

Apport de données télématiques dans la modélisation du risque géographique en assurance automobile

Auteur : Junior ASSI

Ecole : EURIA

Encadrants : Franck VERMET / Nabil RACHDI

Entreprise : Addactis France

Dans un monde qui évolue à une vitesse grand V du fait d'une digitalisation progressive et généralisée des systèmes, les assureurs en général et les assureurs automobiles en particulier dont l'activité dépend inéluctablement de l'exploitation des données, se doivent de rechercher constamment de nouveaux outils (méthodes, sources d'informations, etc) afin d'accroître, voire d'actualiser la connaissance des risques qu'ils se proposent d'assurer. En effet, avec l'émergence de nouvelles mobilités, l'utilisation de voitures beaucoup plus sophistiquées engendrant de nouvelles habitudes de conduite, une veille technologique s'impose à tous les acteurs de ce marché qui souhaitent rester compétitifs tout en assurant la pérennité de leur activité.

C'est donc dans cette quête d'outils innovants, permettant de mieux appréhender ces nouveaux risques en assurance automobile, que s'inscrivent les travaux de ce mémoire. L'étude menée met en interaction deux thèmes à fort impact dans ce domaine. D'une part, les données télématiques, révélatrices de ces nouvelles habitudes de conduite, dont l'utilité ne cesse de s'accroître depuis cette dernière décennie. D'autre part, l'aspect géographique du risque, dont la modélisation s'appuyant sur sa segmentation géospatiale est un facteur significatif de l'amélioration de la qualité des modèles tarifaires. La problématique qui émerge de cette interaction et que l'on doit résoudre est l'évaluation de la pertinence d'une modélisation du risque géographique basée sur des données révélatrices des habitudes de conduite appelées données télématiques. Autrement dit, quel est l'apport des données télématiques dans la modélisation du risque géographique en assurance automobile ?

Pour répondre à cette question, une analyse à deux volets (technique et opérationnel) est effectuée à partir de deux sources de données à savoir des données internes et des données externes. Pour cette étude, les données internes sont les bases contrat et sinistre se rapportant à la garantie responsabilité civile matérielle (RCM) automobile d'un assureur français. Quant aux données externes, la principale base utilisée est une base dénommée *Smart Road Data* et qui contient à la fois des données télématiques et des variables sur les types de routes à la maille commune. Cette base, fournie par MICHELIN, leader mondial de la fabrication et de la commercialisation de Pneumatiques, a l'avantage de contenir des variables contextualisées pour la plupart d'entre elles, de sorte à affiner les informations sur les habitudes de conduite dans les différentes communes.

Le volet technique de l'analyse menée a pour but d'évaluer les gains statistiques découlant de l'utilisation des données télématiques suivant deux différentes approches de modélisation du risque géographique. La première approche mise en place est une modélisation dite « naïve ». Dans cette approche, l'impact des comportements à l'intérieur des zones géographiques sur la fréquence de sinistres est modélisé par une combinaison de

variables télématiques reflétant les habitudes de conduite dans ces zones. Cela revient à modifier un modèle classique de fréquences de sinistres en y injectant directement des données télématiques issues d'une sélection statistique. La seconde approche consiste quant à elle, à synthétiser par un algorithme de *Machine Learning*, ces informations télématiques géographiques en une unique variable appelée zonier représentant une classification des différentes zones du territoire suivant les niveaux de risques qui leur sont rattachés. Ce zonier est par la suite intégré dans le modèle de fréquences de sinistres lui apportant cette fois une unique couche de segmentation géographique.

A l'issue de ces évaluations techniques, il ressort que l'apport des données télématiques dans la modélisation statistique du risque géographique semble permettre de mieux expliquer et discriminer ce risque et par conséquent favoriser l'amélioration de la qualité statistique et de la précision du modèle tarifaire de fréquence de sinistres.

Au niveau du volet opérationnel de l'analyse, il est question d'évaluer l'impact qu'une modélisation du risque géographique se basant sur les données télématiques peut avoir sur les tarifs proposés par l'assureur. Pour y parvenir, une analyse des variations des primes induites par le passage d'une modélisation géographique traditionnelle à une modélisation géographique dite « innovante » prenant en compte des données télématiques a été menée.

Cette analyse a montré de manière opérationnelle qu'il était possible d'ajuster les primes pures proposées par l'assureur en se basant sur les informations apportées par les données télématiques dans la modélisation innovante du risque géographique.

Toutes ces analyses semblent converger vers une unique et même conclusion : la télématique est bel et bien un outil auquel les assureurs automobiles peuvent s'adosser pour mieux appréhender le risque géographique. Ce résultat est d'autant plus encourageant qu'il met en lumière une toute nouvelle manière d'utiliser la télématique en assurance automobile. Cependant, il convient de nuancer cette conclusion pour plusieurs raisons dont le fait que l'étude a été réalisée uniquement sur une garantie RCM et sur un seul portefeuille d'assureur. Ces résultats aussi intéressants et encourageants soient-ils, méritent donc d'être investigués davantage en vue d'élargir la portée de leur pertinence.

Au-delà de la modélisation du risque géographique, les données télématiques ne pourraient-elles pas être considérées comme une solution de simplification des parcours de souscription en assurance automobile ?