

Segmentation et lissage de territoires dans un contexte d'assurance de dommages

soumis par Desjardins Groupe d'assurances générales

Notre compagnie d'assurance IARD (incendie, accidents et risques divers) a recours à des techniques de modélisation statistique afin de quantifier les risques qu'elle assure. Ces modèles nous servent donc à séparer notre clientèle en différents segments de risque. Plus le risque d'un segment est élevé, plus grande sera la prime d'assurance de ce segment.

Parmi les intrants de ces modèles, on trouve, entre autres, les caractéristiques de l'assuré (par exemple l'âge ou le nombre d'années passées comme propriétaire), les caractéristiques du bien assuré (par exemple l'année de construction, le type de toiture ou les sinistres passés), ainsi que des caractéristiques comme le type d'utilisation du bien (par exemple un bureau ou une maison unifamiliale) ou son emplacement géographique (représenté par des variables spatiales, en particulier les données du recensement de Statistique Canada). Il est bien connu que le risque varie en fonction de l'emplacement géographique (par exemple dans le cas du vol) : une bonne quantification de cet effet est primordiale.

Dans un contexte de marché canadien très compétitif, les méthodes de tarification doivent être extrêmement performantes et il faut donc quantifier le risque encouru de manière adéquate. Il est également essentiel de faire évoluer ces modèles (par exemple en ajoutant de nouvelles données) et d'explorer de nouvelles méthodologies.

Dans cette optique, Desjardins Groupe d'assurances générales désire revoir la façon dont les risques sont segmentés géographiquement. À l'heure actuelle, la compagnie utilise une combinaison de modèles de descente de gradient (GBM) et de lissage spatial basé sur les champs aléatoires de Markov, mais cette méthode est difficile à calibrer et manque de robustesse.

Le but de ce projet sera donc de construire un (ou des) modèle(s) statistique(s) s'appuyant sur les variables mentionnées ci-dessus afin de mieux prévoir le degré de risque de nos clients. Pour des raisons de cohérence (vis-à-vis des clients), ces prévisions devront satisfaire une contrainte de continuité spatiale, c'est-à-dire constituer un résultat lisse. Par exemple, toutes choses étant égales par ailleurs, deux clients qui sont voisins devraient avoir des prévisions similaires.

L'approche développée devra également être robuste et offrir des performances semblables pour différents jeux de paramètres et ensembles de données. Quant à l'optimisation d'hyperparamètres, elle devra être stable, reproductible, facile à effectuer et rapide (idéalement). La méthode devra par ailleurs pouvoir être appliquée dans le cas où il existe peu d'exposition à des risques dans un territoire donné.