

Covid-19 et assurances :

Vers une modernisation des méthodes de tarification ?

#BeyondCovid19

EDITORIAL

Comme peut l'illustrer la crise du Covid19, nous sommes dans une société où l'information est centrale et nos actions en sont très dépendantes. En effet, la mauvaise connaissance que nous avons du virus lors de son apparition ne permettait pas de prendre les mesures de protection nécessaires. C'est ainsi que les actions collectives et individuelles se sont adaptées au même rythme que la connaissance du virus et de ses caractéristiques progressait.

Plus généralement, ce lien entre **information** et **comportement** est omniprésent. Les nouvelles technologies rendent en effet possible l'absorption, la gestion et le traitement des données de masse. Ce traitement permet de créer des normes, des objectifs, voire des obligations, et leur atteinte nécessite bien souvent une adaptation de chacun. Ces changements de comportement génèrent alors de nouvelles informations, qui sont à leur tour traitées et viendront nourrir nos prises de décisions futures.

Ce phénomène évolutif n'est pas nouveau, mais fortement accentué par la croissance exponentielle des données disponibles, qui ne pourrait en être qu'à ses débuts. En effet, en 2019, **plus de 25 milliards d'objets connectés**, de tout type, étaient en circulation, et des estimations prévoient, dans un contexte de mise en place de la 5G, la multiplication par trois de ce nombre à l'horizon 2025 [1]. Autant de sources nouvelles susceptibles de générer de l'information personnelle et comportementale.

Comment se positionnent les assureurs face à ces évolutions ?

Le contexte numérique actuel ainsi que les tensions économiques et concurrentielles accentuent les besoins de prédiction pour les entreprises. C'est pourquoi les algorithmes complexes, répondant à ce besoin spécifique, se sont multipliés et tendent même à se généraliser dans certaines industries. Les assureurs ne sont d'ailleurs pas épargnés, notamment par leur spécificité de fonctionnement (cycle de production inversé), et la crise sanitaire actuelle n'en sera très certainement qu'un accélérateur.

En effet, dans la tourmente, des arbitrages ont dû être faits entre contrainte de solvabilité et volonté de solidarité, mais la réelle préoccupation qui sous-tendait les décisions prises concernait la tarification des produits et sa capacité à absorber les chocs de la crise. Une fois la situation d'urgence passée, cette question de fond devrait animer le secteur, de nombreux chantiers sont d'ailleurs déjà initiés ou en passe de l'être au sein de certains acteurs de la place.

Serait-ce le signe annonciateur d'un profond changement de la profession ?

En tout état de cause, la réponse attendue des assureurs ne peut faire fit de l'environnement réglementaire et des réalités du terrain.

Données de masse et avancées technologiques, a priori les principaux ingrédients sont réunis pour construire une réponse au **besoin de précision tarifaire**. Cependant, la réglementation concernant notamment le traitement des données personnelles, les contraintes de transparence imposées par le régulateur ou encore les besoins de compréhension exprimés par les clients et souscripteurs sont autant d'éléments qui doivent être considérés avant toute initialisation d'un chantier de modernisation des méthodes de tarification.

Ainsi, une réponse trop disruptive pourrait mener à de nombreuses difficultés, ce qui implique la nécessité d'avancer par étapes. Dans ce cadre, une solution paraissant équilibrée consiste à optimiser les choix d'hypothèses et de variables utilisées dans les modèles usuels, afin d'en améliorer leurs performances et de répondre aux besoins de **transparence**, d'**interprétabilité** et de **respect des données personnelles** sans pour autant perdre les avantages de ces méthodes.

Le Machine Learning au service de l'amélioration de la compréhension des phénomènes.

S'ils constituent encore aujourd'hui la norme en termes de tarification, les **modèles linéaires généralisés (GLM)** se révèlent parfois limités du fait des hypothèses sur lesquelles ils reposent. A l'inverse, les modèles plus complexes sont généralement plus performants mais souffrent d'un important manque d'interprétabilité.

Or, la singularité de l'activité d'assurance exige que les méthodes de tarification utilisées s'inscrivent dans un cadre strict permettant **transparence** et **auditabilité**. Cette contrainte forte mais nécessaire invite donc à un effort de réflexion quant à l'intégration du *Machine Learning* au sein du processus de tarification.

Heureusement, les différentes étapes nécessaires à l'obtention d'une prime pure offrent plusieurs opportunités, propices à l'intégration de méthodes innovantes, permettant de fiabiliser son tarif sans pour autant en sacrifier la transparence.

Un cas d'usage tiré de l'article [2] de *Henckaerts & al*, permet d'illustrer comment de telles méthodes peuvent être utilisées pour **challenger les jugements d'experts** réalisés dans le cadre de la discrétisation de variables continues en tarification.

L'approche méthodologique employée s'appuie notamment sur les modèles additifs généralisés (GAM) et les arbres de décision évolutifs :

Méthodologie



1 - Ajustement d'un modèle additif généralisé (GAM) sur la fréquence de sinistres et/ou le coût moyen.



2 - Utilisation d'un arbre de décision évolutif pour discrétiser la variable continue à partir de l'effet modélisé par le GAM.



3 - Confrontation des résultats obtenus à l'étape précédente avec l'expertise métier.



4 - Ajustement d'un GLM intégrant la variable discrétisée précédemment obtenue.

Les arbres de décision évolutifs

Les **arbres de décision évolutifs** (*evolutionary trees*) s'appuient sur le principe d'évolution des algorithmes génétiques afin d'influencer la construction des arbres de décision classiques.

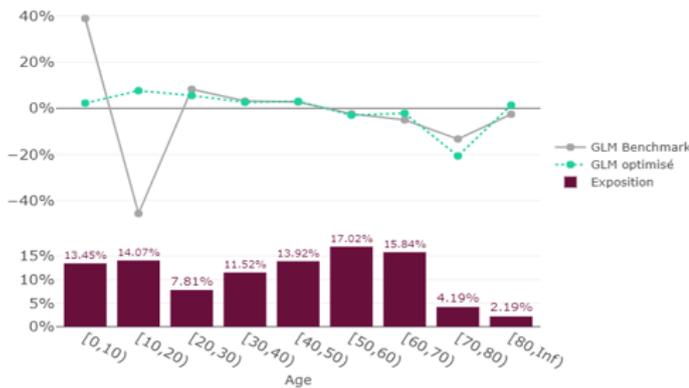
La **combinaison** de ces approches permet de modifier la structure de l'arbre au cours de sa construction : nombre de nœuds, changement de l'ordre des variables explicatives utilisées dans l'arbre, modification des seuils, ...

L'arbre résultant ne se concentre donc plus uniquement sur des choix optimaux locaux et a ainsi plus de chances de fournir une bonne approximation de l'optimum global du problème considéré.

La discrétisation obtenue permet alors une segmentation plus robuste puisqu'elle ne s'appuie plus uniquement sur la perception que peut avoir l'actuaire du phénomène modélisé mais aussi bien sur les relations observées au niveau des données que sur son expérience métier.

Ainsi, **croiser l'expertise métier à l'exploration des données permet d'affiner la modélisation du risque** comme le montre les résultats obtenus dans le cadre de la discrétisation de la variable âge dans un contexte de tarification santé [3] :

Sinistralité moyenne par âge



Un modèle plus performant

Le graphique ci-contre présente l'évolution des écarts relatifs à la sinistralité moyenne observée en fonction des classes d'âge (discrétisation arbitraire utilisée dans le GLM benchmark).

La performance du GLM benchmark surpasse celle du GLM optimisé uniquement sur la classe d'âge (70, 80) qui ne représente que 4,19% de l'exposition totale du portefeuille pour la garantie considérée.

Ainsi, le GLM optimisé permet d'améliorer de façon significative le modèle de prime pure puisqu'il offre un niveau de finesse supérieur ou égal à celui du GLM benchmark sur la quasi-totalité du portefeuille.

Sommes-nous prêts pour de tels changements ?

Bien que les perspectives offertes par le *Machine Learning* dans le cadre de la modernisation des méthodes de tarification soient nombreuses, il est important d'avoir à l'esprit que des risques sont également associés à la généralisation des algorithmes complexes, dont le processus de décision n'est pas toujours aisément transposable en langage naturel.

C'est pourquoi il apparaît essentiel que toutes les personnes impliquées dans la tarification, de la phase de conception à la phase d'exploitation du tarif, soient **formées et sensibilisées** aux algorithmes complexes, à la notion d'interprétabilité mais aussi aux **risques éthiques** induits par l'utilisation de ces modèles. Cela leur permettra ainsi de mieux appréhender l'impact de ces technologies sur leur activité mais aussi d'être acteur de la digitalisation de leur entreprise.

Enfin, le rôle du régulateur et des pouvoirs publics sera prépondérant en cette période de mutation du secteur, notamment pour l'ensemble des problématiques liées à l'éthique et à la gouvernance des données et algorithmes. Leurs actions sont toutes aussi importantes que celles des assureurs et nous ne pouvons que les encourager à travailler de concert afin de définir une stratégie d'ensemble cohérente avec les enjeux de notre époque.

Sources : [1] Statista 2019 (<https://fr.statista.com/>) ; [2] Roel Henckaerts, Katrien Antonio, Maxime Clijsters, and Roel Verbelen. (2018). A Data Driven Binning Strategy for the Construction of Insurance Tariff Classes. *Scandinavian Actuarial Journal* ; [3] L'étude technique a été réalisée sur le portefeuille d'un organisme d'assurance français.

Parlons-en !



Nicolas Servan

Senior Manager

Tél. : 06 61 79 10 90

Mail : nicolas.servan@sia-partners.com



Jordan Marie-Rose

Consultant Actuaire

Tél. : 06 50 67 64 41

Mail : jordan.marierose@sia-partners.com