



LIFE & HEALTH DATA

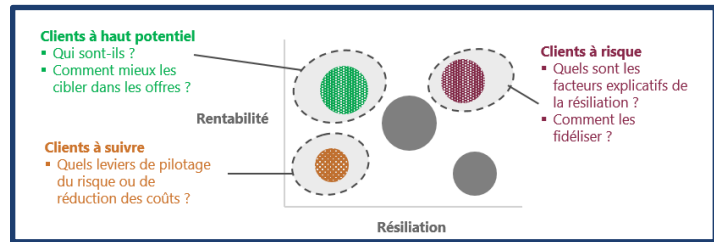
PRÉDICTION DES RÉSILIATIONS EN ASSURANCE SANTÉ

Pourquoi et comment modéliser la résiliation en assurance santé ?

Dans un contexte de démocratisation des comparateurs, les assurés cherchent le meilleur rapport qualité/prix de chacun des produits d'assurance dont la complémentaire santé. Tant sur des contrats individuels que collectifs, **la détection des comportements conduisant à la résiliation est un véritable enjeu pour les assureurs car elle permet un pilotage optimisé du risque et de leur chiffre d'affaire.**

Aujourd'hui, la Data Science apporte de véritables solutions pour déterminer les facteurs de risques de la résiliation. Pour cela, il est important d'exploiter les informations disponibles pour une compagnie d'assurance : prestations, cotisations, caractéristiques de l'assuré mais aussi les données de gestion, de contacts client, de multi-détention... Pour tenir compte de l'évolution du comportement des assurés, ces études sont généralement effectuées sur la dernière année complète à disposition.

Il faut souligner que les résiliations ne représentent en général qu'une faible partie du portefeuille (autour de 10-15% pour les contrats individuels). La modélisation doit ainsi tenir compte de ce contexte de **données déséquilibrées**. Les techniques de rééchantillonnage (SMOTE, ROSE, etc.) peuvent conduire à réduire ce déséquilibre, par exemple en dupliquant certaines données rares.



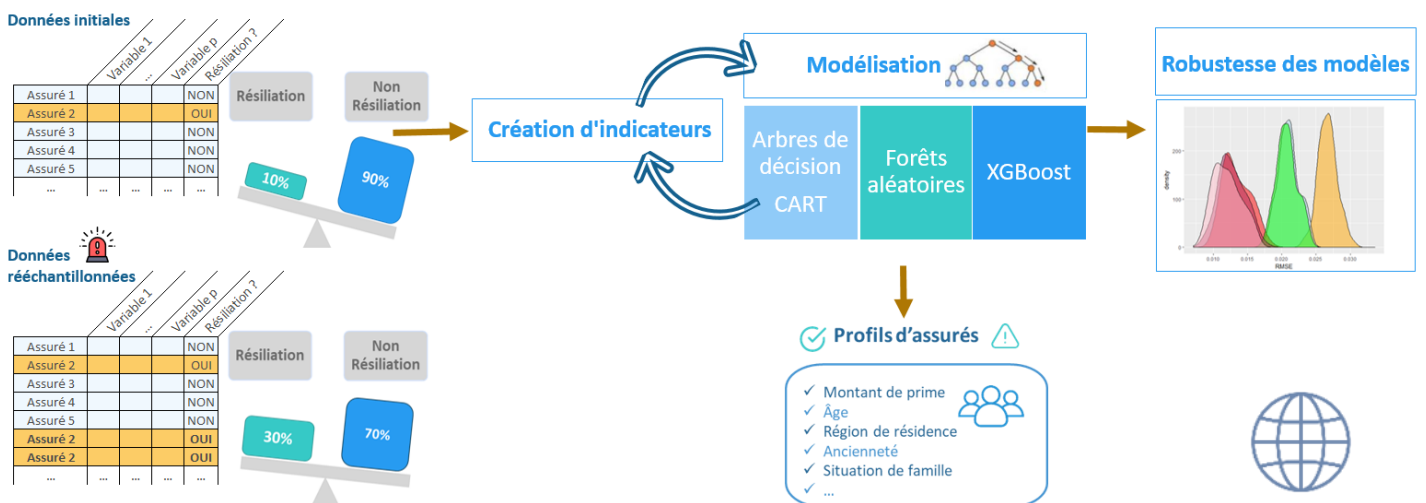
De la construction d'indicateurs à l'exploitation des modèles de machine learning.

À partir des données relatives à un contrat santé, les algorithmes de Data Science (GLM, arbres, forêts aléatoires, XGBoost...) établissent une **probabilité de résiliation** propre à chaque assuré et identifient les **variables discriminantes qualifiant la résiliation**. Il est également possible d'ajouter certains indicateurs de consommation au modèle (part de reste-à-charge moyen, consommation des forfaits de soins ...) afin de l'améliorer. **Cette étape de Feature Engineering nécessite une expertise métier pour créer les indicateurs judicieux.**

Les probabilités de résiliation obtenues par assuré permettent de définir un seuil à partir duquel on considère que l'assuré va résilier : **la probabilité devient alors une règle de décision**. Les courbes ROC et PR déterminent le seuil adapté au modèle. La matrice de confusion quant à elle sert à calibrer les différents modèles et à les comparer entre eux.

Dans un contexte de résiliation infra-annuelle à venir, **la combinaison de modèles de résiliation avec des indicateurs de rentabilité permet à l'assureur d'optimiser le ciblage d'actions commerciales ou de défense de portefeuille** (par ex : chercher à garder les clients qui résilient mais en priorité ceux qui sont les plus rentables).

PROCESS DE PRÉDICTION DES RÉSILIATIONS EN ASSURANCE SANTÉ



Cette publication a été réalisée sous la direction de Nabil RACHDI, Head of Data Science

avec l'expertise de :
Alexandra BARRAL
Eugénie POYET
Jean-Pascal HERMET