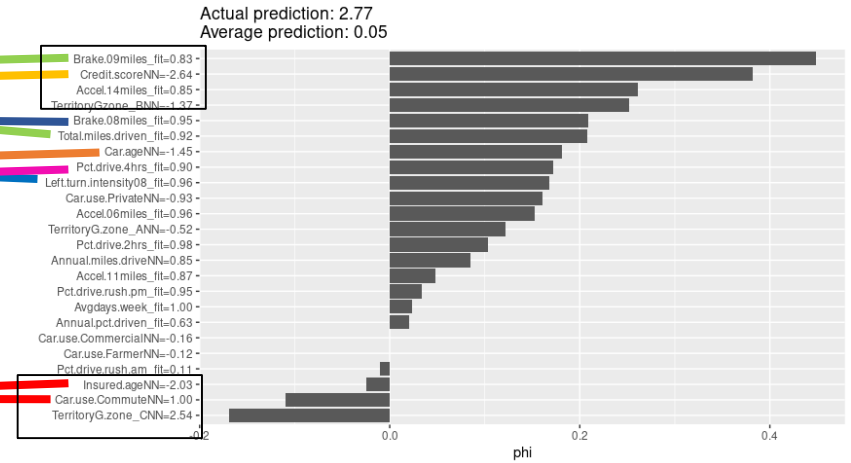
Interprétation locale: illustration (3/5)

LIME ***(Post hoc)***



- **Total.miles.driven_fit, Brake.xxmiles** et **Credit.scoreNN** parmi celles qui contribuent le plus **positivement.**

SHAP ***(Post hoc)***

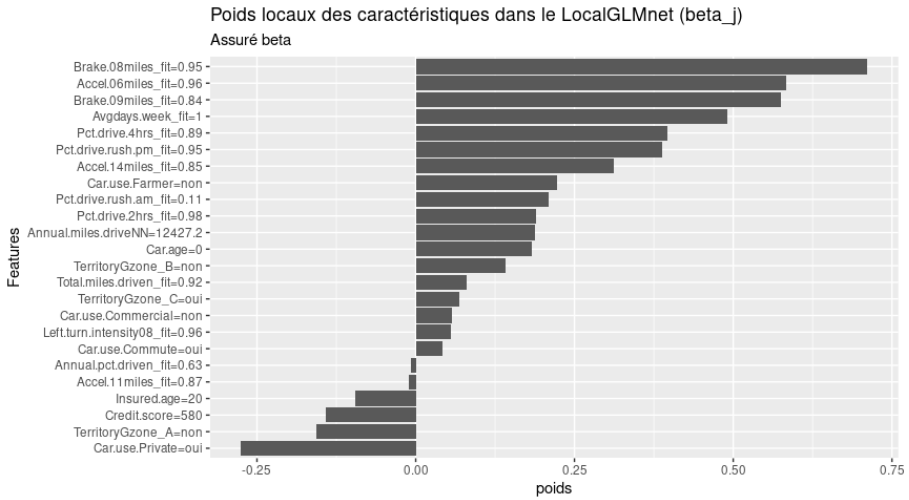


- **Insured.ageNN** contribue **négativement.**

3.CAS D'APPLICATION: RÉSULTATS

Interprétation locale: illustration (4/5)

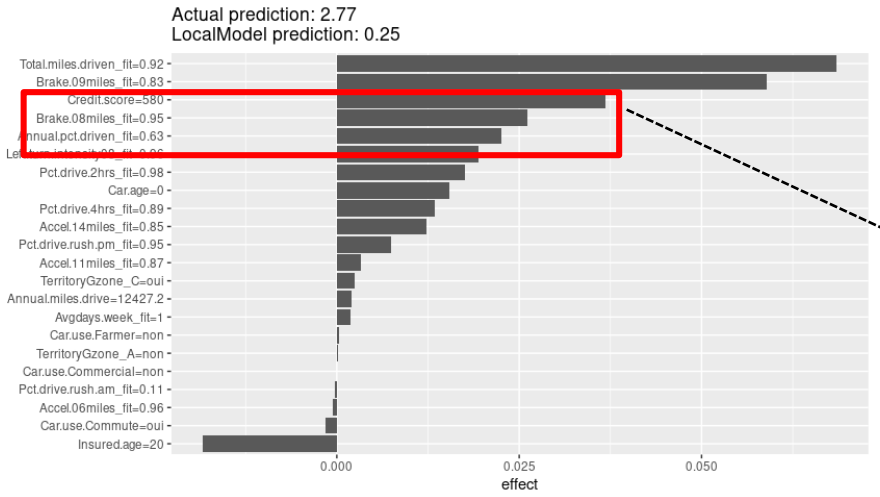
Interprétation Basée sur le Modèle (IBM)



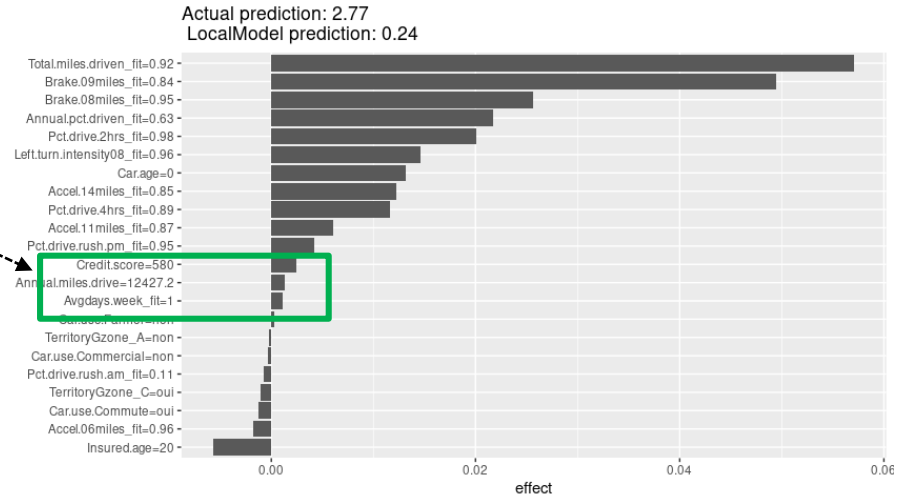
Les interprétations post-hoc sont globalement cohérentes avec l'interprétation basée sur le modèle, à quelques petites exceptions près.

Interprétation locale: limites de la méthode LIME (5/5)

Explications LIME *initiales*



Explications LIME attaquées



$$e(x) = \begin{cases} f(x), \text{ si } is\_OOD(x) \leq 0.8 \\ \psi(x), \text{ si } is\_OOD(x) > 0.8 \end{cases}$$

En dehors du poids de la variable **Credit.score** qui a été modifié, les explications sont quasiment restées inchangées par ailleurs.

3.CAS D'APPLICATION: RÉSULTATS

Ingénierie des caractéristiques: principe (1/2)

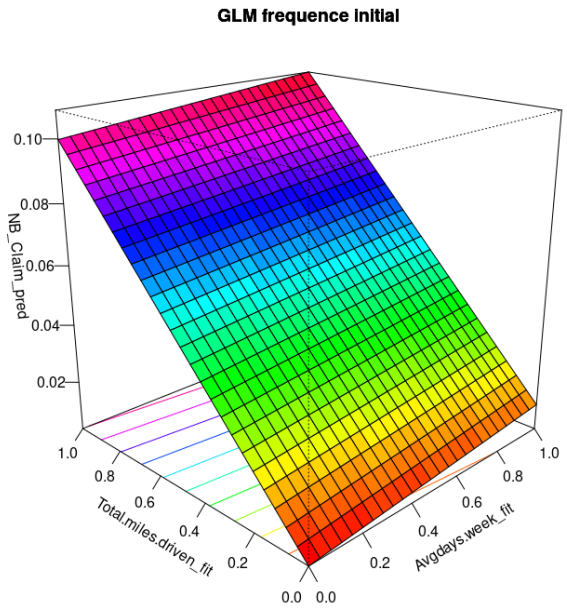
Observation

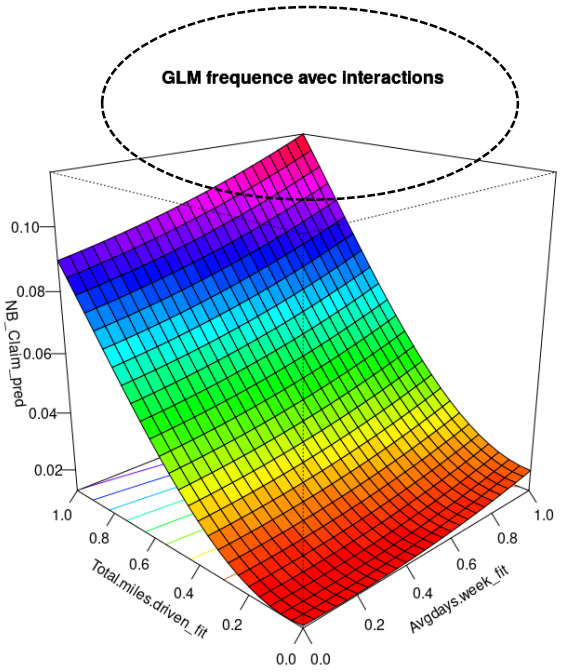
L'interprétabilité peut permettre de réaliser de l'ingénierie des caractéristiques

Ingénierie des caractéristiques: résultats (2/2)

Cas pratique

Sur ces graphiques nous illustrons un cas pratique où nous nous servons des résultats de l'interprétation de notre modèle hybride LocalGLMnet pour enrichir le modèle GLM initial.





LIMITES ET PERSPECTIVES DE L'ÉTUDE

- Mettre en œuvre les explications par contrefactuelle;

- Il est difficile d'évaluer la qualité des interprétations et de les comparer objectivement les unes par rapport aux autres: pas de métriques pertinentes;

- Vers des modèles Hybrides (ce qui fait l'objet d'importantes de recherches scientifiques en ce moment, pas seulement en assurance);

- Automatisation des processus d'interprétation en compagnie d'assurance.

Institut des ACTUAIRES

Institut des ACTUAIRES

INSTITUT Louis Bachelier

UQÀM Université du Québec à Montréal

ENSAE IP PARIS

Merci de votre attention !

Institut des actuaires
4, rue Chauveau-Lagarde – 75008 Paris – tél. 01 44 51 72 72
www.institutdesactuaires.com