

**Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires**

le 16/11/2022

Par : **Afzal ALI ZATCHI**

Titre: **Résiliation et élasticité prix sur le produit MRC**

Confidentialité : NON OUI (Durée : 1 an 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membres présents du jury de la filière

Pierre PICARD

Entreprise : Generali France

Nom : Thibault DALY

Signature :



*Membres présents du jury de l'Institut
des Actuaires*

Directeur de mémoire en entreprise :

Nom : Thibault DALY

Signature :



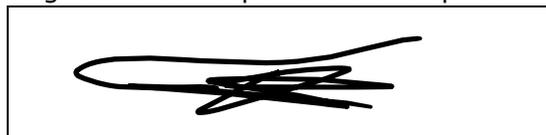
Florence PICARD

Marc JUILLARD

Jean-Marie NESSI

**Autorisation de publication et de
mise en ligne sur un site de
diffusion de documents actuariels**
*(après expiration de l'éventuel délai de
confidentialité)*

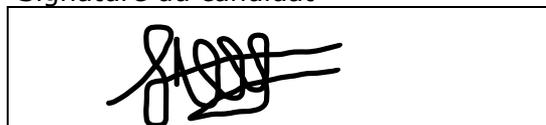
Signature du responsable entreprise



Secrétariat :

Signature du candidat

Bibliothèque :



Résumé

La multirisque commerce MRC de GENERALI est un bon produit d'appel sur le marché des professionnels. Elle protège les commerces des professionnels et des entreprises contre des risques divers. La MRC représente uniquement 10% du Chiffre d'affaires de GENERALI France en IARD mais ce qui rend intéressant ce produit c'est la nature de ses clients avec un fort potentiel à se multi équiper (34% de multi équipement). Le marché étant très concurrentiel, la segmentation tarifaire doit être accompagnée d'une politique de fidélisation des clients à potentiels pour une bonne stabilité et meilleure rentabilité. L'équipe client de GENERALI oriente l'étude de la rentabilité sur un aspect centré sur le client et à travers un indicateur clé qui est la valeur client afin d'harmoniser les différentes actions à mener lors de l'acquisition et durant le parcours du client chez GENERALI. La valeur client est un bon indicateur de la rentabilité actuelle et prospective du client et permet de choisir les profils à fidéliser. Dans le processus de fidélisation, à chaque renouvellement, il faut anticiper les sensibilités des clients à la revalorisation de leurs primes, à plus forte raison les clients créateurs de valeur. Ce mémoire a pour objectif d'étudier les résiliations sur le portefeuille MRC et d'analyser les sensibilités à la majoration de la prime. Ce qui permettra dans un premier temps d'identifier les profils les plus sensibles, d'élaborer des scénarios de revalorisation et de mettre en place des scores de fragilités à l'intention des intermédiaires pour les sensibiliser sur le risque de départ des clients. Dans un deuxième temps, l'étude permettra de refondre la valeur client et de choisir les clients à plus fortes valeurs à défendre contre les fortes revalorisations.

Mots clés : valeur client, renouvellement, revalorisation, majoration, MRC, Segmentation, Résiliation, élasticité, GLM

Abstract

GENERALI's commercial multi-risk MRC is a good appeal product on the professional market. It protects the businesses of professionals and companies against various risks. The MRC represents only 10% of GENERALI France's P&C premiums, but what makes this product interesting is the nature of its customers, with a strong potential for multi-equipment (34% of multi-equipment). The market being very competitive, the pricing segmentation has to go with a retention strategy for high potential clients for a good stability and better profitability. GENERALI's customer team focuses the profitability study on a customer-centric aspect and through a key indicator which is the customer value in order to harmonize the different actions to be carried out during the underwriting and during the customer's experience with GENERALI. The customer value is a good indicator of the current and prospective profitability of the client and enables the selection of profiles to be retained. In the loyalty process, at each renewal, it is necessary to anticipate the sensitivities of the customers to the revaluation of their premium, and especially the customers who create value. The objective of this memoire is to study the cancelations on the MRC portfolio and to analyze the sensitivities to the increase of the premium. This will first allow us to identify the most sensitive profiles, to develop scenarios for premium increases and to set up scores for intermediaries in order to sensitize them to the risk of client termination. Secondly, the study will allow us to review the client value and to choose the clients with the highest values to defend against high revaluation.

Key words: customer value, renewal, revaluation, mark-up, MRC, segmentation, cancellation, elasticity, GLM

Remerciements

Je voudrais exprimer ma plus profonde gratitude à **Thibault DALY** et **Céline CONAN** pour m'avoir donné l'opportunité d'apprendre au sein de l'Equipe Valeur Client de GENERALI et à mon manager **Hamza EL HASSANI** d'avoir rendu cela possible et pour l'accueil qui me fut réservé et la convivialité de nos échanges. Je remercie spécialement **Alice BELLOT**, **Elena TROCCOLI** et **Yacine ABOUTAYBI** pour leurs précieux conseils et directives qui m'ont aidé dans les choix des méthodes employées afin de satisfaire les objectifs de l'étude.

Je tiens à remercier également Monsieur **Pierre PICARD** pour sa disponibilité et son accompagnement.

Sommaire

Résumé.....	ii
Abstract.....	iii
Sommaire.....	v
Introduction.....	1
Partie 1. Cadre de l'étude.....	2
1.1. Département Valeur Client.....	2
1.2. Valeur Client.....	2
1.3. Produit MRC.....	7
1.4. Utilité de l'étude : Modélisation de l'élasticité à la majoration sur le produit MRC :	9
Partie 2. Données et études préliminaires.....	10
2.1. Objectif.....	10
2.2. Variables explicatives.....	11
2.2.1. Données de l'intermédiaire.....	11
2.2.2. Données du contrat.....	11
2.2.3. Données du client.....	12
2.2.4. Données sur le risque.....	12
2.2.5. Données d'interaction.....	12
2.3. Base de données.....	12
2.4. Analyse univariée.....	14
2.5. Analyse multivariée.....	21
Partie 3. Modélisation.....	24
3.1. Modélisation de la résiliation.....	24
3.1.1. Modèles Linéaires généralisés.....	24
a. Rappels Théoriques GLM.....	24
b. Rappels Théoriques sur la régression logistique.....	26
c. Validation des modèles.....	26
d. Application aux données.....	30
3.1.2. Modèles additifs généralisés (GAM).....	36
a. Rappels théoriques.....	36
b. Application.....	38
3.1.3. Gradient Boosting.....	43
a. Rappels théoriques sur les Arbres de décision.....	43
b. Rappels théoriques sur le Gradient Boosting.....	43
c. Applications.....	46

3.1.4. Conclusions sur les 3 modèles :	48
3.2. Modèle d'élasticité.	49
3.2.1. Rappels théoriques	49
3.2.2. Application aux données	50
3.3. Conclusions sur les modèles :	52
3.4. Modèles par produit.	53
3.4.1. Produit 100% Pro Artisans Commerçants	53
3.4.2. Produit 100% Pro Services	59
3.4.3. Produit 100% Pro ACPS	66
Applications possibles de l'étude	70
Conclusions	71
Bibliographie	72
Annexes	74
Note de synthèse	II
Executive Summary	V

Introduction

Chaque année, les assureurs sont confrontés lors du renouvellement des contrats d'assurance à la question récurrente de réévaluation des primes. Suivant les taux de revalorisation que l'assureur choisit, il fait le compromis entre ses marges plus importantes à travers les fortes revalorisations et la stabilité de ses parts de marché à travers les faibles revalorisations. C'est un compromis à trouver d'autant plus que le secteur assurantiel est très concurrentiel.

L'assouplissement par ailleurs des règles de détachement du client depuis la Loi Hamon représente une contrainte supplémentaire sur le choix de la politique de l'assureur lors du renouvellement des contrats. En effet, la Loi Hamon du 17 mars 2014 offre la possibilité de résiliation, sans frais et sans pénalité et donc sans contraintes des assureurs, de certains contrats d'assurances et des adhésions tacitement reconductibles à l'expiration d'un délai d'un an. Autrement dit, cette loi confère aux assurés le droit de résilier leur contrat passé un an.

Face à un tel environnement, l'assureur se doit de mettre en place une politique de ciblage de ses clients pour une meilleure rentabilité et stabilité de son portefeuille. Le choix des revalorisations s'inscrit aussi dans une politique de fidélisation des clients qui à court ou long terme sont des clients rentables. Une façon de répondre au choix des clients potentiellement rentables est l'utilisation d'un indicateur de rentabilité comme la valeur client.

La valeur client prend en compte la rentabilité actuelle et potentielle du client et centre l'interprétation de la rentabilité d'un point de vue client plutôt que produit. Elle permet donc de piloter le parcours du client chez l'assureur, entre autres la protection des clients contre les fortes revalorisations, l'harmonisation des revalorisations des différents contrats du client.

Un autre paramètre à prendre en compte lors du renouvellement est l'inflation qui est de nouveau très importante. Il faut préciser que les majorations des primes se font principalement au gré de l'inflation. Le choix des revalorisations doit permettre de couvrir les sinistres dont les coûts sont de plus en plus importants.

Anticiper les résiliations par suite d'un changement de la prime représente la problématique de l'élasticité au prix. La problématique de l'élasticité au prix se pose au sein de l'Equipe Client de GENERALI pour le renouvellement des contrats de la multirisque commerce MRC. La multirisque commerce est l'assurance des professionnels et représente un marché très intéressant pour le profil de client qu'il fait entrer chez GENERALI. Comment utiliser les données pour identifier les clients plus sensibles que d'autres afin de piloter les revalorisations ? C'est l'objectif de ce mémoire qui est d'étudier les résiliations et les élasticités prix des contrats de la multirisque commerce GENERALI. Dans ce mémoire, nous faisons également une analyse en profondeur des diverses familles d'activités de la MRC vis-à-vis du risque de résiliation.

Dans la suite de ce mémoire, une première partie sera consacrée à la présentation de la valeur client, ses composantes et le lien avec notre étude, ensuite dans une deuxième partie nous délimiterons notre périmètre d'étude et les données utilisées. Dans une dernière partie nous présenterons les différents modèles construits sur le portefeuille MRC complet et sur les sous-produits MRC.

Partie 1. Cadre de l'étude

1.1. Département Valeur Client

Partie intégrante de la Technique Assurance de GENERALI, l'équipe se charge principalement de la production d'un indicateur clé : la valeur client. Les différentes missions de l'équipe sont la valorisation des contrats et des clients, la production de scores de pilotage de type évolution du portefeuille, multi équipement, la mise en place de modèle de tarification, d'élasticité pour le renouvellement des tarifs.

1.2. Valeur Client

La valeur client dénote une nouvelle vision de la rentabilité du client. Elle permet d'avoir une vision consolidée de la rentabilité des clients ayant plus d'un contrat GENERALI. Complexe comme vision de la rentabilité, il est important d'envisager toutes les positions actuelles et futures du client. Ainsi, il est envisageable qu'une mauvaise rentabilité sur certains produits peut à long terme favoriser une très bonne rentabilité dans le sens où cette situation va permettre de fidéliser le client et de lui faire monter en équipement et ou en gamme. On peut alors piloter les coûts d'acquisition et les rabais commerciaux à accorder à la souscription en déterminant leurs limites supérieures.

La valeur client permet donc de rendre compte des avantages de la rétention des clients et la nécessité de fidéliser les clients pouvant à long terme apporter de fortes rentabilités.

La valeur client se compose de deux parties : la valeur à date ou actuelle et la valeur potentielle

- La valeur Actuelle

Elle correspond à l'espérance actualisée des profits techniques et financiers du contrat. Il est donc nécessaire de projeter les 2 marges technique et financière sur un terme très variable suivant le produit et le profil client considéré et pouvant atteindre 30ans.

Le résultat technique : C'est la projection des primes et des coûts (frais et sinistres) pour l'année i considérée.

$$RT_i = Primes_i - Sinistres_i - Frais_i$$

où

$$Primes_i = Primes_{i-1} * (1 + T\chi_i^{\text{croissance}}) * (1 - T\chi_i^{\text{sortie}})$$

avec

$$Primes_1 = Primes$$

$T\chi_i^{\text{evolution}}$ la probabilité sachant le contrat en cours durant l'exercice i-1 de résilier en i et
 $T\chi_i^{\text{croissance}}$ un taux de croissance de la prime qui prend en compte toutes les majorations probables de la prime.

Ces indicateurs sont fournis par des Bureaux d'Etudes Techniques. Pour la sinistralité, la projection dépend de l'évolution de la prime et de la résiliation également. Les primes sont fixées et les montants n'évoluent qu'à travers les résiliations et les majorations (beaucoup plus dépendantes de l'inflation).

Afin de garder une projection robuste, les S/P sont utilisés pour déterminer $Sinistres_i$ selon la relation suivante :

$$Sinistres_i = Sinistres_{i-1} * (1 + Tx_i^{inflation}) * (1 - Tx_i^{sortie})$$

Avec

$$Sinistres_1 = S/P * Primes$$

Les frais regroupent les frais de gestions et les commissions relativement stables et indépendant de l'année d'évaluation.

$$Frais_i = Primes_i * (Tx^{FGCOM})$$

avec

Tx^{FGCOM} le taux de frais de gestion et de commission.

Le résultat technique est donné par la formule suivante :

$$RT = \sum_{i=1}^{\max.projection} (RT_i * Déflateur_i * (1 - Tx_i^{impots}))$$

Le résultat financier est basé sur la provision ; montant provisionné et investi suivant la formule :

$$Provision_i = \sum_{j=1}^i (Tx_j^{PSAP} * Sinistres_{i-j+1})$$

avec

Tx_j^{PSAP} le taux pour sinistres à Payer correspondant au cout total estimé de tous les sinistres survenus jusqu'à l'exercice net des sommes déjà payées pour ces sinistres.

Le résultat financier est donné par la formule :

$$RF = \sum_{i=1}^{\max.projection} (Provision_i * Tx_i^{Rendement} - MargeSolva_i) * Déflateur_i * (1 - Tx_i^{impots}) * (1 - Tx_i^{sortie})$$

Où $Tx_i^{Rendement}$ est le taux de rendement des primes réinvesties.

$MargeSolva_i$ est l'exigence de capital pour sécuriser la position financière de l'entreprise vis à vis de ce risque.

$$MargeSolva_i = Primes_i * (Tx^{MargeSolva}) * (Tx^{Cot.Capital})$$

avec

$T\chi^{\text{MargeSolva}}$ le taux règlementaire de marge de solvabilité et

$T\chi^{\text{Cot.Capital}}$ le cout du capital correspondant à la rentabilité annuelle attendue par les actionnaires et les créanciers.

La valeur actuelle du contrat va correspondre à la somme du résultat technique et du résultat financier.

$$\text{Valeur. Actuelle} = RT + RF$$

Pour un client multi équipé, détenteur de plusieurs contrats, la valeur actuelle du client correspond à la somme des différentes valeurs actuelles des contrats.

- **La valeur potentielle**

C'est l'espérance de la valeur de chaque contrat futur que le client suivant son profil va potentiellement souscrire. L'historique des contrats et des clients permet d'établir le niveau d'appétence à se multi équiper suivant chaque profil du client. Ce qui permet de limiter les contrats potentiels à un nombre limité de produits. Par exemple, l'entrée avec un contrat multirisque commerce peut-être une voie d'entrée également pour les produits Auto, Epargne et Prévoyance par exemple pour les professionnels.

$$\text{Valeur. Potentielle} = \sum_{\omega \in \Omega} p_{\omega} * VC_{\omega}$$

avec

p_{ω} la probabilité de souscrire à un produit ω et

VC_{ω} la valeur espérée du produit ω .

Il faut donc modéliser la duration client et l'appétence client à souscrire à d'autres produits pour déterminer les différentes probabilités sachant le profil client de souscrire à d'autres produits.

La valeur client est donc :

$$\text{Valeur. Client} = \text{Valeur. Actuelle} + \text{Valeur. Potentielle}$$

La valeur client telle que définie est un montant en Euro. La valeur potentielle est une valeur très discutable étant donné la difficulté et la complexité de prédire l'évolution du portefeuille du client. Les deux composantes sont calculées mensuellement pour chaque contrat mais la valeur actuelle constitue la principale composante de la valeur client. Nous précisons que la valeur actuelle d'un client détenteur de plusieurs produit n'est qu'une addition des différentes valeurs actuelles de ses contrats.

Diffusion de la valeur

La valeur client permet d'avoir une vision actuelle et prospective de la rentabilité client sur tous les produits détenus par le client. De la même manière, les comportements des clients n'étant pas indépendants, il est important de prendre en compte l'effet que le comportement d'un client peut provoquer sur un autre en particulier ceux appartenant à un même entourage. C'est dans cette logique que la notion de valeur client est étendue au-delà d'un client à la sphère. La sphère d'un client va correspondre à l'entourage le plus proche du client qui peut être :

- Lien matrimonial (conjoint ou conjointe de)

- Lien filial (parent/enfant)
- Lien entreprise

D'autres liens peuvent être envisagés mais compte tenu de la complexité que cela pourrait engendrer de les étudier, nous allons nous limiter à ces trois liens.

De la même manière qu'un client peut à court terme avoir une mauvaise rentabilité, si à long terme cela peut permettre de fortes rentabilités alors ce client est désirable. Le même raisonnement peut être appliqué à la notion de sphère dans la fidélisation des clients appartenant à la même sphère. On peut ainsi classer les clients de la même sphère entre les promoteurs et les détracteurs respectivement ceux qui avantagent la sphère et ceux qui la desservent. L'idée de la diffusion de la valeur dans la sphère va constituer à faire profiter des avantages d'une sphère à tous les membres ou nouveaux adhérents de la sphère. Par exemple, considérons une sphère de 2 personnes. Si un nouveau client intègre cette sphère par le lien filial, alors la diffusion de la valeur va consister à faire bénéficier à cette personne des avantages de sa sphère. Ainsi, un rabais plus ou moins important peut être accordé à la souscription, un traitement privilégié peut être donné au client à travers les interactions d'avec l'entreprise.

Une classification de la valeur client (Valeur Actuelle) permet de la rendre plus fonctionnelle. On distingue 5 classes de la valeur suivant le tableau de répartition ci-après :

TABLEAU 1 : CLASSIFICATION DES CLIENTS SUIVANT LEUR VALEUR ACTUELLE.

CONDITION	STANDING
Valeur actuelle < 0€ sans potentiel de multi-équipement	1*
Valeur actuelle < 0€ avec potentiel de multi-équipement	2*
Valeur actuelle entre 0€ et 1600€ sans potentiel de multi-équipement	3*
Valeur actuelle supérieure à 1600€ sans potentiel de multi-équipement	4*
Valeur actuelle entre 0€ et 750€ avec potentiel de multi-équipement	
Valeur actuelle >750€ avec potentiel de multi-équipement	5*

Les profils 4* et 5* sont les plus rentables et donc les clients à plus forte valeur : ils sont dits clients à valeur. La valeur client va permettre de cibler ces profils afin de leur offrir un niveau de service client plus élevé et une priorité à travers les interactions d'avec l'entreprise. La valeur client d'une sphère va permettre dès la croissance de cette sphère de faire profiter à l'entrant du standing des

autres clients déjà existants dans la sphère et de tous les avantages qui vont avec. Les clients à valeur bénéficient de rabais plus importants à la souscription d'autres produits et sont aussi protégés contre les fortes revalorisations des primes lors des renouvellements. Il est donc très important de déterminer en amont la sensibilité au prix des clients, et de se servir de la valeur comme deuxième couche de protection pour adapter les niveaux de majoration au profil client.

Elasticité prix et Valeur client :

La valeur client telle que précédemment définie est croissante en fonction de la prime.

$$\text{Primes}_i = \text{Primes}_{i-1} * (1 + T x_i^{\text{croissance}}) * (1 - T x_i^{\text{sortie}})$$

Plus le taux de croissance de la prime est important et plus la valeur actuelle va être très importante. Mais ce fort taux va aussi dégrader la rétention du client. Le client préférera toujours payer moins cher d'où l'importance de contrôler les majorations en prenant en compte la rétention du client.

1.3. Produit MRC

La Multirisque Commerce est l'assurance des biens et de l'activité de l'indépendant. Elle est destinée aux professionnels et entreprises souhaitant couvrir leurs commerces. La multirisque professionnelle comprend les couvertures suivantes :

- La protection des biens contenus dans le local en cas de sinistre
- La responsabilité civile professionnelle de l'indépendant et de ses salariés
- Les conséquences financières d'un sinistre avec la garantie perte d'exploitation

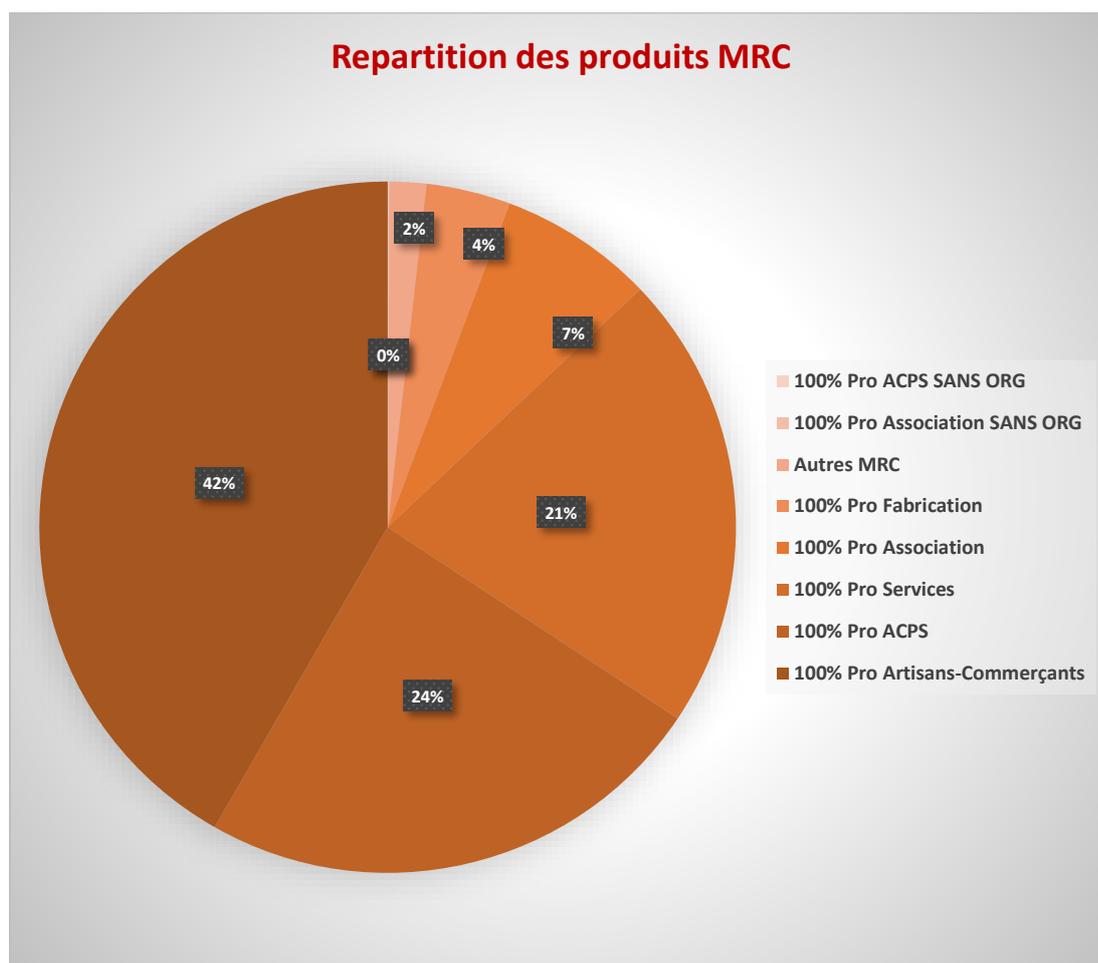


FIGURE 1 : LES DIFFERENTS PRODUITS MRC DE GENERALI

Les produits MRC peuvent être regroupés comme suit :

- Le Pro Artisans Commerçants : destinés au Artisans-Commerçants
- Le Pro Fabrication : une gamme spéciale destinée aux Artisans Commerçants aussi
- Le Pro Services : Pour principalement les bureaux
- Le Pro Associations : destinés aux associations
- Le Pro ACPS : produits de l'ancienne générations destinés à la fois aux Commerçants Artisans et les métiers de services. Ce produit n'est plus commercialisé mais les détenteurs représentent une part significative du portefeuille MRC.

Les produits sont distribués à travers deux réseaux Generali pour les professionnels : les réseaux agents et les réseaux courtiers. Les produits MRC représentent 151000 contrats pour une prime moyenne de 1200€ en 2022. Les professionnels ont l'un des taux de multi équipement les plus élevés du portefeuille de GENERALI avec 34% des clients multi équipés, c'est-à-dire qu'ils possèdent plus d'un contrat. En effet, les professionnels commencent par des contrats multirisques professionnels avant de monter en gamme sur d'autres produits GENERALI. Tout comme l'Auto, le produit MRC est un excellent produit d'appel des Professionnels chez GENERALI. Les autres contrats détenus par les clients MRC sont :

- Auto (21%),
- Multirisque Habitation (7%),
- Responsabilité civile (11%),
- Prévoyance santé (7%),
- Retraite individuelle (5%),
- Construction (5%) ...

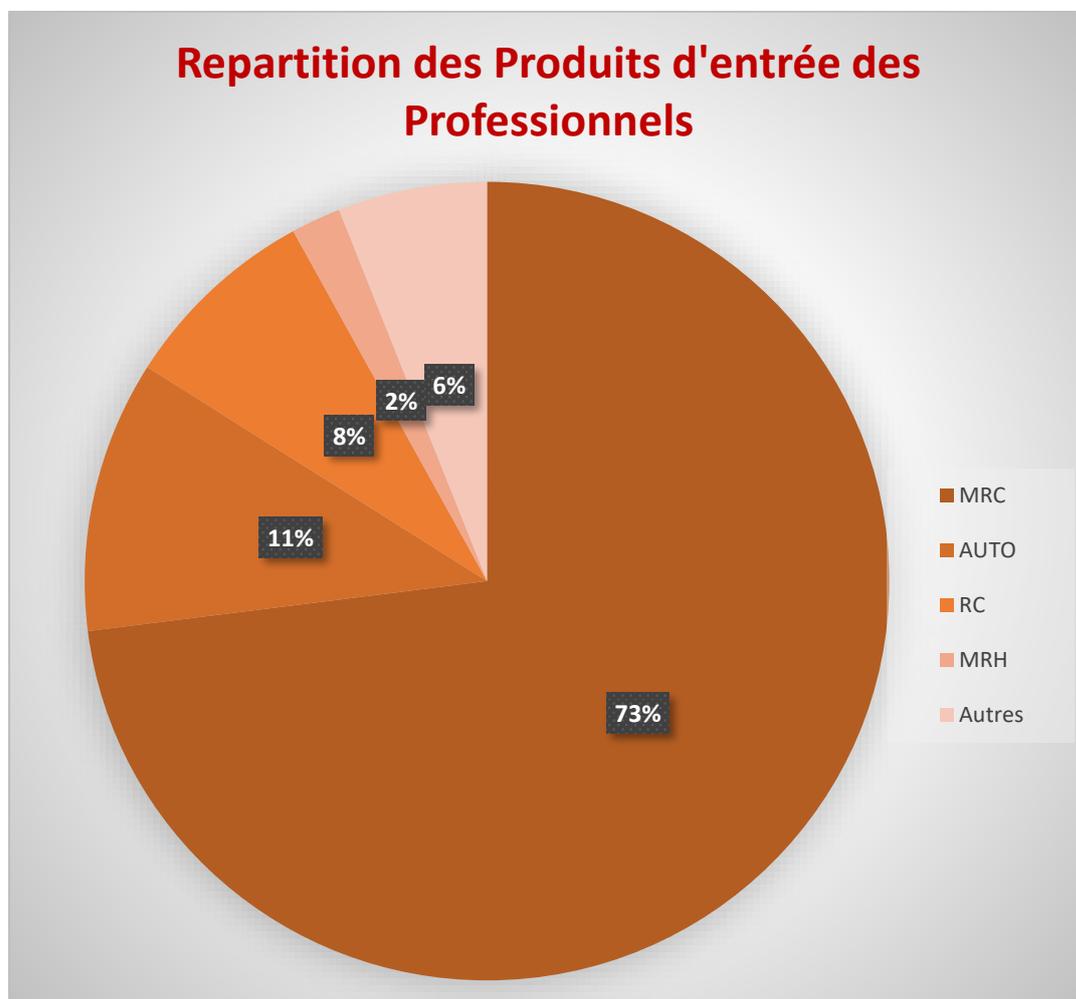


FIGURE 2 : REPARTITION DES PRODUITS D'ENTREE DES PROFESSIONNELS CHEZ GENERALI

1.4. Utilité de l'étude : Modélisation de l'élasticité à la majoration sur le produit MRC :

De façon générale, les primes d'assurance seront amenées à croître, en particulier sur le marché des professionnels, résultant de l'inflation et de sortie de la COVID 19. En effet, les exercices 2020 et 2021 ont été marqués par la fermeture de plusieurs commerces des professionnels avec une perte importante de leurs chiffres d'affaires. Si la décision de fermeture n'est pas très directement due à l'épidémie mais plutôt à une décision gouvernementale dans l'optique de mieux gérer l'épidémie, plusieurs assureurs considèrent que ces pertes d'exploitation ne sont pas liées à l'épidémie tel que spécifié dans les contrats pour être remboursées. Le gouvernement a incité les assureurs à une médiation avec les assurés pour le remboursement des pertes d'exploitation au cas par cas. Ainsi, si chez certains assureurs, les couvertures multirisques commerces sont rééditées en excluant certaines garanties (devenues des options) telles que les pertes d'exploitation pour risque épidémique, d'autres assureurs comme GENERALI augmentent fortement les primes de la multirisque professionnelle.

Pour GENERALI, tout d'abord un plan de sauvegarde des clients a été mis en place pendant la pandémie pour diminuer les primes des assurés. Ensuite s'en est suivies d'importantes pertes d'exploitations qui ont dégradées le S/P du produit MRC. Le retour de l'inflation augmente fortement les coûts des sinistres qui doivent être couverts par les primes acquises. Pour le renouvellement des produits MRC, le paramètre majeur du taux de revalorisation moyen est le niveau d'inflation. Ainsi, les exercices 2022 et 2023 seront accompagnés de fortes revalorisations moyennes minorées par le niveau d'inflation qui est déjà très important. D'où la nécessité de bien segmenter le portefeuille afin de cibler les profils les moins sensibles aux majorations à travers des modèles de résiliation et d'élasticité. En effet, les effets de la pandémie sont à relativiser. Même si la valeur client a été fortement impactée par la covid, la valeur potentielle des clients MRC reste importante puisque les clients MRC sont des clients avec un fort taux de multi équipement. L'étude de la sensibilité prix permettra également de sauvegarder cette valeur potentielle en protégeant les clients à valeur potentielle.

Partie 2. Données et études préliminaires

2.1. Objectif

L'objectif dans cette étude est d'étudier la résiliation des contrats MRC et de calculer l'élasticité à la majoration de la prime d'un contrat MRC. Pour ce faire plusieurs définitions sont possibles pour la résiliation :

- Le non-renouvellement : le plus commun pour les contrats très courts-termes de type IARD tel que le MRC
- La résiliation par suite d'un contentieux : éventuellement due à un défaut de paiement, une mauvaise foi de l'assuré.
- La résiliation à tout moment à partir d'une année d'ancienneté

La première définition est celle qui nous intéresse, en particulier les résiliations qui sont dues à une revalorisation de la prime. Les majorations des primes sont transmises au client 1mois à 45 jours avant échéance du contrat. Plusieurs scénarios sont possibles :

- Le contrat est résilié entre autres en raison de la majoration dans les 45 jours avant échéance
- Le contrat est renouvelé mais résilié dans les 6 mois suivants le renouvellement
- Le contrat est renouvelé mais résilié avant la prochaine échéance, juste avant que la prochaine prime ne soit transmise au client, et donc avant les futures majorations.

Les deux derniers cas de figures peuvent être regroupés, mais l'objectif en faisant cette distinction est de mettre en lumière l'effet non immédiat que la majoration des primes peut avoir sur le comportement du client.

Pour chaque exercice, nous définissons ainsi la variable indicatrice **top_resil_1_11mois** qui prend la valeur 1 si le client résilie son contrat un mois avant l'échéance et au plus tard onze mois après l'échéance. Cette variable telle que définie représente notre variable cible que nous cherchons à expliquer dans l'acte de résiliation. Nous considérons donc que l'impact la revalorisation se fait sur la fenêtre temporelle comprise entre deux revalorisations. Nous aurions pu faire le choix d'une autre fenêtre d'observation telle que 1 mois avant l'échéance et 4 mois au plus tard après l'échéance du contrat, mais nous considérons la fenêtre temporelle entre deux revalorisations comme étant la vision la plus complète.

2.2. Variables explicatives

Suivant les caractéristiques du contrat, de la nature du risque couvert et des caractéristiques propres aux clients, les comportements face à la revalorisation de la prime et à la résiliation peuvent varier. Nous distinguons 5 types de variables explicatives :

- Les données de l'intermédiaire
- Les données du contrat
- Les données du client
- Les données sur le risque
- Les données d'interactions

2.2.1. Données de l'intermédiaire

Les contrats MRC sont issus du marché de Professionnels entièrement souscrits par l'intermédiaire des deux principaux réseaux GENERALI qui sont les Agences GENERALI et les réseaux de courtages de GENERALI. La demande de devis et la souscription sont faites sur les sites de ces deux réseaux. Ainsi, les données de l'intermédiaire sont :

- Code réseau : Agent, Courtier
- Classement de l'intermédiaire : Standard, Etendu et Supérieur. Le classement de l'intermédiaire est fait en fonction de sa performance commerciale telle que le nombre d'affaires nouvelles apportées et en fonction de son expertise qui est la capacité de l'intermédiaire à souscrire et à suivre des risques plus ou moins complexes.
- Top code affaire : indicatrice de la présence d'un code affaire à la souscription. Le code affaire correspond au code attribué à certains intermédiaires performants leur permettant d'accorder des réductions à leurs clients afin de promouvoir le produit MRC pendant une période donnée, ou lors d'un lancement de produit.
- BCOM : Rabais commercial à la souscription en pourcentage.

2.2.2. Données du contrat

Les données contrats sont constatées avant échéance. On a donc

- Ancienneté contrat : exprimée en année, est l'ancienneté à l'échéance du contrat.
- Produit de couverture : **100% Pro Artisans Commerçants** pour les artisans commerçants, **100% Pro Service** pour les professionnels de services...
- La prime actuelle (prime avant majorations)
- Fractionnement de la prime : annuelle, mensuelle, trimestrielle ou semestrielle
- Nombre de sinistres dans les 12 mois précédant l'échéance
- Nombre total de sinistres
- Charge totale de sinistres
- Charge de sinistre sur 12 mois
- Date de dernier sinistre
- Les niveaux de majorations termes passées (une majoration terme correspond à la variation finale de la prime principale entre deux échéances du contrat). Par abus, nous appellerons les coefficients des écarts aux tarifs des majorations. Ainsi dans nos modèles les variables majorations ne seront autres que des coefficients d'écarts au tarif avant revalorisation. Par exemple pour une prime de **1000€** et un **taux de revalorisation** de **10%**, le coefficient d'écart au tarif est **(1+10%) = 1.1**.

- Le nombre de fortes revalorisations passées : Pour une profondeur d'historique limitée à 2 années, cette variable prend 3 valeurs : 0, 1 et 2 pour un seuil fixé à 6% de revalorisation. Ainsi pour les affaires nouvelles, cette variable prendra la valeur 0.

2.2.3. Données du client

La définition que nous donnons au client correspond au payeur de la prime. Les données clients sont donc :

- Ancienneté client : elle correspond à l'ancienneté contrat pour les mono détenteurs, cependant pour les multi détenteurs, elle sera l'ancienneté contrat du plus ancien contrat du client.
- Nombre de contrats détenus par le client
- Nombre de famille de produits détenus : IARD/ER/VIE
- Evolution du portefeuille client dans les 12 mois précédents l'échéance : Montée en équipement, portefeuille stable ou dés équipement.

2.2.4. Données sur le risque

- Le nombre de sites couverts : pouvant être plus de 1 pour les contrats qui couvrent plusieurs emplacements
- La surface du commerce en m²
- La qualité de l'occupant : Propriétaire ou Locataire pour simplifier car on peut noter que pour les multisites couverts par un seul contrat, le propriétaire peut posséder certains locaux et pas d'autres, dans ce cas il est considéré ici comme locataire.
- Famille d'activité du commerce.

2.2.5. Données d'interaction

Il s'agit des différentes données récoltées durant le parcours du client à travers les échanges entre l'assuré et l'assureur. On peut lister entre autres :

- Top interaction : indicatrice de retour ou de réponse du client à travers divers plateformes.
- Impayés sur contrat GENERALI : Elle prend 3 modalités : aucun, préavis et contentieux. Cette variable renseigne sur la qualité du client dans le paiement de la prime. Elle indique si un préavis a été transmis au client sur un de ces contrats et ou si le client a eu un contentieux avec Generali sur un de ces contrats.

2.3. Base de données

Nous disposons de 3 bases annuelles 2019, 2020 et 2021. Chaque base de données est constituée d'environ 130 000 contrats auxquels les majorations ont été transmises. Notre fenêtre temporelle d'observation des résiliations est entre 1 mois avant l'échéance et 11 mois au plus tard après l'échéance. La base de 2021 va contenir moins de contrats car l'observation des résiliations par suite de la majoration de 2021 est incomplète au moment de l'étude. Nous utilisons la base de 2019 comme base de données de référence sur laquelle nous allons entraîner nos modèles et la base de 2020 comme base de données d'entraînement.

Notre base de données ne contient pas les résiliations précoces ou de non-renouvellement qui émanent directement de la volonté de l'assureur de ne pas garder un contrat en portefeuille. Ainsi,

tous les contrats que nous étudions sont les contrats qui ont été revalorisés, et qui par suite de la revalorisation ont été renouvelés ou résiliés. Les résiliations étudiées sont celles de l'assuré après revalorisation et non celles de l'assureur.

Nos données d'étude ont de très bonnes complétudes globalement à l'exception de certaines informations sur le risque (telles que le chiffre d'affaires, la surface du site, le nombre d'employés) qui sont manquantes pour les contrats plus anciens que nous préciserons par la suite.

Dans la suite, nous présenterons quelques analyses descriptives de notre base de données.

2.4. Analyse univariée

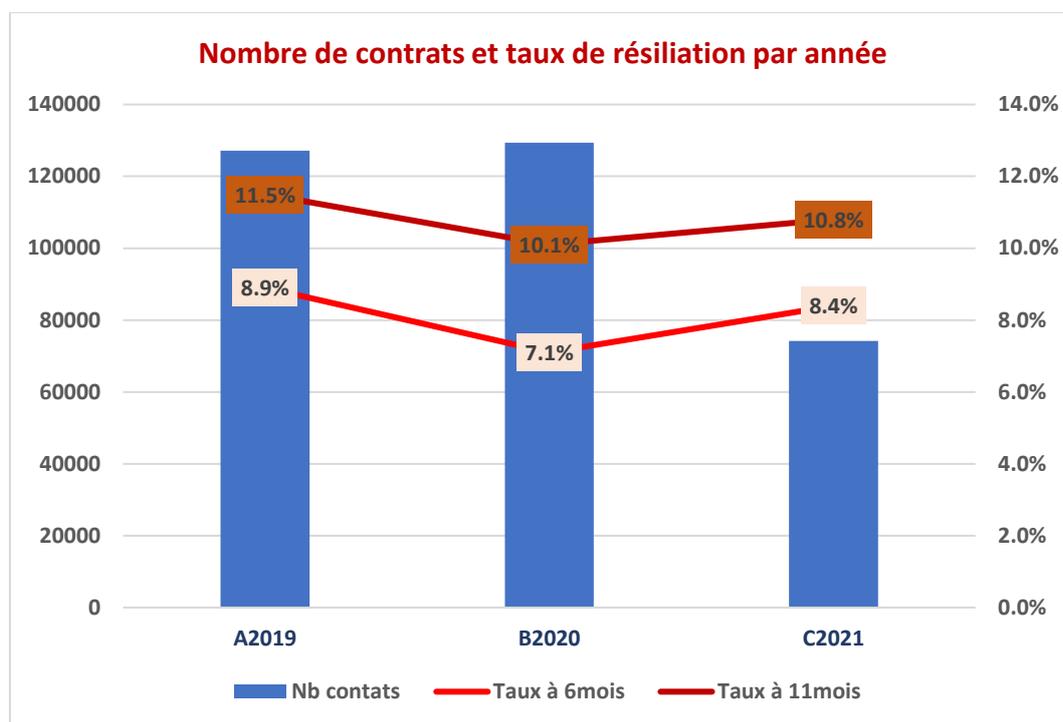


FIGURE 3 : NOMBRE DE CONTRATS ET TAUX DE RESILIATION PAR ANNEE

Sur nos trois données, on remarque que le taux de résiliation est historiquement croissant en fonction des années. Sur les 3 années que nous étudions, le taux baisse sensiblement en 2020 ce qui est dû à la particularité de cette année notamment le ralentissement de l'activité pour cause du Covid. Le taux se redresse en 2021 comme attendu.

☐ Le réseau Courtier résilie plus que le réseau Agent.

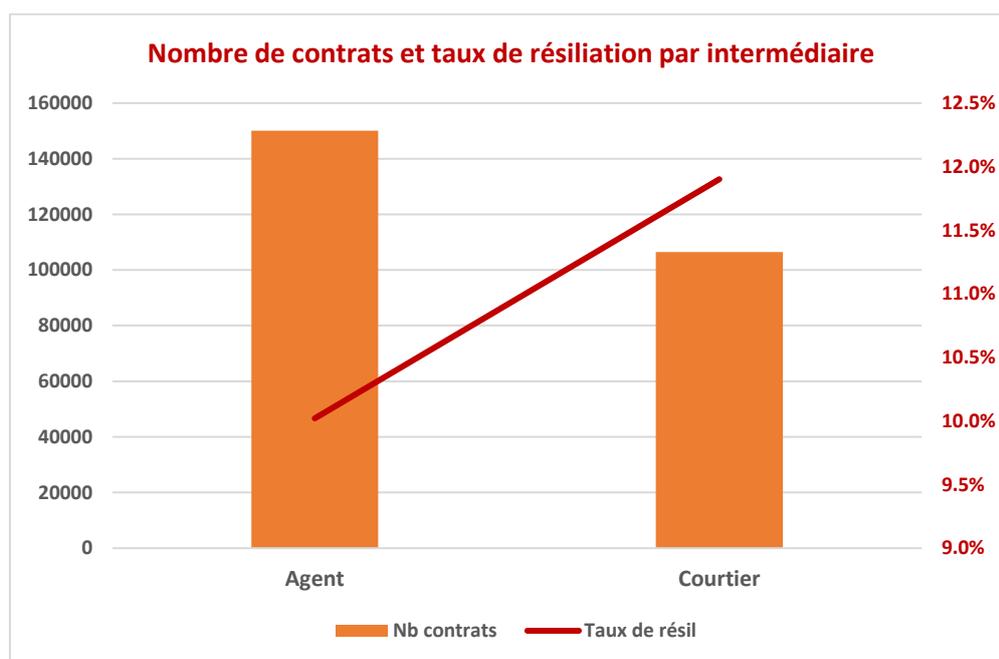


FIGURE 4 : NOMBRE DE CONTRATS TAUX DE RESILIATION PAR RESEAU

Sur les deux années 2019 et 2020, nous remarquons que le réseau courtier résilie 2 points de plus que le réseau Agent. Le classement des intermédiaires montre également que les réseaux classés Etendus et Supérieurs résilient moins que les réseaux classés Standards.

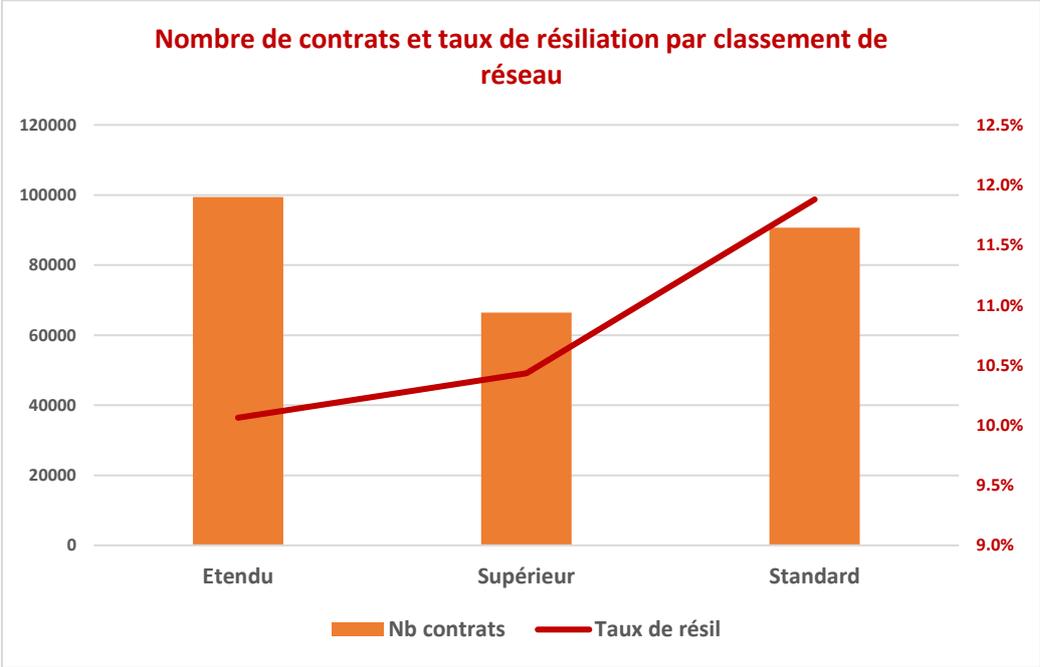


FIGURE 5 : NOMBRE DE CONTRATS ET TAUX DE RESILIATION PAR CLASSEMENT DU RESEAU

- ❑ Les produits nouveaux Pro Artisans Commerçants et Pro Service résilient jusqu’à 3 points de plus que les Produits Autres.

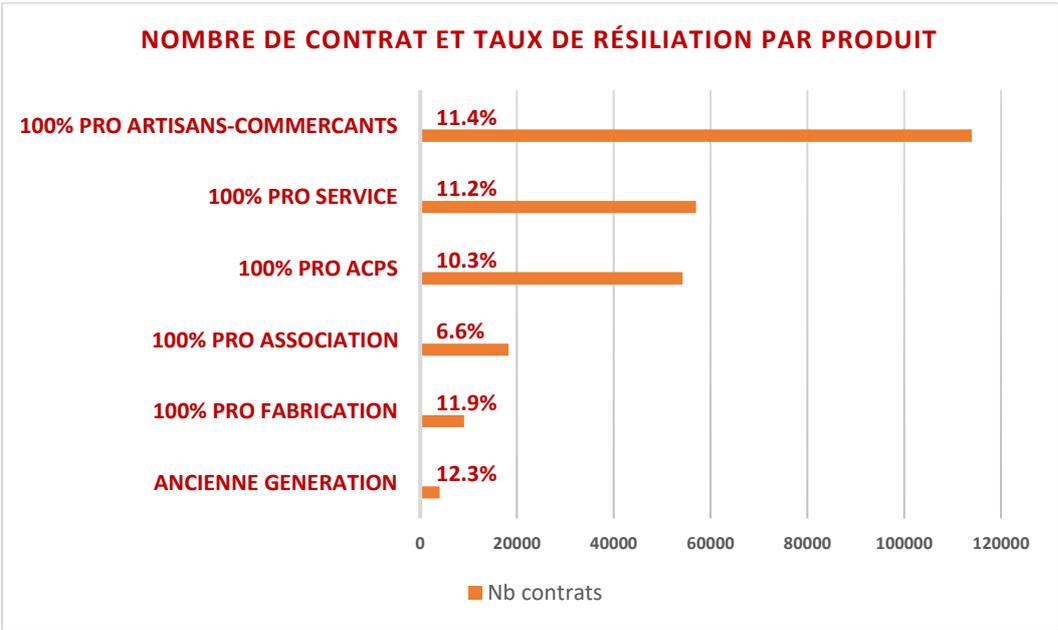


FIGURE 6 : NOMBRE DE CONTRATS ET TAUX DE RESILIATION PAR TYPE DE PRODUIT

Les produits Associations sont ceux qui résilient le moins de tous les produits MRC. Aussi, les produits de l'ancienne génération semblent résilier plus que les autres produits.

- ❑ Les clients multi équipés résilient moins que les mono-détenteurs.

TABLEAU 2 : EVOLUTION DU MULTI EQUIPEMENT DES CLIENTS MRC

Année	Nb contrats	Taux de résiliation	Multi équipement
2019	127124	11.5%	30.2%
2020	129356	10.1%	33.0%
2021	134000	10.8%	36.6%

Le tableau précédent montre l'évolution du multi équipement des clients MRC sur les années de 2019 à 2021. On remarque que les clients MRC sont de plus en plus multi équipés. Les multi détenteurs sont des clients de longue durée. Les clients multi équipés représentent 43 % du portefeuille MRC en 2022. Ils résilient beaucoup moins que les clients monos équipés. En particulier ceux qui détiennent plusieurs gammes (clients *multigammés*) de produits résilient encore moins. Les détenteurs mixtes de type MRC – Epargne résilient moins que les détenteurs MRC – MRH.

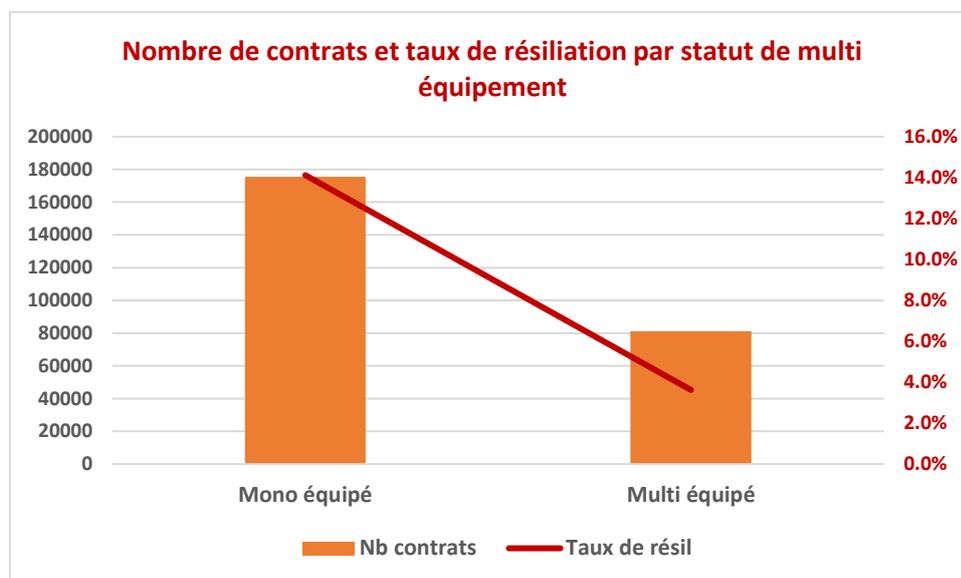


FIGURE 7 : NOMBRE DE CONTRATS ET TAUX DE RESILIATION PAR NOMBRE DE CONTRATS DU CLIENT

- ❑ Les clients sinistrés ne résilient presque pas.

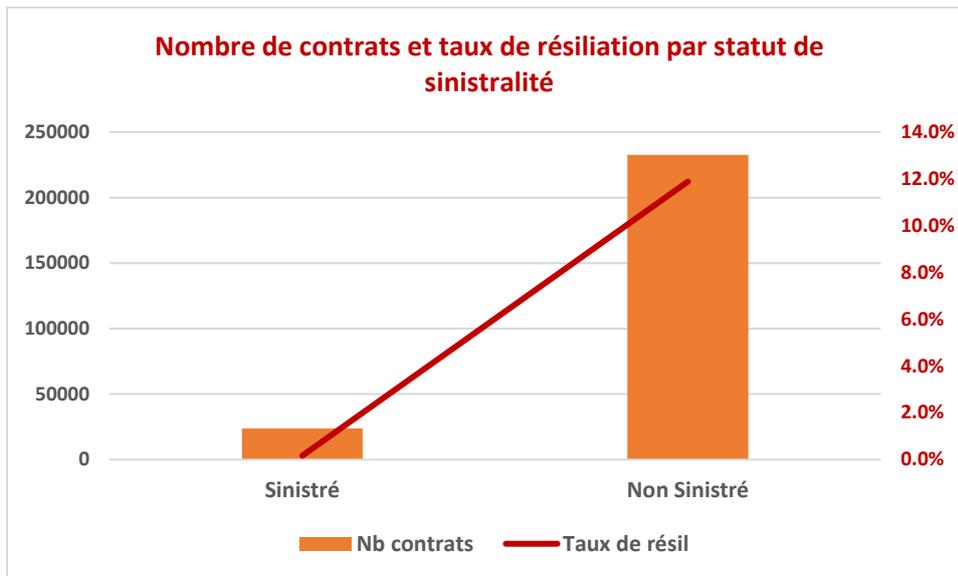


FIGURE 8 : NOMBRE DE CONTRATS ET TAUX DE RESILIATION PAR STATUT DE SINISTRALITE DU CONTRAT

Le taux de contrats sinistrés représente 9% de 2019 à 2020 et le taux de résiliation des contrats sinistrés est de 0.3%.

- Les commerces provinciaux résilient moins que les commerces parisiens.

20% des contrats de 2019 et 2020 sont sur des commerces parisiens et le taux de résiliations des commerces parisiens est 12% contre 10.5% pour les commerces provinciaux.

- Le Chiffre d’Affaires, le Nombre d’Employés et la prime ont tous un effet négatif sur la résiliation.

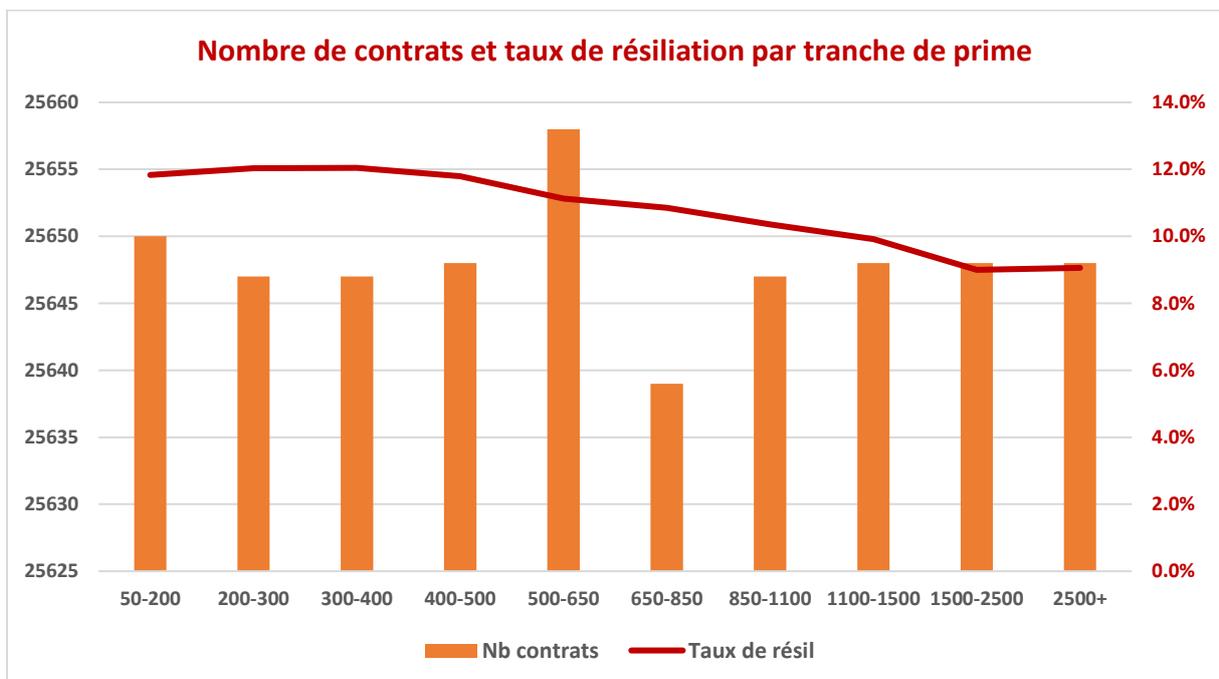


FIGURE 9 : NOMBRE DE CONTRATS ET TAUX DE RESILIATION PAR TRANCHE DE LA PRIME

Plus la prime est importante ce qui implique un niveau de risque plus élevé et moins le client résilie.

☐ Cependant la résiliation croit avec la prime au m².

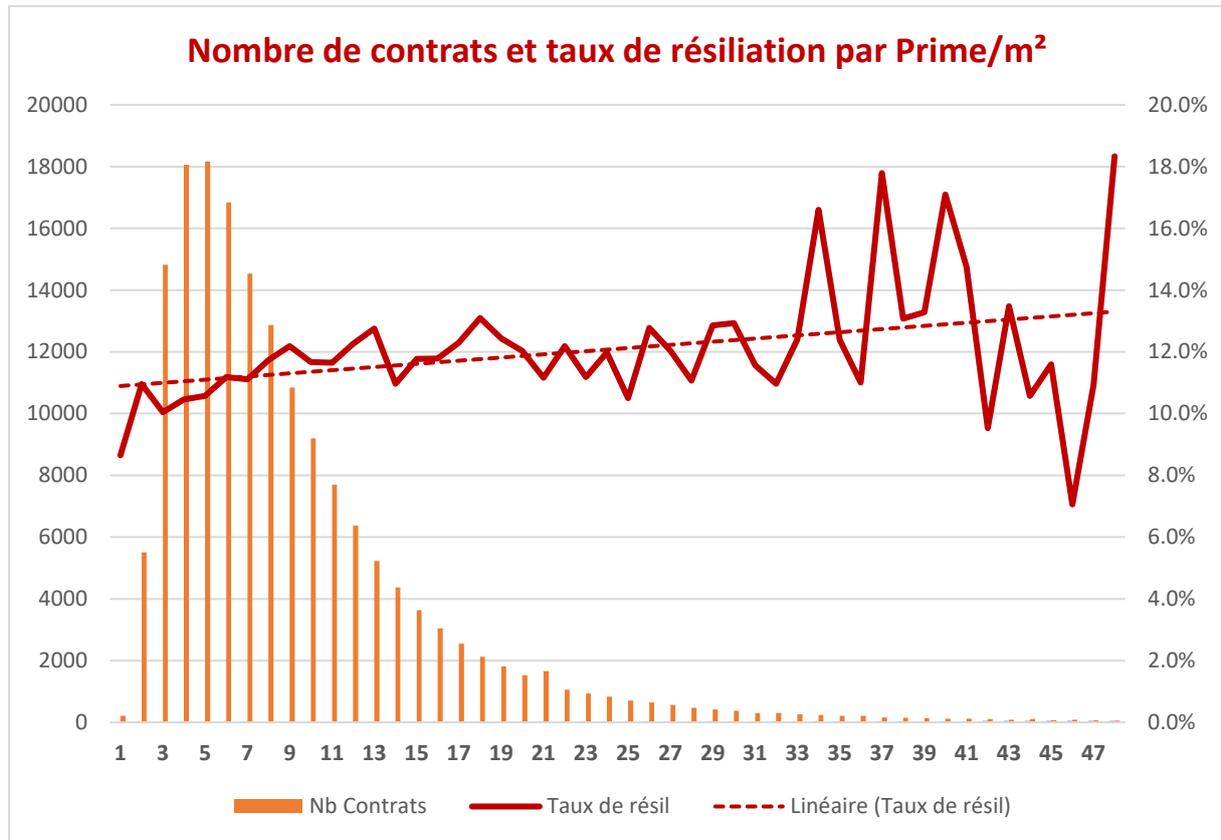


FIGURE 10 : NOMBRE DE CONTRATS ET TAUX DE RESILIATION EN FONCTION DE LA PRIME

L'ajustement linéaire du taux de résiliation en fonction de la prime au m² donne le résultat suivant :

```
Call:
lm(formula = TR ~ Prime_Sur, data = Prime_Surface)

Residuals:
    Min       1Q   Median       3Q      Max
-0.057266 -0.007548  0.001025  0.007193  0.053416

Coefficients:
(Intercept)  Estimate Std. Error t value Pr(>|t|)
Prime_Sur    0.0003729  0.0001963    1.90  0.0639 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01825 on 45 degrees of freedom
Multiple R-squared:  0.07426, Adjusted R-squared:  0.05369
F-statistic:  3.61 on 1 and 45 DF, p-value: 0.06386
```

L'ajustement n'est globalement pas significatif au seuil de 5% en revanche au seuil de 10% oui et dans ce cas la prime au mètre carré n'expliquerait que 7% de la variation du taux de résiliation.

☐ Le taux de résiliation est croissant en fonction de la majoration (écart au tarif)

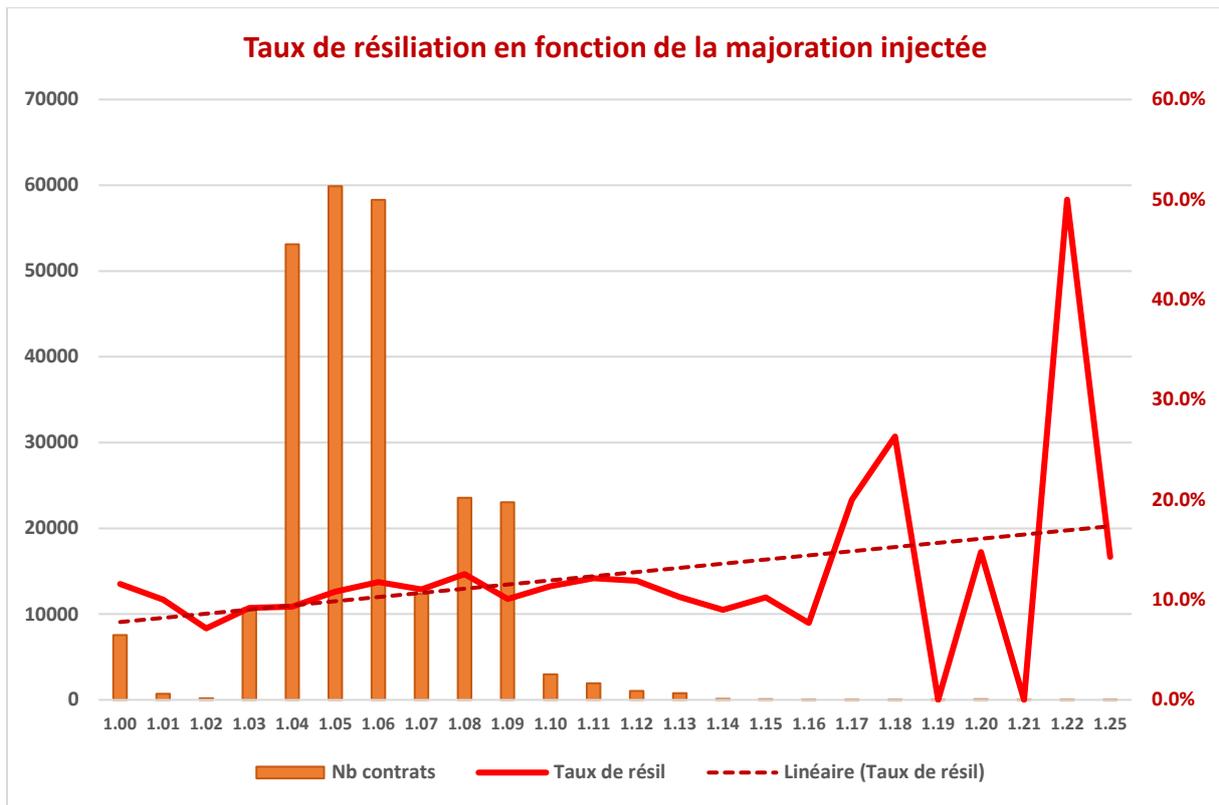


FIGURE 11 : NOMBRE DE CONTRATS ET TAUX DE RESILIATION EN FONCTION LA MAJORATION

Plus le niveau de majoration est élevé moins le renouvellement se fait. Le lien semble linéaire entre le taux de résiliation et la majoration. L'ajustement linéaire permet d'avoir les résultats suivants.

```
Call:
lm(formula = TR ~ majo, data = S)

Residuals:
    Min       1Q   Median       3Q      Max
-0.16352 -0.03054 -0.00384  0.01441  0.33246

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.3229    0.2996  -1.078   0.293
majo           0.4020    0.2680   1.500   0.148

Residual standard error: 0.09281 on 22 degrees of freedom
Multiple R-squared:  0.09282,    Adjusted R-squared:  0.05158
F-statistic: 2.251 on 1 and 22 DF,  p-value: 0.1478
```

Le modèle n'est globalement pas significatif ce qui permet de conclure la non-linéarité de la majoration en fonction du taux de résiliation.

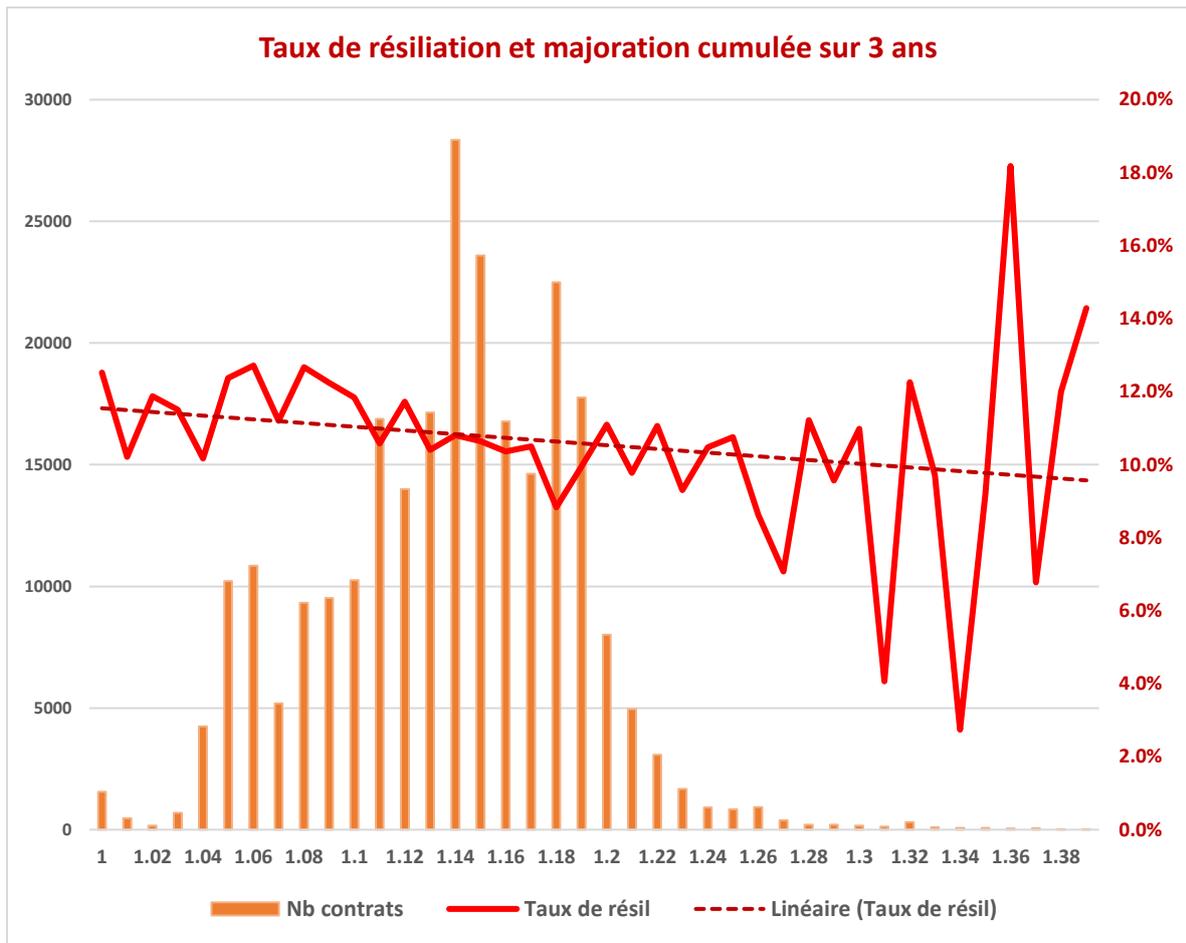


FIGURE 12 : NOMBRE DE CONTRATS ET TAUX DE RESILIATION EN FONCTION DE LA MAJORATION CUMULEE SUR 3 ANS

La variable majoration n'est pas une variable aléatoire vu qu'elle émane d'un tableau de majoration en suivant certains critères de revalorisation. On remarque également une tendance à la décroissance du taux de résiliation en fonction de la majoration cumulée sur trois ans. Pour les contrats de 2020, cette majoration va correspondre au cumule de 2018 à 2020. Ce qui est contre intuitif (cette tendance à la baisse) puisqu'à priori, un contrat ayant subi de fortes revalorisations successives va plus avoir tendance à résilier. L'ajustement linéaire du taux de résiliation avec la majoration cumulée sur trois ans n'est pas statistiquement significatif comme le montre le résumé suivant.

```
Call:
lm(formula = TR ~ majo_cumulee, data = SS)

Residuals:
    Min       1Q   Median       3Q      Max
-0.14632 -0.02844 -0.00577  0.01104  0.86627

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.03869   0.14113   0.274   0.785
majo_cumulee 0.06294   0.10967   0.574   0.568

Residual standard error: 0.1382 on 53 degrees of freedom
Multiple R-squared:  0.006176, Adjusted R-squared:  -0.01258
F-statistic: 0.3294 on 1 and 53 DF, p-value: 0.5685
```

2.5. Analyse multivariée

Dans cette partie, nous étudions principalement le lien entre les variables.

La liste des variables est indiquée dans le tableau suivant :

TABEAU 3 : CORRESPONDANCE DES VARIABLES

Variables	Label	Type
top_resil_1_11_mois	Indicatrice de résiliation du contrat sur la fenêtre - 1/+11mois	Binaire
majo	Ecart au tarif / majoration (ou majoration N)	Continue
bcom	Rabais commercial à la souscription	Continue
prmpft	Prime avant revalorisation	Continue
ancclient2	Ancienneté client	Continue
anccontrat	Ancienneté contrat	Continue
majo_1	Majoration N-1	Continue
majo_2	Majoration N-2	Continue
top_sinistre12mois	Indicatrice de sinistre sur 12 mois	Binaire
top_multiequipe	Indicatrice de multi équipement du client	Binaire
top_multigamme	Indicatrice de multi gamme du PF client (IARD, VIE, Epargne-Retraite)	Binaire
top_multisites	Indicatrice de multisites couverts	Binaire
topcodeaffaire	Indicatrice de présence de code affaire à la souscription	Binaire
topNonProprio	Indicatrice de propriété des locaux du commerce	Binaire
topres12mois	Indicatrice de résiliation d'un autre contrat sur 12 mois	Binaire
topLOLR12mois	Indicatrice de retard de paiement sur un contrat du client	Binaire
toSGRSREC12mois	Indicatrice de contentieux sur un des dossiers client	Binaire
reseau	Réseau d'appartenance du contrat	Catégorielle
Classement	Classement du réseau	Catégorielle
top_mensuel	Indicatrice de paiement mensuel de la prime	Binaire
type_produit	Type de produit de couverture	Catégorielle
famille_activite	Famille d'activité du commerce	Catégorielle
gamme-entree	Gamme du dernier contrat souscrit	Catégorielle
gamme-sortie	Gamme du dernier contrat résilié	Catégorielle
laps_dernier_sous	Nombre de mois depuis dernière souscription	Continue
laps_dernier_resil	Nombre de mois depuis dernière résiliation	Continue
mnt_ca_france	Chiffre d'affaires	Continue
nb_effectif	Nombre d'employés	Discrète
surface_sites	Surface du site	Continue
top_interaction	Indicatrice d'interaction récente avec le client	Binaire
nsin_courant	Nombre de sinistres courant	Discrète

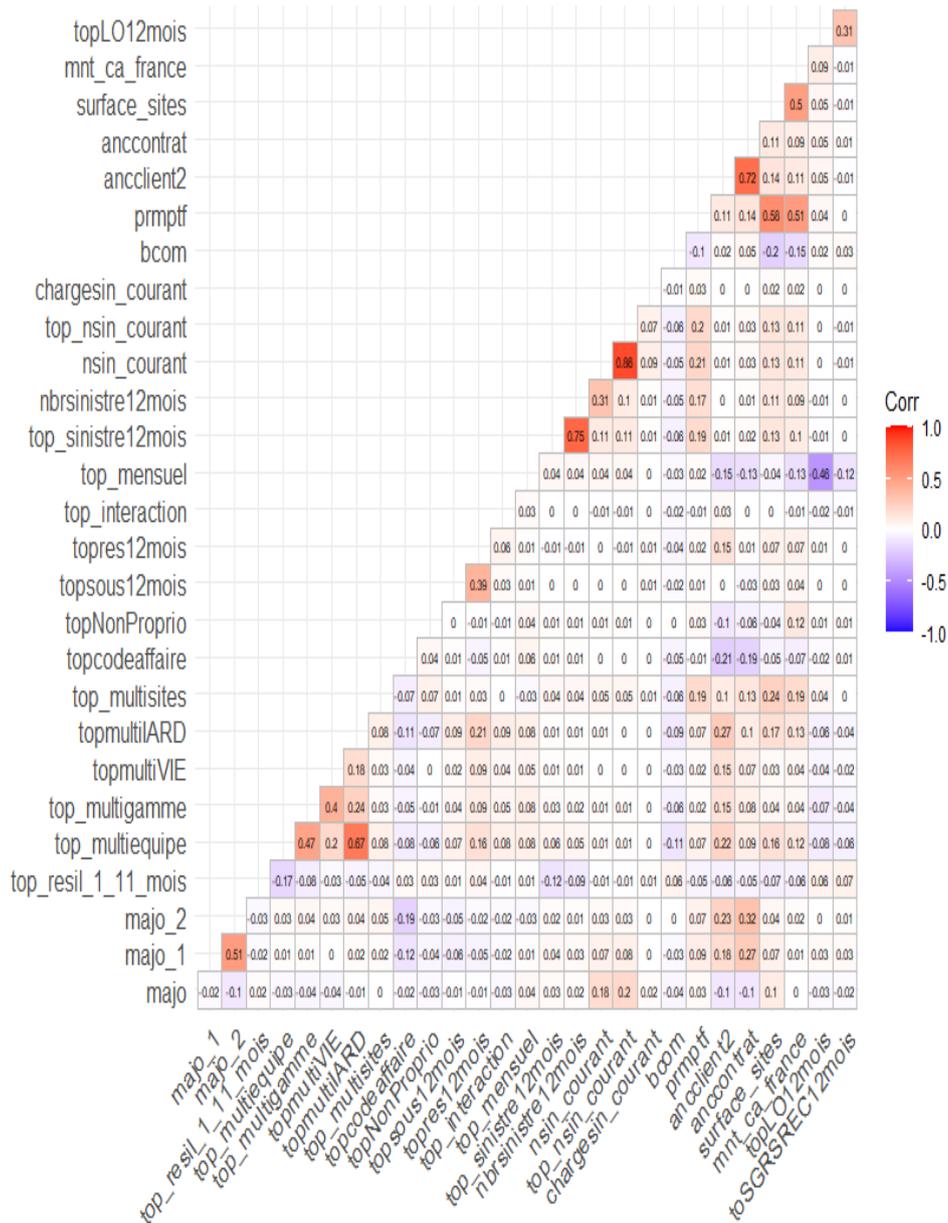


FIGURE 13 : CORRELATION DES VARIABLES QUANTITATIVES

Les variables qui sont plus corrélées à la variable cible (top_resil_1_11_mois) sont :

- Le top_multiéquipe : indicatrice de multi équipement du client et la corrélation est négative bien que faible.
- Le top_sinistre12 mois : indicatrice de sinistralité du contrat sur les 12 mois précédant l'échéance. La corrélation est aussi négative ici.

Aussi, on peut remarquer que la dernière majoration dépend beaucoup moins de la précédente(majo_1) que de celle d'il y a 2 ans(majo_2) dans nos données. Les fortes corrélations d'avec la majoration sont celles d'avec la sinistralité, l'ancienneté client et contrat et de la majoration N-2(majo_2).

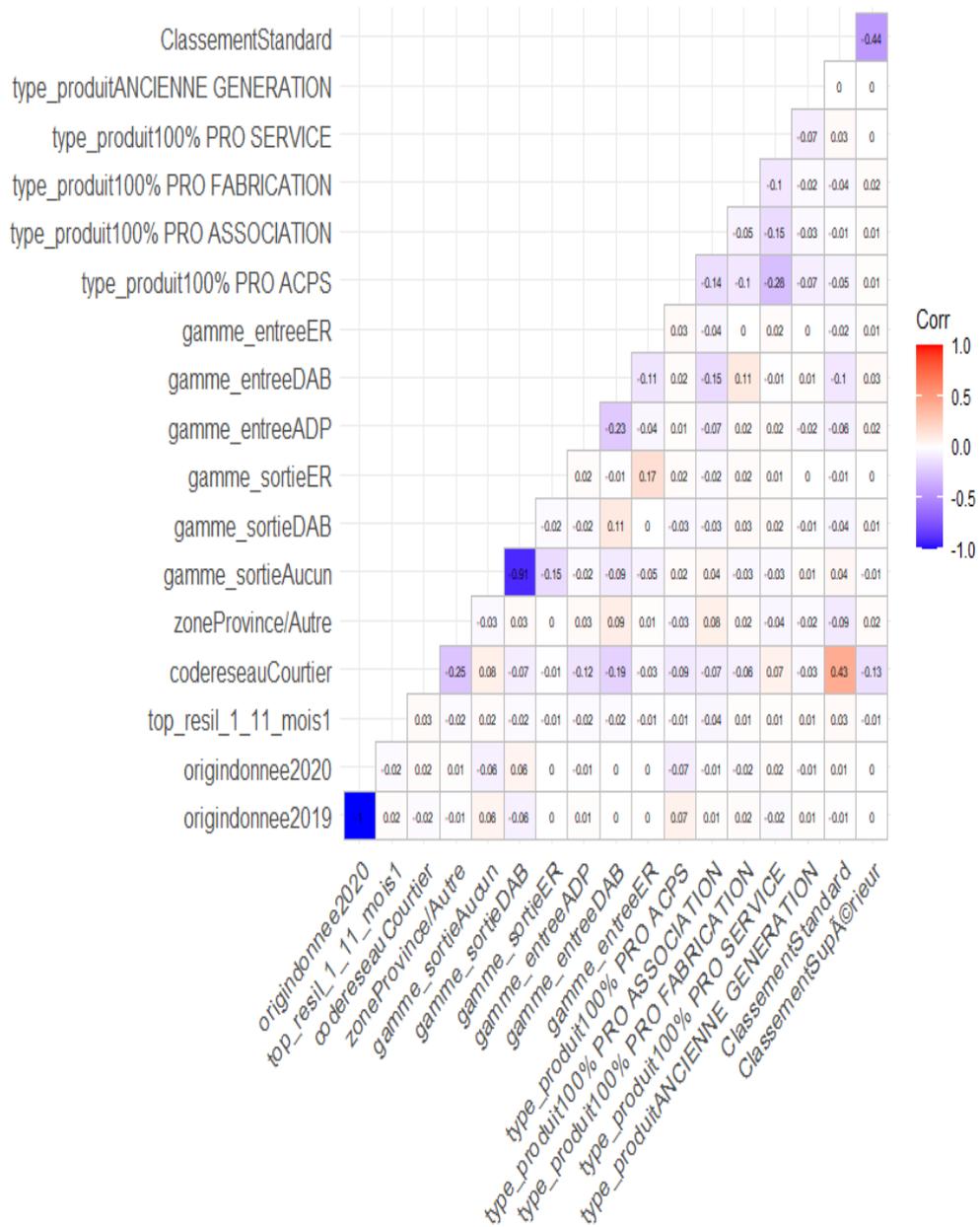


FIGURE 14 : CORRELATION DES VARIABLES CATEGORIELLES

Partie 3. Modélisation

L'objectif de la modélisation est de permettre finalement de comprendre l'effet des variables explicatives dans l'acte de résiliation par suite d'une revalorisation. Mais nous souhaitons aussi mesurer la sensibilité de chaque profil client à la prime ou à la revalorisation de sa prime. Pour ce faire, nous devons estimer la probabilité de rétention du client et l'effet de la majoration sur chaque caractéristique du contrat. Les probabilités de rétentions peuvent être estimées par des modèles de classifications tels que les modèles linéaires généralisés, les modèles additifs généralisés, Les arbres de décisions mais toutes ces méthodes ne sont pas paramétriques et donc ne permettent pas de mesurer l'élasticité implicitement. Ainsi nous divisons notre modélisation en deux parties :

- Modélisation de la résiliation pour estimer les probabilités de rétention : nous utilisons les algorithmes paramétriques et non paramétriques.
- Estimation des coefficients de l'élasticité : par les algorithmes paramétriques uniquement.

Les modèles que nous construisons sont entraînés sur les données de 2019 qui est l'année de référence que nous choisissons puis les prédictions sont faites sur les données de 2020 et 2021.

3.1. Modélisation de la résiliation.

3.1.1. Modèles Linéaires généralisés

Dans cette partie, nous utilisons les modèles linéaires généralisés pour modéliser l'acte de résiliation par suite de la revalorisation de la prime.

a. Rappels Théoriques GLM

Les modèles linéaires généralisés en abrégé GLM sont une méthode d'estimation paramétrique très utilisée en machine Learning pour prédire des sorties binaires, discrètes et continues. Les GLM sont une généralisation des modèles linéaires classiques. Ils permettent aussi de modéliser des comportements non linéaires, non normaux. L'intérêt premier de ces méthodes est celui de permettre de capter les variables dites significatives qui sont les plus intéressantes dans l'explication de la variable cible. Elles permettent ainsi de comprendre la variable cible.

Soit Y la variable cible aléatoire qu'on cherche à expliquer par le groupe de variables (X_1, \dots, X_k) dites variables déterministes explicatives de Y .

La relation entre la variable cible et le groupe de variables explicatives est établie au travers d'une fonction de lien suivant cette équation :

$$g[E(Y_i)] = \beta_0 + \sum_{j=1}^k \beta_j X_{ij}$$

où g est la fonction de lien et le vecteur $(\beta_0, \beta_1, \dots, \beta_k)$ ce qu'on cherche à estimer.

La loi de probabilité de Y appartient à la famille exponentielle de densité de forme générale :

- $f(y, \theta, \phi) = \exp [(y\theta - b(\theta))/a(\phi) + c(y, \phi)]$
 – $\forall y \in \mathbb{N}$ ou \mathbb{R}
 – $\theta \in \mathbb{R}$ paramètre canonique (ou paramètre de la moyenne)
 – $\phi \in \mathbb{R}$, paramètre de dispersion
 – $a \in \mathbb{R}^*$
 – b : fonction définie sur \mathbb{R} et deux fois dérivable
 – c est une fonction définie sur \mathbb{R}^2

L'estimation des $(\beta_0, \beta_1, \dots, \beta_k)$ se fait par maximum de vraisemblance.

Si on suppose que $\mu_i = E(Y_i)$, $\mu \equiv E(Y)$ et $\eta_i = g(\mu_i) = X_i\beta$, on définit la fonction de vraisemblance comme suit :

$$\mathcal{L}(\beta) = \prod_{i=1}^n f_{\theta_i}(y_i)$$

Le log de vraisemblance correspondant est :

$$\begin{aligned}
 l(\beta) &= \sum_{i=1}^n \log[f_{\theta_i}(y_i)] \\
 &= \sum_{i=1}^n \left[\frac{y_i\theta_i - b_i(\theta_i)}{a_i(\phi)} + c_i(y_i, \phi) \right]
 \end{aligned}$$

avec n le nombre d'observations.

Dans le cas où $a_i(\phi) = \phi/w_i$

$$\frac{\partial l}{\partial \beta_j} = \frac{1}{\phi} * \sum_{i=1}^n w_i * \left[y_i \frac{\partial \theta_i}{\partial \beta_j} - b_i'(\theta_i) \frac{\partial \theta_i}{\partial \beta_j} \right]$$

$$\frac{\partial \theta_i}{\partial \beta_j} = \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j}$$

$$\frac{\partial \mu_i}{\partial \beta_j} = b_i''(\theta_i)$$

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b_i''(\theta_i)}$$

$$\frac{\partial l}{\partial \beta_j} = \frac{1}{\phi} * \sum_{i=1}^n \frac{[y_i - b_i'(\theta_i)]}{b_i''(\theta_i)/w_i} \frac{\partial \mu_i}{\partial \beta_j}$$

$$\frac{\partial l}{\partial \beta_j} = (g^{-1})'(g(\mu_i)) \times \frac{y_i - \mu_i}{\text{Var}(Y_i)} X_{ij}$$

Les paramètres $(\beta_0, \beta_1, \dots, \beta_k)$ qui maximisent le log de vraisemblance respectent la condition de premier ordre :

$$\sum_{i=1}^n \frac{[y_i - \mu_i]}{\text{Var}(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0 \text{ pour tout } j.$$

Sous R, les paramètres $(\beta_0, \beta_1, \dots, \beta_k)$ sont déterminés de façon itérative par l'algorithme itératif de Newton Raphson.

b. Rappels Théoriques sur la régression logistique

Nous cherchons à modéliser une indicatrice de résiliation de contrat sachant les caractéristiques du contrat. $Y = \mathbf{1}_{(\text{Résiliation})}$ qui est une variable binaire et la loi adaptée est de type Bernoulli qui appartient à la famille exponentielle. Afin d'obtenir une sortie binaire, le fonction *logit* va être utilisée comme lien.

L'équation

$$g[E(Y_i)] = \beta_0 + \sum_{j=1}^k \beta_j X_{ij}$$

se réécrit en prenant $g(x) = \ln[x/(1-x)]$ dans ce cas:

$$\ln[p/(1-p)] = \beta_0 + \sum_{j=1}^k \beta_j X_j$$

où p est le paramètre de la loi de Bernoulli. La probabilité de résilier s'écrit comme :

$$\text{Proba de résilier} = Tx(p) = \exp(\beta_0 + \sum_{j=1}^k \beta_j X_j) / [1 + \exp(\beta_0 + \sum_{j=1}^k \beta_j X_j)]$$

Le log de vraisemblance dans le cas de la régression logistique se réécrit comme :

$$\ln(\mathcal{L}(p_1, \dots, p_n, \phi, y_1, \dots, y_n)) = \sum_{i=1}^n y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)$$

c. Validation des modèles

Pour valider le modèle, plusieurs étapes et critères sont à suivre et à prendre en compte.

- La Déviance :

Permet de comparer le modèle estimé au modèle saturé comportant autant de paramètres à estimer que d'observations (toute la donnée). Mathématiquement, c'est une différence de log de vraisemblance qui est calculée et on s'attend à ce qu'elle soit aussi petite que possible si le modèle est de bonne qualité. C'est aussi un critère de sélection de variables. En effet, plus on ajoute de variables à notre modèle et plus on complexifie notre modèle donnant un gain en précision mais pas en robustesse. Donc il est préférable d'avoir un modèle moins complexe mais qui prédit bien la variable de sortie.

$$\text{Deviance} = -2(l - l_{\text{saturé}})$$

avec l le log vraisemblance du modèle estimé et $l_{\text{saturé}}$ le log vraisemblance du modèle saturé.

- **Critère d'information d'Akaike (AIC) et Bayésien (BIC) :**

Mesurent la qualité d'un modèle statistique et servent à comparer plusieurs modèles également. L'AIC utilise le maximum de vraisemblance en pénalisant les modèles avec plus de variables explicatives. On note par k le nombre de paramètres du modèle. Le critère AIC est défini de la manière suivante :

$$AIC = 2k - 2l.$$

Le critère bayésien à la différence de l'AIC prend en compte en plus du nombre de paramètres la taille du modèle :

$$BIC = -2l + \log(n) * k$$

avec k le nombre de paramètres, l le log vraisemblance du modèle et n le nombre d'individus composant l'échantillon.

- **Test de significativité des variables**

Ce test mesure la significativité statistique du modèle (des variables). Nous utilisons comme test de significativité le Test de Wald :

Le test confronte deux hypothèses :

La statistique de $\hat{\beta}_j$ est nul contre $\hat{\beta}_j$ est non nul. ce test

$$z = \frac{\hat{\beta}_j}{\sqrt{\text{Var}(\hat{\beta}_j)}} \sim N(0,1)$$

Les GLM donnent en output les valeurs suivantes :

Estimate : Les coefficients estimés

Std. Error : Erreur d'estimations des coefficients

$z \text{ value} = \frac{\text{Estimate}}{\text{Std. Error}}$: Statistiques de Wald des coefficients

$P(> |z|)$: P-value

- **Sélection des variables : Stepwise selection**

Cette méthode consiste à ajouter et retirer de manière progressive les variables dans le modèle en fonction de leur pouvoir explicatif et leur contribution dans la performance du modèle pour retenir finalement le groupe de variables qui donne le modèle avec la meilleure performance.

Il existe trois façons de procéder :

Forward selection : consiste à démarrer par un modèle avec uniquement la constante à estimer. Ensuite de façon progressive ajouter les variables et s'arrêter lorsque la métrique choisie ne s'améliore plus de façon significative.

Backward selection : consiste à démarrer avec toutes les variables et retirer progressivement les variables qui n'améliorent pas significativement les performances du modèle.

Stepwise selection : n'est autre que la combinaison du Backward et Forward selection. Consiste à démarrer par un modèle sans variables prédictives et d'ajouter ensuite de façon séquentielle les variables les plus contributives à l'explicabilité du modèle. Lors de l'ajout de chaque variable, retirer toutes les autres variables qui n'améliorent pas les performances prédictives du modèle.

- **Matrice de confusion :**

S'agissant d'une régression logistique, nous définissons à partir de la matrice de confusion du tableau ci-dessous les critères suivants :

Matrice de confusion	Actual 1	Actual 0
Predicted 1	True Positive (TP) $TPR = \frac{TP}{TP + FN}$	False Positive (FP) $FPR = \frac{FP}{TP + FN}$
Predicted 0	False Negative (FN) $FNR = \frac{FN}{TP + FN}$	True Negative (TN) $TNR = \frac{TN}{TP + FN}$

Pour comparer les modèles, nous utilisons :

ROC AUC (AUC) : AUC signifie aire sous la courbe ROC, courbe qui représente le compromis entre le TPR et le FPR pour chaque seuil de décision. Bien sûr plus le TPR est élevé et plus le FPR est faible pour chaque seuil. Les classificateurs dont les AUC sont proches de 1 (100%) sont les meilleurs.

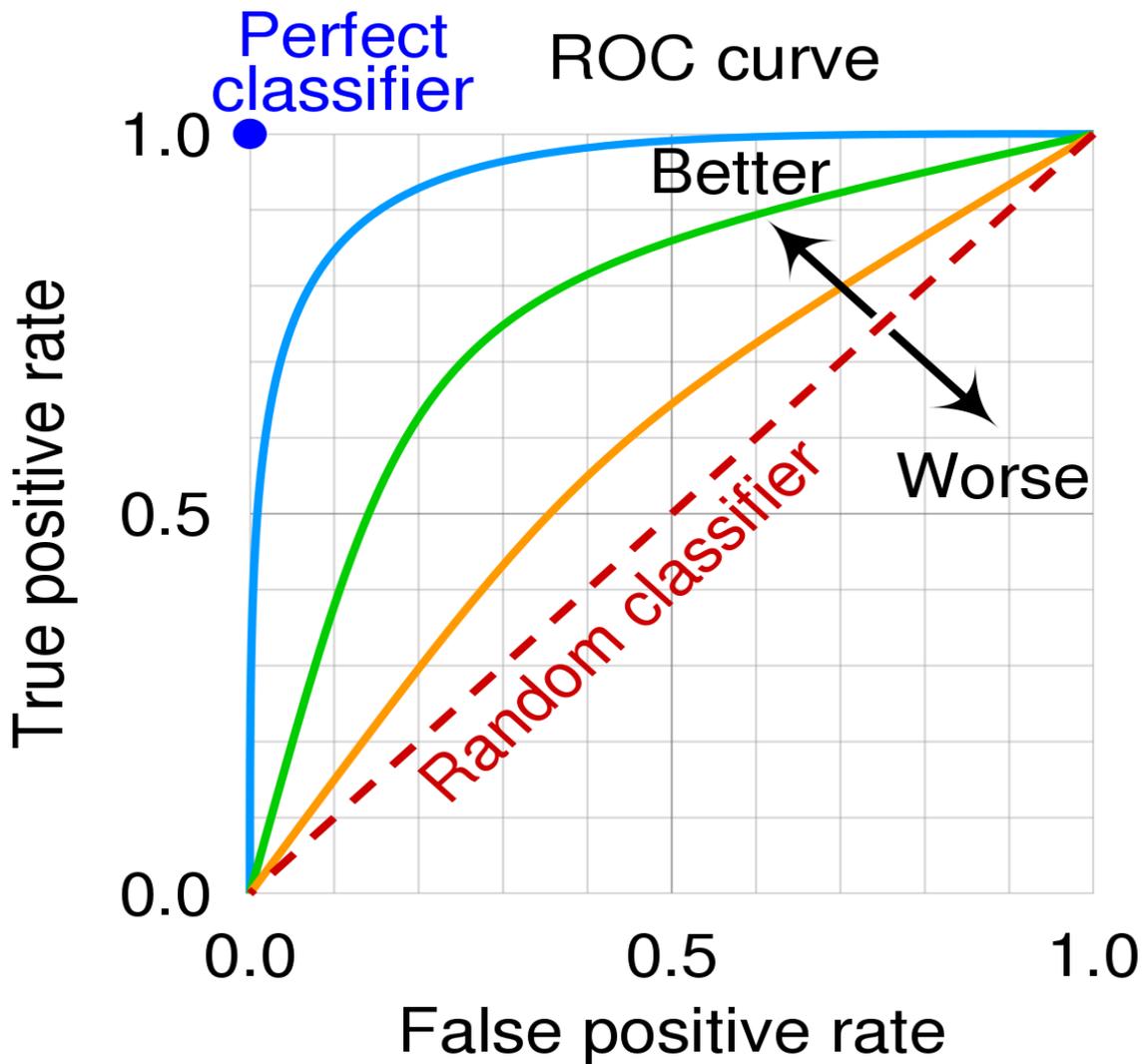


FIGURE 15 : EXEMPLE DE COURBE ROC

Imprécision = 1- Précision, où la précision est le pourcentage de prédictions positives correctes effectuées. Elle dépend du seuil choisi qui par défaut pour les données équilibrées est de 0.5.

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1\ Score = 2 * \frac{Précision * Recall}{Précision + Recall}$$

- Les odds Ratios

Un des avantages des modèles linéaires généralisés est qu'ils permettent d'estimer les odds ratios qui représentent les rapports de chances pour les modalités des variables afin d'identifier les caractéristiques qui influencent le plus le phénomène étudié. Dans le cas de la résiliation les odds ratios permettent de déterminer pour toutes les variables les modalités présentant un risque de résiliation plus élevé.

Pour une variable indicatrice du fait d'avoir eu au moins un sinistre sur une année ($X = 1$ si au moins 1 sinistre a été enregistré), l'odds ratio est le rapport entre la proba de résiliation et la proba de rétention lorsqu'on passe de l'état $X = 0$ à l'état $X = 1$.

d. Application aux données

Les résultats d'un premier modèle sont présentés dans le résumé suivant.

```
Call:
glm(formula = top_resil_1_11_mois ~ majo + bcom + prmptf + anconcontrat +
     majo_1 + majo_2 + top_sinistre12mois + top_multiequipe +
     top_multigamme + topmultiIARD + top_multisites + topcodeaffaire +
     topNonProprio + topres12mois + topLOLR12mois + toSGRSREC12mois +
     codereseau + Classement + top_mensuel + type_produit + famille_activite_red,
     family = binomial, data = train_1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4607  -0.5744  -0.4131  -0.2373   3.8620

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.957e+00  7.914e-01  -6.264 3.74e-10 ***
majo         3.562e+00  5.210e-01   6.838 8.02e-12 ***
bcom        5.536e-01  6.182e-02   8.955 < 2e-16 ***
prmptf      -2.685e-05  8.818e-06  -3.045 0.002327 **
anconcontrat -2.223e-02  2.108e-03 -10.543 < 2e-16 ***
majo_1       7.992e-01  5.455e-01   1.465 0.142936
majo_2      -2.083e+00  5.848e-01  -3.561 0.000369 ***
top_sinistre12mois -5.869e+00  5.002e-01 -11.734 < 2e-16 ***
top_multiequipe -1.936e+00  3.814e-02 -50.761 < 2e-16 ***
top_multigamme  3.326e-01  5.652e-02   5.884 4.00e-09 ***
topmultiIARD  4.418e-01  2.321e-02  19.036 < 2e-16 ***
top_multisites -3.045e-01  4.260e-02  -7.148 8.82e-13 ***
topcodeaffaire  1.415e-01  2.966e-02   4.770 1.84e-06 ***
topNonProprio  2.567e-01  4.305e-02   5.962 2.49e-09 ***
topres12mois  5.954e-01  2.927e-02  20.341 < 2e-16 ***
topLOLR12mois  4.407e-01  2.309e-02  19.085 < 2e-16 ***
toSGRSREC12mois 6.351e-01  4.512e-02  14.076 < 2e-16 ***
codereseauCourtier -1.413e-01  2.174e-02  -6.503 7.90e-11 ***
ClassementStandard 4.846e-02  2.354e-02   2.059 0.039530 *
ClassementSupA@rieur 6.306e-03  2.369e-02   0.266 0.790091
top_mensuel  2.436e-01  2.182e-02  11.163 < 2e-16 ***
type_produit100% PRO ARTISANS-COMMERCANTS -2.739e-02  2.907e-02  -0.942 0.346102
type_produit100% PRO ASSOCIATION -1.041e+00  5.132e-02 -20.283 < 2e-16 ***
type_produit100% PRO FABRICATION  3.011e-01  5.299e-02   5.682 1.33e-08 ***
type_produit100% PRO SERVICE -2.163e-02  3.304e-02  -0.655 0.512664
type_produitANCIENNE GENERATION  3.470e-01  7.141e-02   4.859 1.18e-06 ***
famille_activite_redComptabilit@ - Finance -6.733e-01  1.630e-01  -4.131 3.61e-05 ***
famille_activite_redExpertise -9.414e-01  6.042e-01  -1.558 0.119171
famille_activite_redJuridique -2.584e-01  1.205e-01  -2.144 0.031995 *
famille_activite_redSoins aux animaux -6.825e-01  2.908e-01  -2.347 0.018928 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 90791 on 127123 degrees of freedom
Residual deviance: 81537 on 127094 degrees of freedom
AIC: 81597

Number of Fisher Scoring iterations: 10
```

Les variables retenues dans le modèle sont celles qui font suite à la procédure Stepwise décrite dans la partie validation et qui donnent les meilleures performances prédictives des résiliations. Les variables telles que la dernière gamme de produit souscrit et la dernière gamme de contrat résilié ne se sont pas révélées significatives dans notre modèle.

Les variables les plus importantes sont :

- Le multi équipement
- Les retards de paiement sur un des contrats du client dans les 12 mois précédant
- Les contentieux sur un des contrats du client dans les 12 mois précédant

- La résiliation d'un autre contrat du client dans les 12 mois précédant l'échéance
- Le nombre de sinistre sur 12 mois précédant l'échéance.

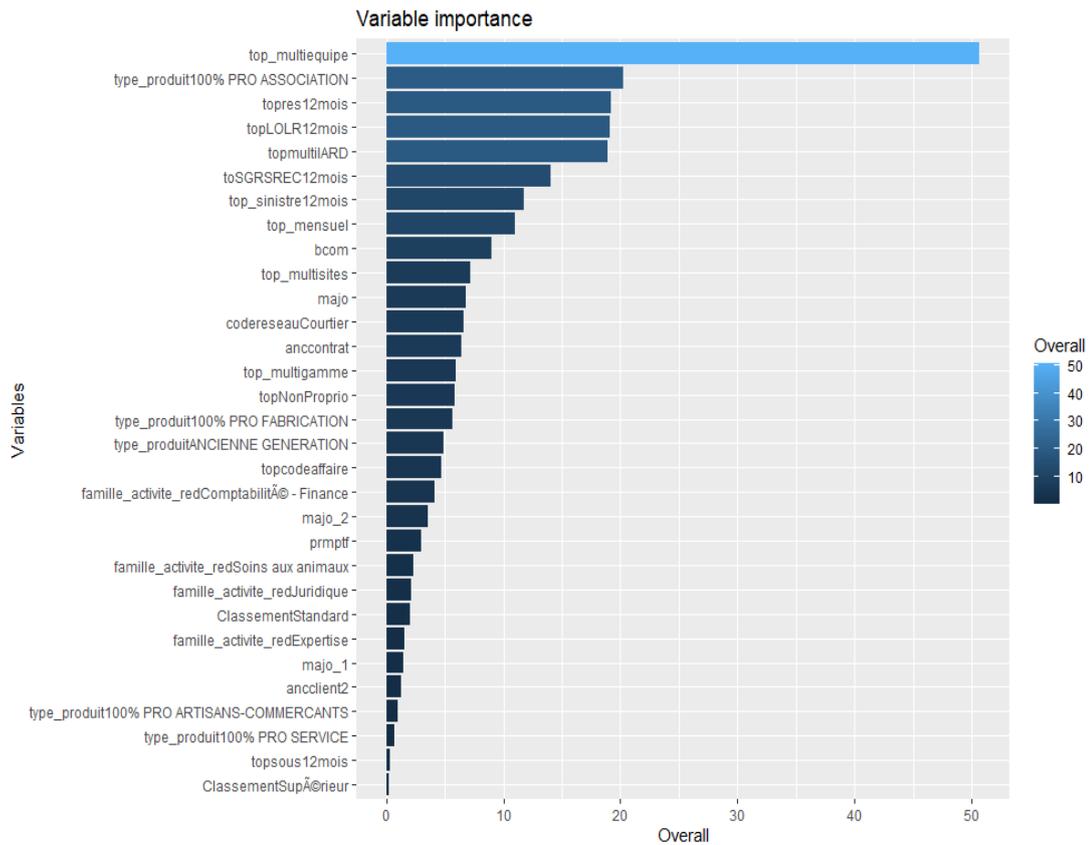


FIGURE 16 : CLASSEMENT DES VARIABLES PAR IMPORTANCE DANS LE MODELE GLM

Nous représentons la courbe ROC de notre modèle dans le graphique suivant :

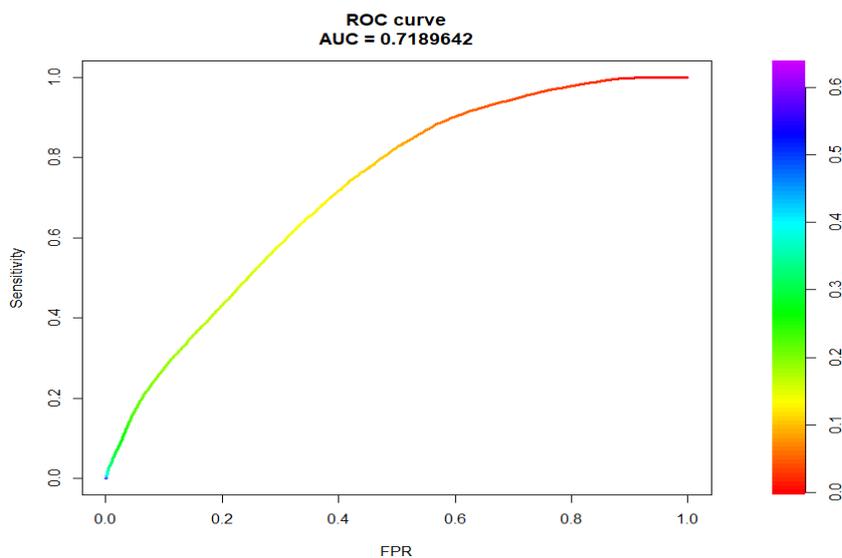


FIGURE 17 : COURBE ROC DU MODELE GLM SUR DONNEES D'EVALUATION

Avec un AUC de 72 % et un taux d'erreur de classification de 24%, nous pouvons dire que nous avons une très bonne performance.

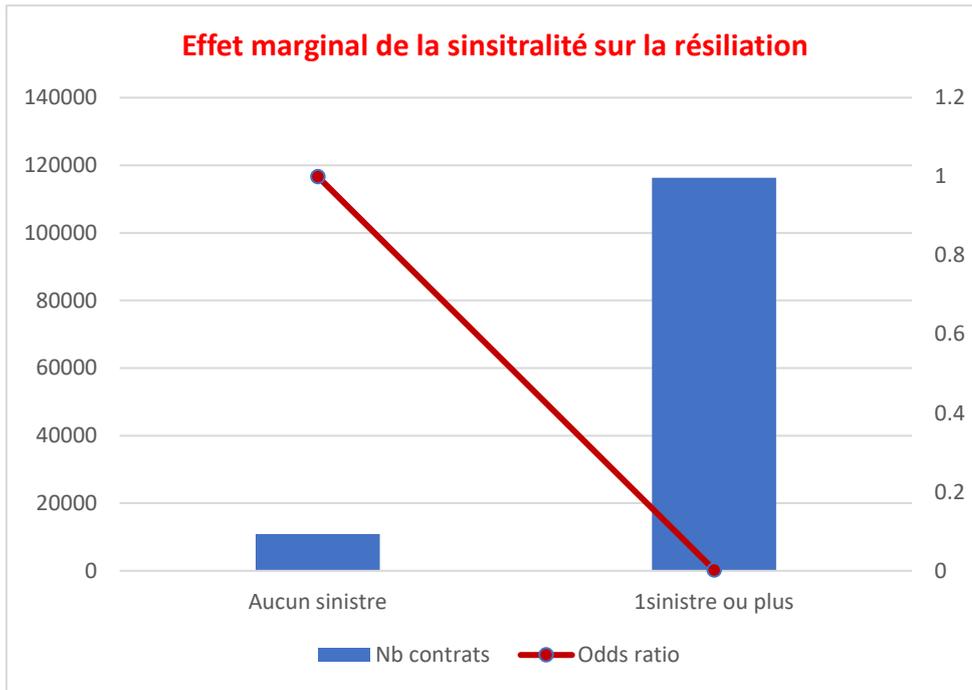
Un des avantages des GLM est qu'ils permettent d'estimer les odds ratios qui sont présentés dans la figure suivante.

Variables	Odds ratio	2.5 %	97.5 %	Pr(> z)	Significativité
(Intercept)	0.009727850	0.002065145	0.045822972	4.66e-09	***
majo	29.356769532	10.607279994	81.247965343	7.68e-11	***
bcom	1.697049566	1.504022091	1.914850352	< 2e-16	***
prmpft	0.999977072	0.999959834	0.999994310	0.009138	**
ancclient2	1.005035547	1.000439629	1.009652578	0.031720	*
anconcontrat	0.974756783	0.969104717	0.980441813	< 2e-16	***
majo_1	2.345628593	0.808427097	6.805775707	0.116725	
majo_2	0.116744451	0.037172917	0.366645069	0.000235	***
top_sinistre12mois	0.002746325	0.001030378	0.007319939	< 2e-16	***
top_multiequipe	0.187771765	0.175044154	0.201424811	< 2e-16	***
top_multigamme	1.309967000	1.172811427	1.463162365	1.71e-06	***
top_multisites	0.760990852	0.700258078	0.826990926	1.22e-10	***
topcodeaffaire	1.126768277	1.063190004	1.194148502	5.63e-05	***
topNonProprio	1.260871194	1.159001325	1.371694866	6.93e-08	***
topres12mois	1.983375589	1.872724120	2.100564992	< 2e-16	***
topLOLR12mois	1.562419146	1.493414951	1.634611723	< 2e-16	***
toSGRSREC12mois	1.881536975	1.722645119	2.055084561	< 2e-16	***
codereseauCourtier	0.831926183	0.797410282	0.867936103	< 2e-16	***
ClassementStandard	1.059134820	1.011503516	1.109009062	0.014399	*
ClassementSupAncien	1.014443292	0.968539291	1.062522916	0.543880	
top_mensuel	1.312671480	1.257695451	1.370050606	< 2e-16	***
type_produit100% PROARTISANS- COMMERCANTS	0.968517834	0.914972067	1.025197192	0.270296	
type_produit100% PROASSOCIATION	0.330431426	0.298903270	0.365285156	< 2e-16	***
type_produit100% PROFABRICATION	1.449101477	1.307020222	1.606627851	1.85e-12	***
type_produit100% PROSERVICE	0.950836126	0.892014969	1.013536061	0.121791	
type_produitANCIENNEGENERATION	1.391191827	1.209858065	1.599703929	3.60e-06	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

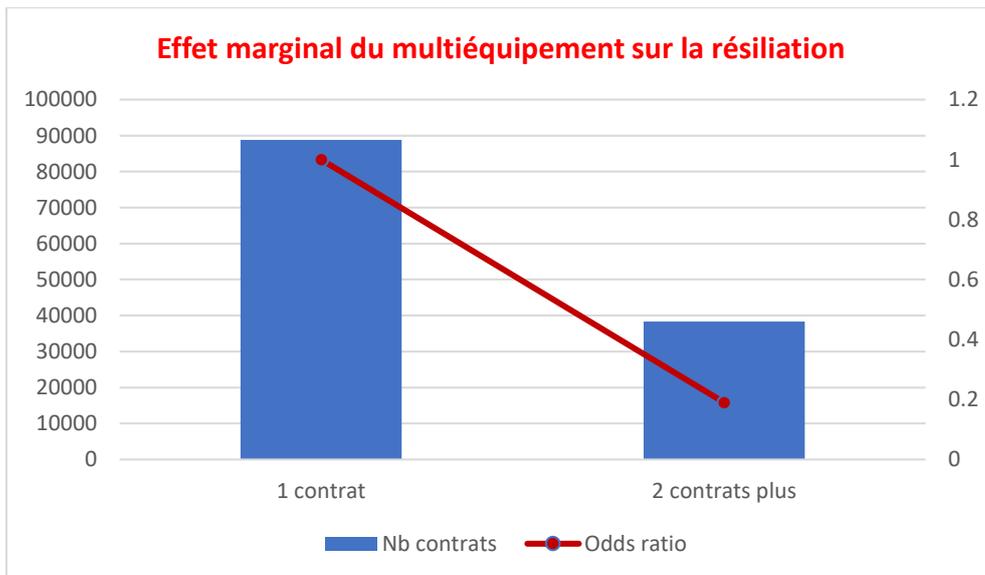
FIGURE 18 : ODDS RATIOS DU MODELE GLM SUR TOUS LES PRODUITS

Toutes choses égales par ailleurs :

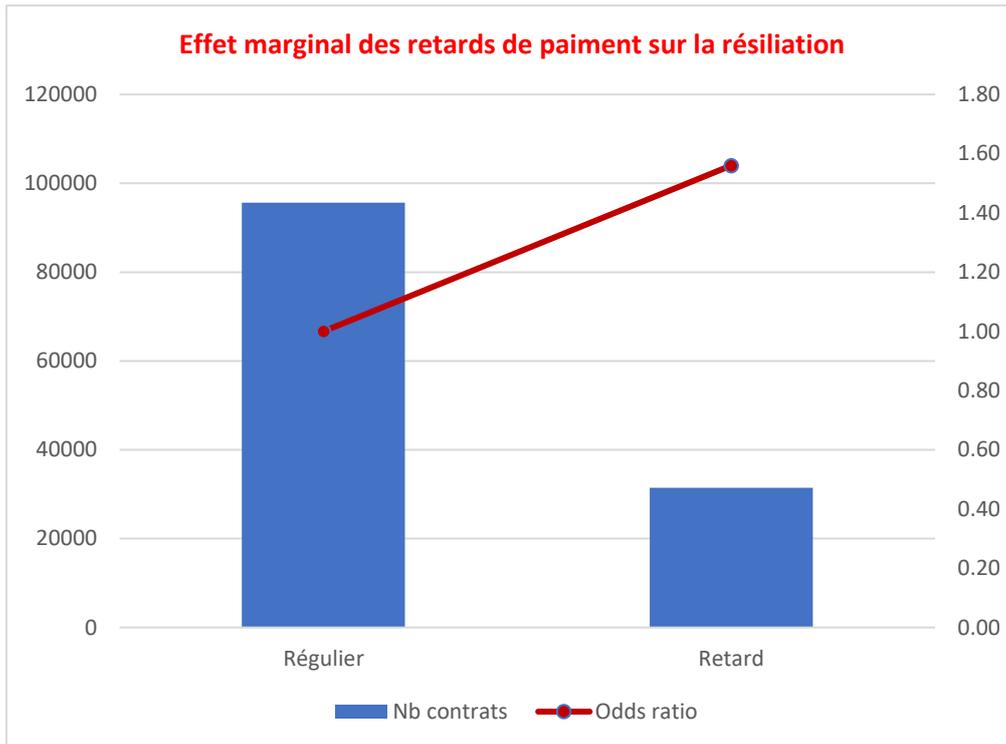
- La sinistralité : Le fait d'avoir un sinistre annule pratiquement la probabilité de résilier par suite d'une revalorisation. Le taux de client sinistré représente 8.6% sur la base d'entraînement. Cet effet est attendu du fait que les clients sinistrés s'attendent à une revalorisation de leur prime et sont moins enclins à résilier par suite d'une revalorisation. Les clients sinistrés s'ils résilient sont dans l'obligation de déclarer leurs sinistres chez le nouvel assureur potentiel. Ce qui peut leur coûter cher en termes de nouvelle prime à payer ailleurs. Ce qui pourrait expliquer le ralentissement des résiliations des clients sinistrés. On peut également attribuer cela à une bonne satisfaction des clients sinistrés qui sont bien indemnisés.



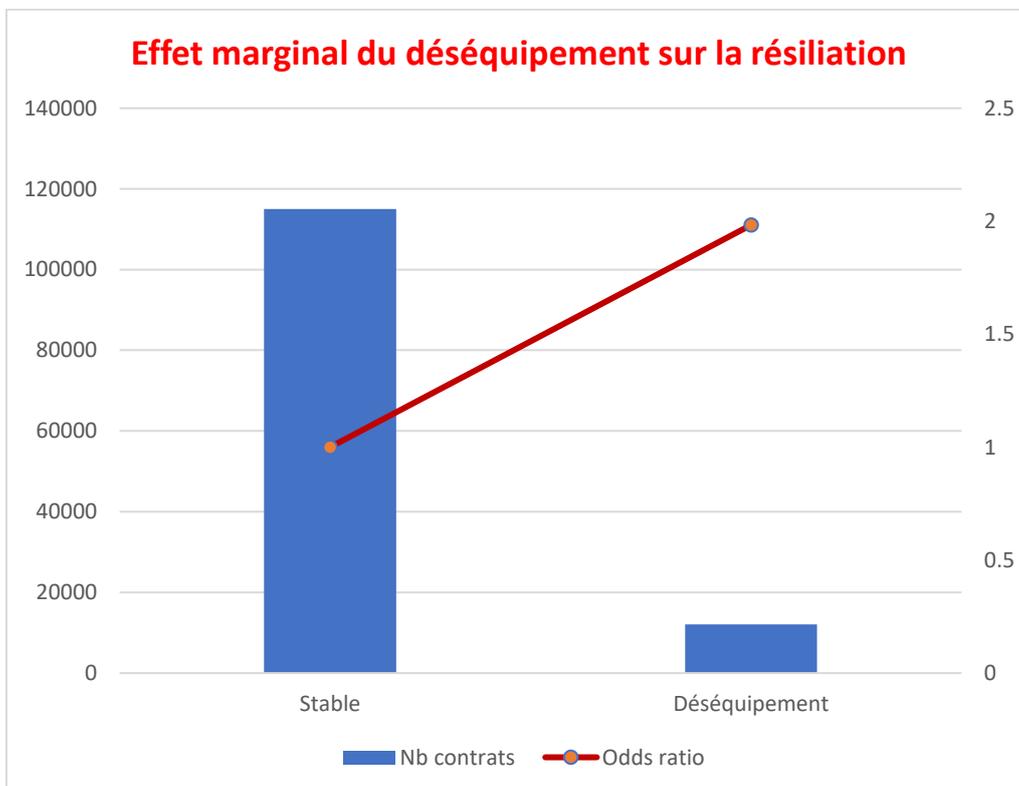
- Les clients multi équipés résilient jusqu'à 80% moins que les clients monos équipés.



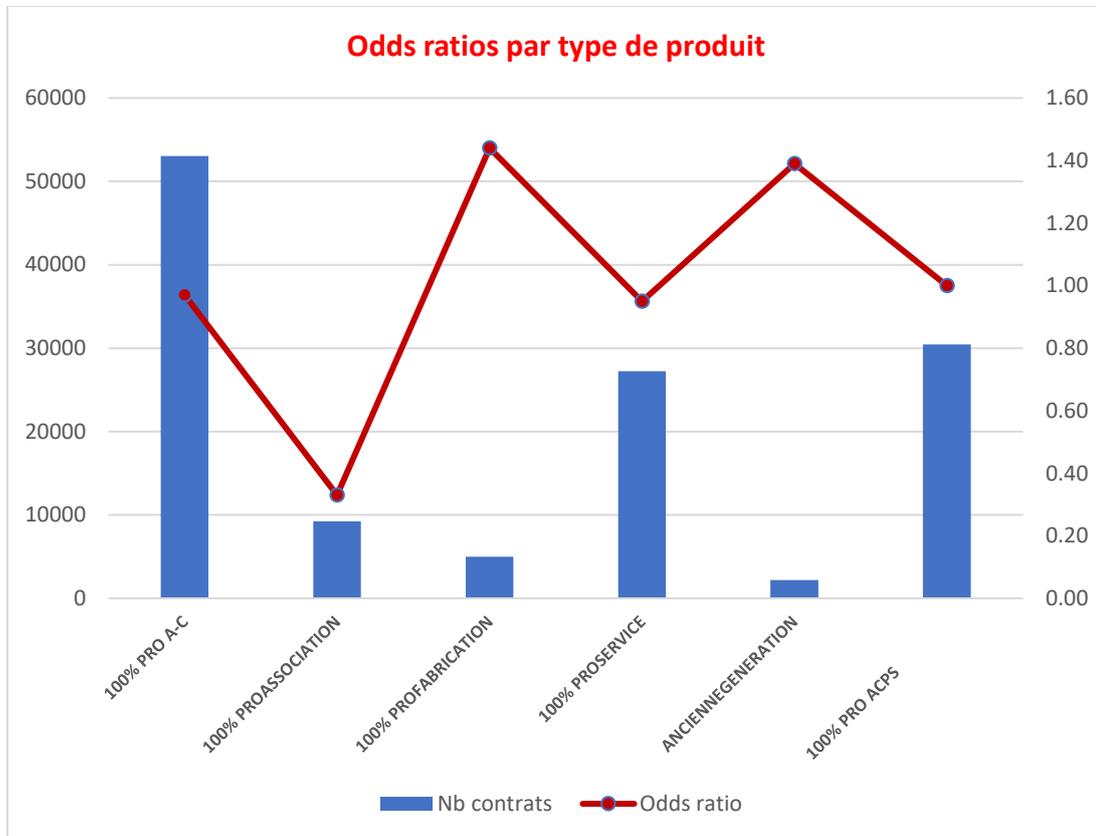
- Les clients qui ont du retard de paiement ou qui ont des contentieux sur d'autres contrats résilient de 1.5 à 2 fois plus que les clients sans retard de paiement et sans contentieux.



- Le fait de se déséquiper augmenterait de 80 à 92% la probabilité de résiliation par rapport à une situation stable du portefeuille client. Ce qui est également attendu puisque le dés équipement est un annonciateur de départ du client. Aussi, suivant le type de contrat résilié, longue durée ou courte durée, le comportement peut à priori différer même si l'effet croisé du dés équipement avec le type de contrat résilié n'est pas significatif dans notre modèle.



- Pour les produits, pas de différence significative entre les produits ACPS, Artisans Commerçants et Pro Service. Cependant, les produits Pro Association résilieraient jusqu'à 64% moins que les 3 produits précédents. Les produits de l'ancienne génération et les produits Pro Fabrication résilieraient jusqu'à 60% plus que les produits ACPS.



Le coefficient de la variable majoration dans le modèle est **+3.56** ce qui représente l'élasticité moyenne du portefeuille à la majoration (au prix). Ce qui signifierait qu'une augmentation de l'écart de l'écart au tarif de 1% augmenterait le taux de résiliation de **3.56%**. Le taux de résiliation du portefeuille étant **11.5%** et la moyenne de majoration du portefeuille de **6.7%**, 1 % d'augmentation de l'écart au tarif sur le portefeuille correspond à un passage du taux de revalorisation de 6.7% à 8%. Le taux de résiliation deviendrait 11.5% (1+ 3.56%) = 11.9% ce qui représente une faible variation du taux de résiliation.

Afin d'améliorer l'interprétabilité de nos coefficients, nous allons discrétiser les variables continues à partir d'un modèle additif généralisé.

3.1.2. Modèles additifs généralisés (GAM)

a. Rappels théoriques

Un modèle additif généralisé est un modèle linéaire généralisé qui intègre des fonctions *lisses* ou de transformation des variables explicatives. Les fonctions lisses permettent d'estimer les relations non linéaires entre les variables explicatives et la variable cible.

Le modèle prend la structure simplifiée suivante :

$$g[E(Y_i)] = \beta_0 + \sum_{j=1}^k f_j(X_{ij})$$

Avec g la fonction de lien comme dans le cas des GLM, f_j les fonctions *lisses* et la loi de Y appartenant à la famille exponentielle.

La fonction lisse peut être univariée et être représentée à l'aide d'une *spline* cubique sous la forme générale :

$$f(x) = \sum_{i=1}^q b_i(x) * \beta_i$$

Une spline cubique est une courbe composée de sections de polynômes cubiques, reliées entre elles de manière qu'elles soient continues et que les dérivées premières et secondes de la fonction restent continues. Les points d'intersection des polynômes cubiques sont appelés les nœuds de la spline.

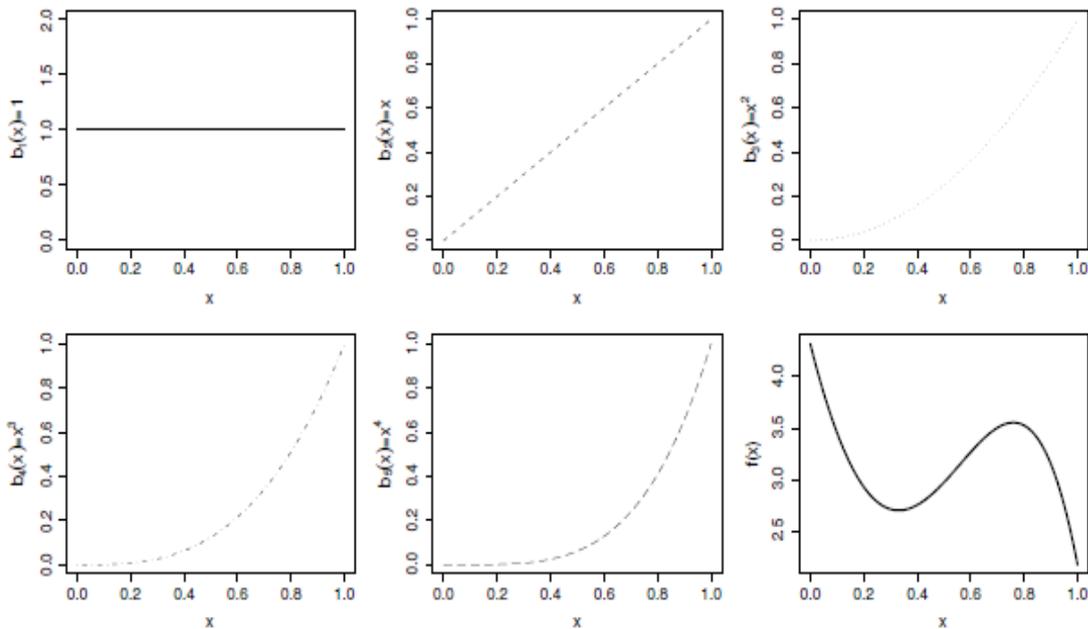


FIGURE 19 : EXEMPLE DE REPRESENTATION D'UNE FONCTION LISSE A L'AIDE D'UNE SPLINE CUBIQUE.

Les GAM sont une généralisation des modèles additifs comme les GLM pour les modèles linéaires.

Une structure simplifiée d'un modèle additif est :

$$y_i = f_1(x_i) + f_2(z_i) + \epsilon_i$$

L'estimation du modèle additif peut se faire par les moindres carrés pénalisés. La fonction de cout à minimiser dans ce cas est :

$$\| \mathbf{y} - \mathbf{X}\beta \|^2 + \lambda_1 \beta^T \mathbf{S}_1 \beta + \lambda_2 \beta^T \mathbf{S}_2 \beta$$

avec λ_1 et λ_2 les paramètres de lissage contrôlant les poids à accorder au lissage comparé à l'ajustement linéaire simple. Donc plus le lien entre X et Y n'est pas linéaire, et plus le poids du lissage va être important dans la fonction de cout.

Si on pose

$$S \equiv \lambda_1 S_1 + \lambda_2 S_2$$

alors la fonction de cout à minimiser devient :

$$\| \mathbf{y} - \mathbf{X}\beta \|^2 + \beta^T \mathbf{S}\beta = \left\| \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} - \begin{bmatrix} \mathbf{X} \\ \mathbf{B} \end{bmatrix} \beta \right\|^2$$

pour toute matrice B telle que $\mathbf{B}^T \mathbf{B} = \mathbf{S}$.

Si pour les modèles linéaires généralisés l'estimation se fait par maximum de vraisemblance, l'estimation des modèles additifs généralisés se fait par maximum de vraisemblance pénalisée encore appelée quasi-vraisemblance. Dans la pratique, il n'existe pas de procédé simple pour effectuer un GLM non pénalisé qui serait équivalent à une vraisemblance pénalisée. Donc une façon d'approcher la solution est l'utilisation des moindres carrés itératifs pénalisés (plus connus sous l'abréviation **P-IRLS**) jusqu'à la convergence.

Tout comme les modèles linéaires généralisés, nous allons utiliser les mêmes critères d'AIC, de BIC pour comparer les modèles. Également, une équivalence du Stepwise (StepGAM) permet sous R d'effectuer la procédure de sélection des variables.

b. Application.

L'application des modèles additifs généralisés nous donnent les premiers résultats suivants. Nous gardons les mêmes variables que dans le GLM tout en ajoutant des transformations de toutes les variables continues par des fonctions splines. Le résumé du modèle est présenté dans la figure suivante.

```
Family: binomial
Link function: logit

Formula:
top_resil_1_11_mois ~ s(majo) + s(bcom) + s(prmptf) + s(ancclient2) +
s(anccontrat) + s(majo_1) + top_sinistre12mois + top_multiequipe +
top_multigamme + top_multisites + topcodeaffaire + topNonProprio +
topres12mois + topLOLR12mois + toSGRSREC12mois + codereseau +
Classement + top_mensuel + type_produit + famille_activite_red

Parametric coefficients:

```

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.021753	0.086858	-23.277	< 2e-16	***
top_sinistre12mois	-5.881302	0.500273	-11.756	< 2e-16	***
top_multiequipe	-1.673545	0.035920	-46.591	< 2e-16	***
top_multigamme	0.284591	0.056486	5.038	4.70e-07	***
top_multisites	-0.217917	0.042825	-5.089	3.61e-07	***
topcodeaffaire	0.102628	0.030232	3.395	0.000687	***
topNonProprio	0.244661	0.043358	5.643	1.67e-08	***
topres12mois	0.710669	0.029529	24.067	< 2e-16	***
topLOLR12mois	0.457333	0.023084	19.812	< 2e-16	***
toSGRSREC12mois	0.628832	0.045131	13.934	< 2e-16	***
codereseauCourtier	-0.178786	0.021767	-8.214	< 2e-16	***
ClassementStandard	0.041993	0.023598	1.780	0.075157	.
ClassementSupA@rieur	0.001521	0.023690	0.064	0.948822	
top_mensuel	0.283122	0.022039	12.847	< 2e-16	***
type_produit100% PRO ARTISANS-COMMERCANTS	-0.075065	0.032439	-2.314	0.020668	*
type_produit100% PRO ASSOCIATION	-1.236523	0.053686	-23.033	< 2e-16	***
type_produit100% PRO FABRICATION	0.347734	0.053651	6.481	9.09e-11	***
type_produit100% PRO SERVICE	-0.171910	0.034899	-4.926	8.40e-07	***
type_produitANCIENNE GENERATION	0.238233	0.073012	3.263	0.001103	**
famille_activite_redAutos cycles bateaux a@ronautique	0.007141	0.098919	0.072	0.942447	
famille_activite_redAutres	-0.014309	0.069363	-0.206	0.836559	
famille_activite_redElectronique A@lectricit@	-0.105889	0.145955	-0.725	0.468151	
famille_activite_redEnseignement	0.089759	0.200673	0.447	0.654668	
famille_activite_redHabillement mode et accessoire	-0.042923	0.080940	-0.530	0.595900	
famille_activite_redInformatique Electrom@nager T@l@phonie	-0.222131	0.112524	-1.974	0.048374	*
famille_activite_redLoisirs musique	-0.057701	0.102198	-0.565	0.572344	
famille_activite_redPapier Impression	-0.172780	0.129265	-1.337	0.181342	

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

```

	edf	Ref.df	Chi.sq	p-value
s(majo)	1.086	1.168	41.60	<2e-16 ***
s(bcom)	5.994	7.097	82.00	<2e-16 ***
s(prmptf)	6.161	7.312	97.81	<2e-16 ***
s(ancclient2)	4.609	5.668	74.93	<2e-16 ***
s(anccontrat)	1.478	1.817	50.21	<2e-16 ***
s(majo_1)	4.132	5.050	13.00	0.0232 *

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nous retrouvons les mêmes variables significatives que dans le cas du GLM. L'ajout des fonctions splines permettent d'établir les liens suivants de la résiliation et les variables continues.

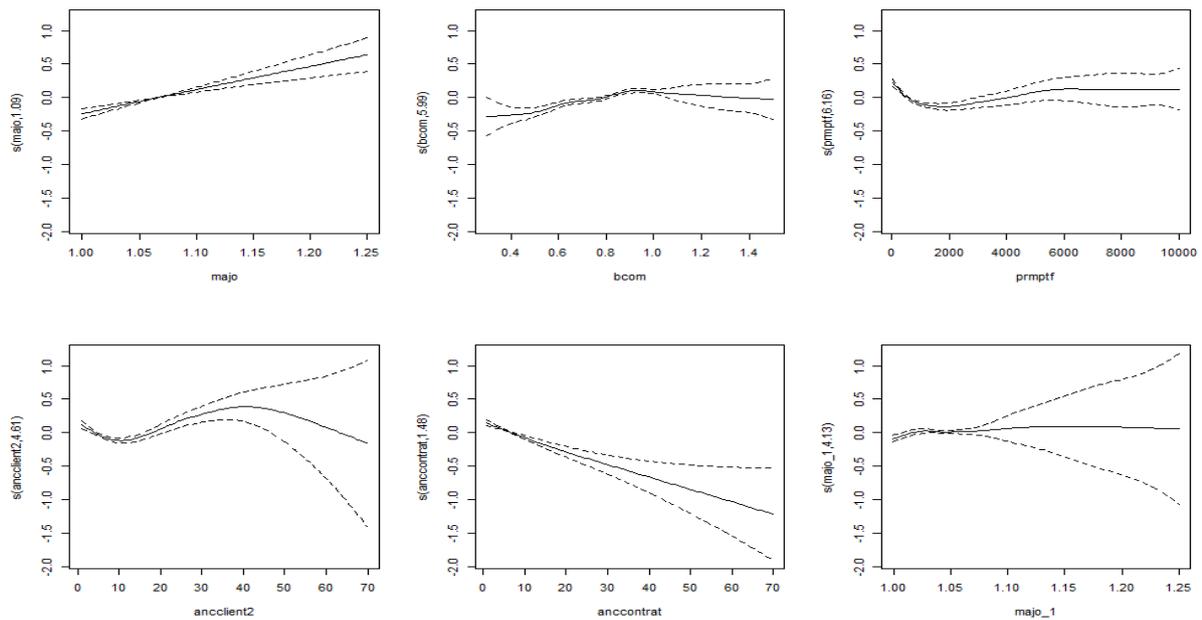


FIGURE 20 : TAUX DE RESILIATION EN FONCTION DES VARIABLES CONTINUES (GAM)

Le lien de la majoration et la résiliation semble être linéaire. Donc nous gardons la variable majoration telle qu'elle est dans le modèle. Cependant, le lien du rabais commercial initial avec la résiliation n'est pas parfaitement linéaire. On a une tendance générale à la croissance du taux de résiliation en fonction du rabais initial accordé mais la relation n'est pas parfaitement linéaire. De façon optimale, le rabais commercial peut être regroupé en 5 tranches de rabais. La résiliation décroît linéairement avec l'ancienneté contrat mais la tendance en fonction de l'ancienneté client n'est pas monotone. Aussi, la résiliation décroît en fonction de l'ancienneté contrat mais n'a pas de tendance monotone avec l'ancienneté client.

Nous découpons les variables en tranches et le nouveau résumé du modèle est présenté dans la figure suivante.

Formula:
top_resil_1_11_mois ~ majo + BCOM + PRMPTF + ANCCclient + anconcontrat +
MAJO_1 + top_sinistres12mois + top_multiequipe + top_multigamme +
top_multisites + topcodeaffaire + topNonProprio + topres12mois +
topLQLR12mois + toSGRSREC12mois + codereseau + Classement +
top_mensuel + topres12mois + majo:(ANCCclient + type_produit)

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-7.534948	0.740857	-10.171	< 2e-16 ***
majo	5.248944	0.690555	7.601	2.94e-14 ***
BCOM20 à 40%	0.123932	0.031354	3.953	7.73e-05 ***
BCOMjusqu'à 20%	0.262776	0.031902	8.237	< 2e-16 ***
BCOMpas de rabais	0.167285	0.072919	2.294	0.021784 *
PRMPTF600-1000€	-0.118057	0.024850	-4.751	2.03e-06 ***
PRMPTF1000-1500€	-0.186930	0.030635	-6.102	1.05e-09 ***
PRMPTF1500-2000€	-0.237945	0.040736	-5.841	5.18e-09 ***
PRMPTF2000€ +	-0.110897	0.032503	-3.412	0.000645 ***
ANCCclient5-100ans	2.542076	1.231050	2.065	0.038926 *
ANCCclient10-20ans	3.818470	1.330147	2.871	0.004095 **
ANCCclient20ans +	5.273863	2.064600	2.554	0.010636 *
anconcontrat	-0.021772	0.002698	-8.069	7.09e-16 ***
MAJO_1plus de 5%	0.029750	0.021446	1.387	0.165374 .
top_sinistres12mois	-5.883905	0.500185	-11.763	< 2e-16 ***
top_multiequipe	-1.668335	0.035800	-46.601	< 2e-16 ***
top_multigamme	0.276097	0.056466	4.890	1.01e-06 ***
top_multisites	-0.231644	0.042599	-5.438	5.39e-08 ***
topcodeaffaire	0.115805	0.029619	3.910	9.23e-05 ***
topNonProprio	0.234914	0.043067	5.455	4.91e-08 ***
topres12mois	0.700540	0.029239	23.959	< 2e-16 ***
topLQLR12mois	0.455586	0.023054	19.762	< 2e-16 ***
toSGRSREC12mois	0.627152	0.045093	13.908	< 2e-16 ***
codereseauCourtier	-0.182853	0.021677	-8.435	< 2e-16 ***
ClassementStandard	0.043564	0.023574	1.848	0.064612 .
ClassementSupA@rieur	0.006449	0.023655	0.273	0.785128 .
top_mensuel	0.276878	0.021898	12.644	< 2e-16 ***
majo:ANCCclient5-100ans	-2.529346	1.152214	-2.195	0.028149 *
majo:ANCCclient10-20ans	-3.693063	1.246762	-2.962	0.003055 **
majo:ANCCclient20ans +	-4.788668	1.939673	-2.469	0.013557 *
majo:type_produit100% PRO ARTISANS-COMMERCANTS	-0.084622	0.028864	-2.932	0.003371 **
majo:type_produit100% PRO ASSOCIATION	-1.069194	0.047912	-22.316	< 2e-16 ***
majo:type_produit100% PRO FABRICATION	0.351502	0.049106	7.158	8.19e-13 ***
majo:type_produit100% PRO SERVICE	-0.113323	0.032536	-3.483	0.000496 ***
majo:type_produitANCIENNE GENERATION	0.228572	0.067922	3.365	0.000765 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.0663 Deviance explained = 9.95%
UBRE = -0.35635 Scale est. = 1 n = 127124

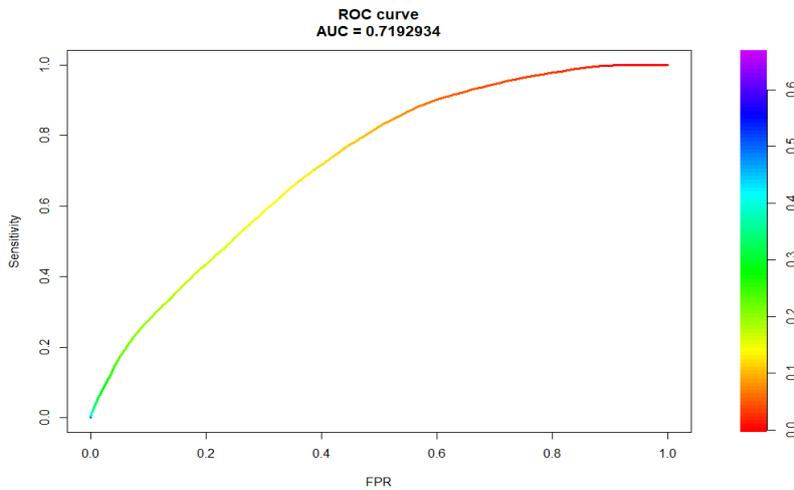


FIGURE 21 : COURBE ROC DU MODELE GAM SUR DONNEES D'EVALUATION

TABEAU 4 : ODDS RATIOS DU MODELE GAM SUR TOUS LES PRODUITS

Variables	Odds ratio	2.50%	97.50%	Pr(> z)	Significativité
(Intercept)	0	0	0.01	< 2e-16	***
majo	36.9	13.22	102.98	5.52E-12	***
BCOM20 à 40%	1.13	1.07	1.2	6.84E-05	***
BCOMJusqu'à 20%	1.3	1.22	1.39	< 2e-16	***
BCOMpas de rabais	1.18	1.03	1.37	0.020024	*
PRMPTF600-1000€	0.89	0.85	0.93	2.43E-06	***
PRMPTF1000-1500€	0.83	0.78	0.88	1.48E-09	***
PRMPTF1500-2000€	0.79	0.73	0.85	6.01E-09	***
PRMPTF2000€ +	0.9	0.84	0.96	0.000773	***
ANCclient5-10ans	0.85	0.81	0.9	7.52E-09	***
ANCclient10-20ans	0.88	0.83	0.95	0.000442	***
ANCclient20ans +	1.19	1.07	1.32	0.001451	**
ancontrat	0.98	0.97	0.98	6.40E-16	***
MAJO_1plus de 5%	1.03	0.98	1.07	0.226605	
top_sinistre12mois	0	0	0.01	< 2e-16	***
top_multiequipe	0.19	0.18	0.2	< 2e-16	***
top_multigamme	1.32	1.18	1.48	6.89E-07	***
top_multisites	0.79	0.73	0.86	6.40E-08	***
topcodeaffaire	1.12	1.05	1.18	0.000194	***
topNonProprio	1.26	1.16	1.38	4.93E-08	***
topres12mois	2.02	1.91	2.14	< 2e-16	***
topLOLR12mois	1.58	1.51	1.65	< 2e-16	***
toSGRSREC12mois	1.87	1.71	2.04	< 2e-16	***
codereseauCourtier	0.83	0.8	0.87	< 2e-16	***
ClassementStandard	1.04	1	1.09	0.066219	.
ClassementSupÃ©rieur	1.01	0.96	1.05	0.769858	
top_mensuel	1.32	1.26	1.37	< 2e-16	***
type_produit100% PRO ARTISANS-COMMERCANTS	0.91	0.86	0.97	.003763 *	*
type_produit100% PRO ASSOCIATION	0.31	0.28	0.34	< 2e-16 *	**
type_produit100% PRO FABRICATION	1.43	1.29	1.59	.16e-11 *	**
type_produit100% PRO SERVICE	0.88	0.82	0.94	.000140 *	**
type_produitANCIENNE GENERATION	1.29	1.12	1.48	0.000478	***

Le découpage en tranche des variables primes et ancienneté permet d'avoir des Odds ratio par groupes. Toutes choses égales par ailleurs :

- La tranche des petites primes (moins de 600€) résilie plus que les autres.
- Les clients d'ancienneté inférieure à 5 ans résilient plus que la tranche 5-20 ans mais moins que les clients plus anciens (20ans+).

L'AUC est 72% sur les données d'entraînement et de test. L'élasticité moyenne à la majoration reste de 3.6% sur le portefeuille MRC.

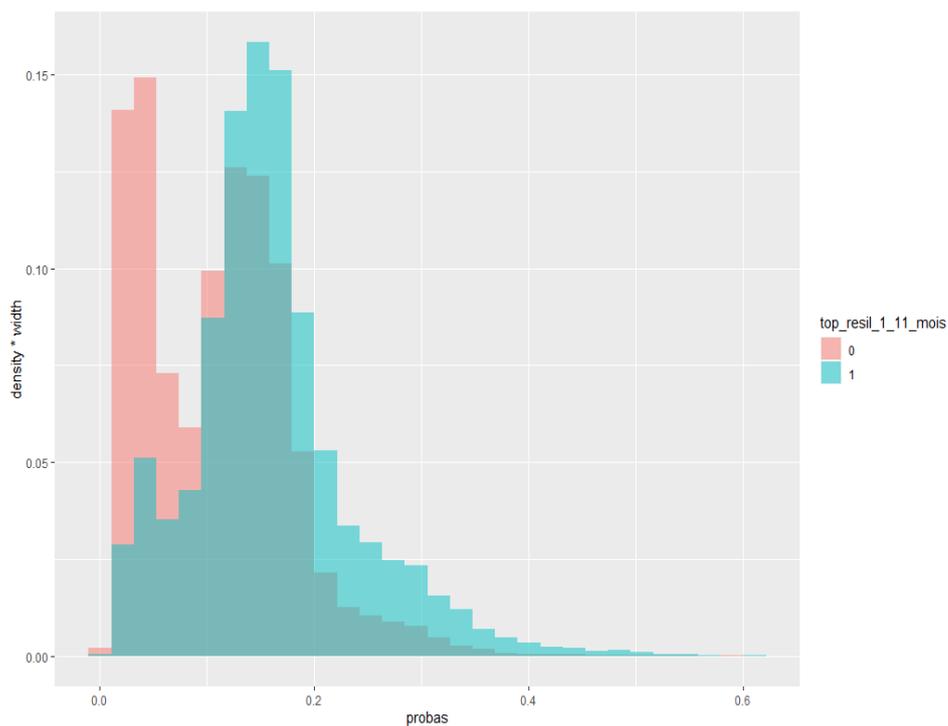


FIGURE 22: COMPARAISON DES PROBABILITES DE RESILIATION POUR LES CONTRATS RESILIES ET RETENUS

Les probabilités estimées par les GLM et les GAM restent cependant faibles bien pour les vraies résiliations que les rétentions. Pour les GLM et les GAM la probabilité moyenne estimée des contrats retenus est de 10.7% contre 17.7% pour les vraies résiliations. Nous utilisons par la suite les Gradients boosting pour améliorer la détection des résiliations.

3.1.3. Gradient Boosting

a. Rappels théoriques sur les Arbres de décision

Les arbres de décision sont des algorithmes qui divisent l'ensemble de données en des coupes successives dans l'espace des variables de l'ensemble de données. Par analogie à un arbre, la feuille représente la prédiction de Y (la variable cible) et les nœuds sont construits sous des conditions $X_i \leq \alpha$ contre $X_i \geq \alpha$, où X_i est la i -ème composante de la matrice des variables explicatives et α un seuil de découpage. À chaque nœud, l'algorithme détermine la meilleure variable (X_i) à utiliser et le point à couper (le seuil α) afin d'obtenir le meilleur critère de découpage de la population en deux sous-ensembles les plus homogènes. Les arbres de décision que nous utilisons sont binaires.

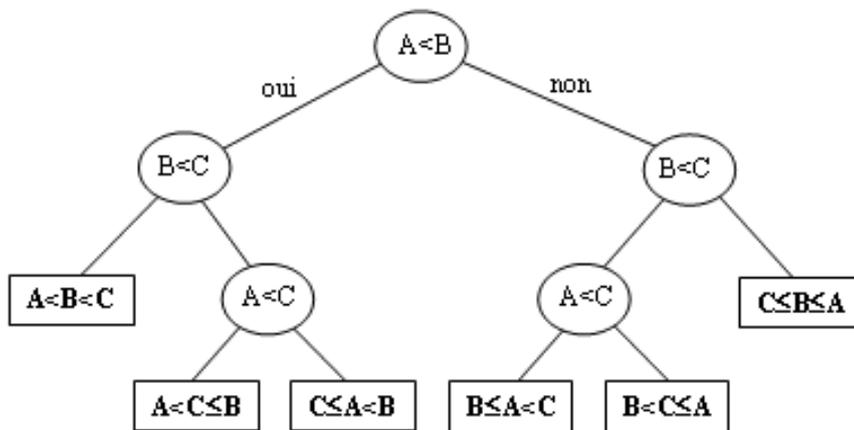


FIGURE 23 : EXEMPLE D'ARBRE DE DECISION

b. Rappels théoriques sur le Gradient Boosting

Le gradient boosting est l'une des méthodes de prédictions les plus performantes des méthodes de machine Learning. L'idée du boosting consiste à convertir des hypothèses ou des modèles ou des apprenants faibles en de très bonnes hypothèses ou de très bons modèles. Un apprenant faible correspond à un modèle qui performe un peu mieux que le hasard. Le boosting se réfère donc à la conversion d'un ensemble de règles ou de prédictions approximatives en de règles plus précises dans l'explication du problème étudié. Par exemple, les modèles faibles sont des arbres de décisions (Un exemple est présenté dans la Figure 24).

Le problème d'optimisation du gradient boosting est la minimisation d'une fonction de coût en ajoutant des règles faibles par une procédure de descente du gradient. L'algorithme comporte trois paramètres :

- La fonction de coût (loss function) :

La fonction de coût dépend du problème étudié. Elle doit être différentiable.

- Un apprenant faible pour les prédictions

Les modèles utilisés sont les arbres de décisions le plus souvent mais peuvent être également des modèles linéaires.

- Un modèle additif pour ajouter de nouvelles règles et donc de nouveaux modèles pour minimiser la fonction de cout :

Les modèles sont ajoutés de façon itérative. La procédure de la descente du gradient est utilisée pour minimiser la fonction de cout.

Les gradient boosting sont des algorithmes qui peuvent tomber très rapidement dans le surapprentissage ou l'overfitting d'où la nécessité de limiter le nombre d'arbres et la profondeur des arbres utilisés. Pour ce faire, il existe une suite d'améliorations du Gradient Boosting.

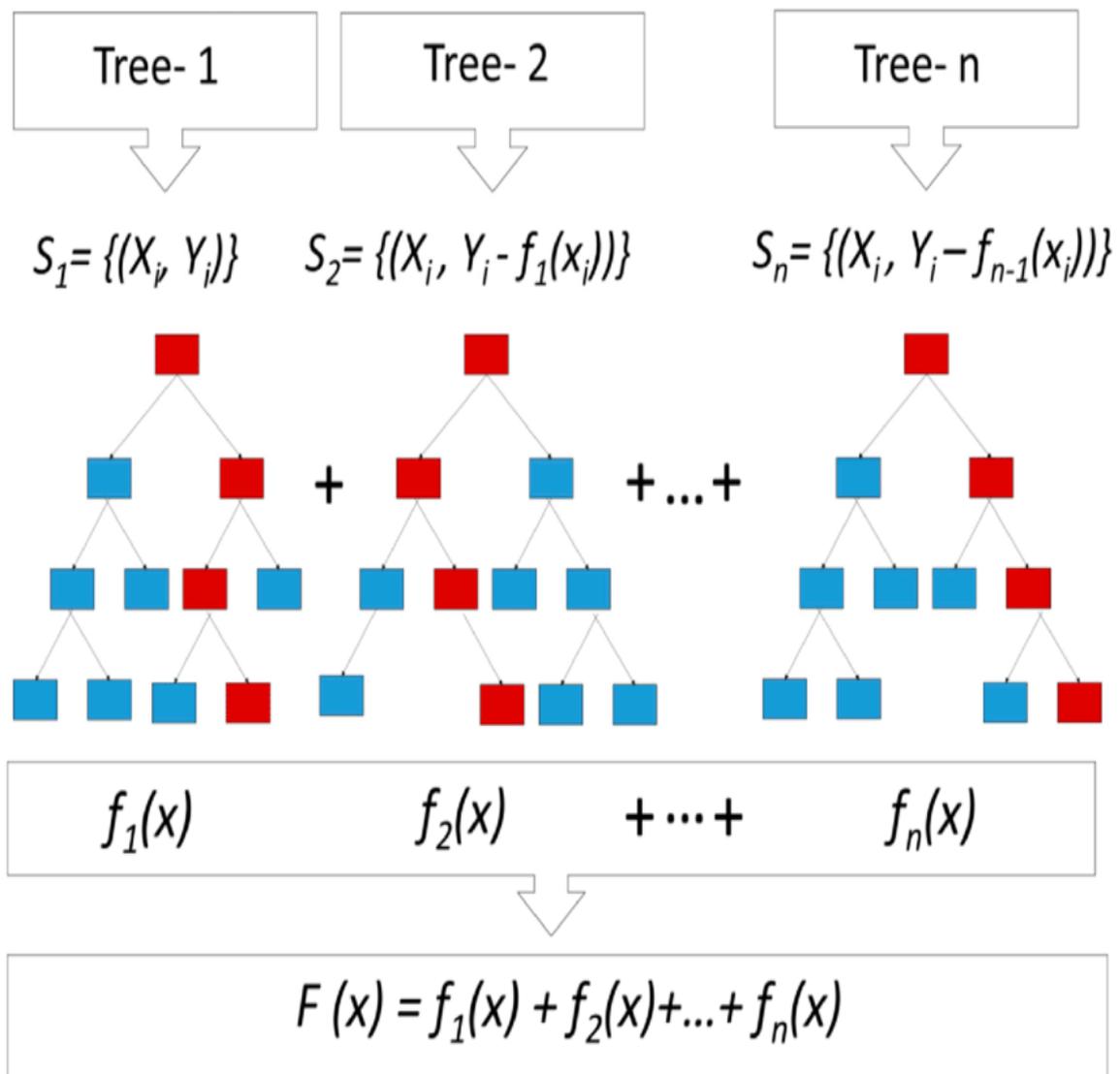


FIGURE 24 : DIAGRAMME D'UN GRADIENT BOOSTING BASE SUR LES ARBRES DE DECISION

Améliorations du gradient boosting

Le gradient boosting est un algorithme gourmand qui peut rapidement surapprendre sur les données d'entraînement. D'où l'utilisation des méthodes dites de régularisation qui pénalisent diverses parties de l'algorithme et améliorent généralement les performances de l'algorithme en réduisant l'overfitting. Les points de régularisation sont :

1. Contraintes des arbres.

Il est important que les apprenants faibles aient des compétences mais restent faibles. Il existe plusieurs façons de contraindre les arbres et plus la création de l'arbre est contrainte, plus le nombre d'arbres différents créés est élevé.

Aussi, plus la création d'un arbre est contrainte, plus il faudra d'arbres dans le modèle, et inversement, ou moins les arbres individuels sont contraints, moins il y aura d'arbres dans le modèle. Les contraintes qui peuvent être imposées à la construction d'arbres de décision :

- Le nombre d'arbres : généralement l'ajout de plusieurs arbres au modèle peut réduire l'overfitting.
- La profondeur de l'arbre : les arbres plus profonds sont des arbres plus complexes et les arbres plus courts sont préférables.
- Le nombre de nœuds ou nombre de feuilles : comme la profondeur, cela peut limiter la taille de l'arbre et réduire l'overfitting.
- Le nombre d'observations par fractionnement impose une contrainte minimale sur la quantité de données de formation à un nœud de formation avant un fractionnement.

2. Rétrécissement.

Les prédictions de chaque arbre sont additionnées séquentiellement. La contribution de chaque arbre à la somme des prédictions peut être pondérée pour ralentir l'apprentissage par l'algorithme. Cette pondération est appelée le rétrécissement. Le rétrécissement réduit l'influence de chaque arbre individuel et laisse de la place aux arbres futurs pour améliorer le modèle.

3. Échantillonnage aléatoire.

Afin de réduire la corrélation entre les arbres, à chaque itération du gradient, un sous échantillon de l'échantillon est tiré au hasard pour l'entraînement du modèle.

4. Apprentissage pénalisé.

Consiste en une régularisation supplémentaire permettant de lisser les poids finaux appris afin d'éviter l'overfitting. Intuitivement, l'objectif régularisé aura tendance à sélectionner un modèle plus simple que complexe.

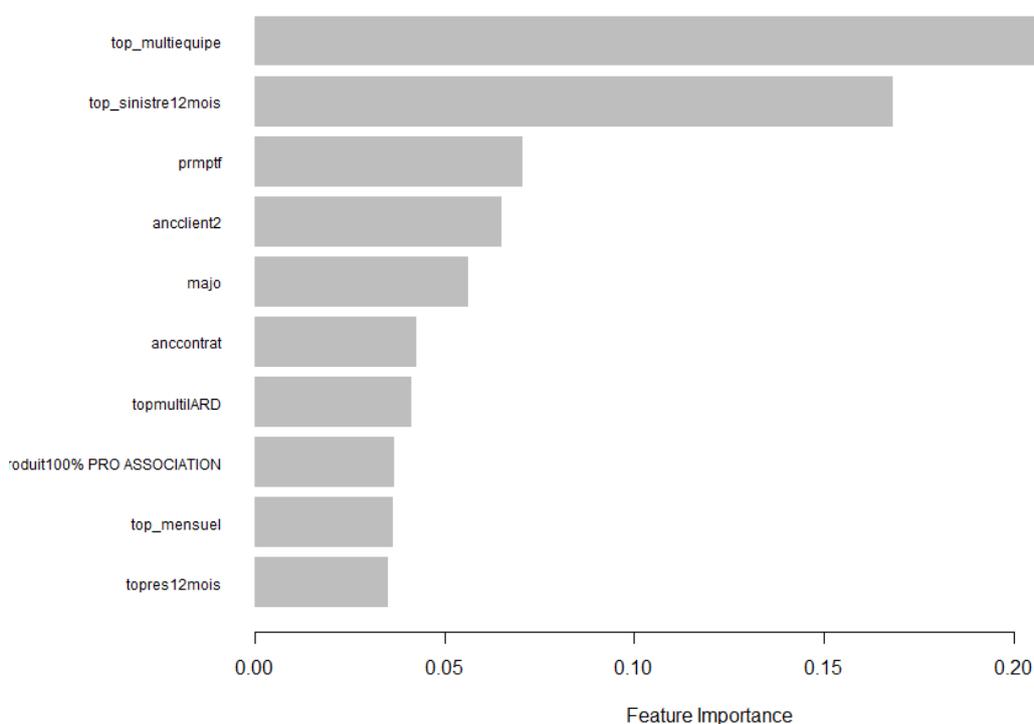
L'un des avantages de l'utilisation du boosting de gradient est qu'une fois les arbres boostés construits, il est relativement simple de récupérer les scores d'importance pour chaque variable. En général, l'importance fournit un score qui indique l'utilité ou la valeur de chaque caractéristique dans la construction des arbres de décision boostés du modèle. Plus une variable est utilisée pour prendre des décisions clés avec des arbres de décision, plus son importance relative est élevée. Cette importance est calculée explicitement pour chaque variable, ce qui permet de classer les variables et de les comparer entre elles.

L'importance d'une variable est calculée pour un seul arbre décisionnel et la mesure de performance peut être l'indice de Gini. La moyenne des importances des caractéristiques est ensuite calculée sur tous les arbres de décision du modèle. Dans notre cas la classification des variables nous permet de comprendre les caractéristiques les plus discriminantes dans la résiliation.

c. Applications.

L'application de la méthode du boosting sur nos données nous permet d'améliorer globalement l'estimation des probabilités de rétention des profils les plus sensibles et d'améliorer l'estimation de la probabilité de résiliation.

Les variables les plus importantes du modèle sont :



Dans ce modèle, nous observons pratiquement le même ordre d'importance des variables que dans le cas du GLM précédent. Le taux de résiliation prédit est de 11.5 % sur tout le portefeuille ce qui correspond au taux observé. L'erreur de classification est de 25% et l'AUC est de 76% sur données de test et 78% sur données d'entraînement.

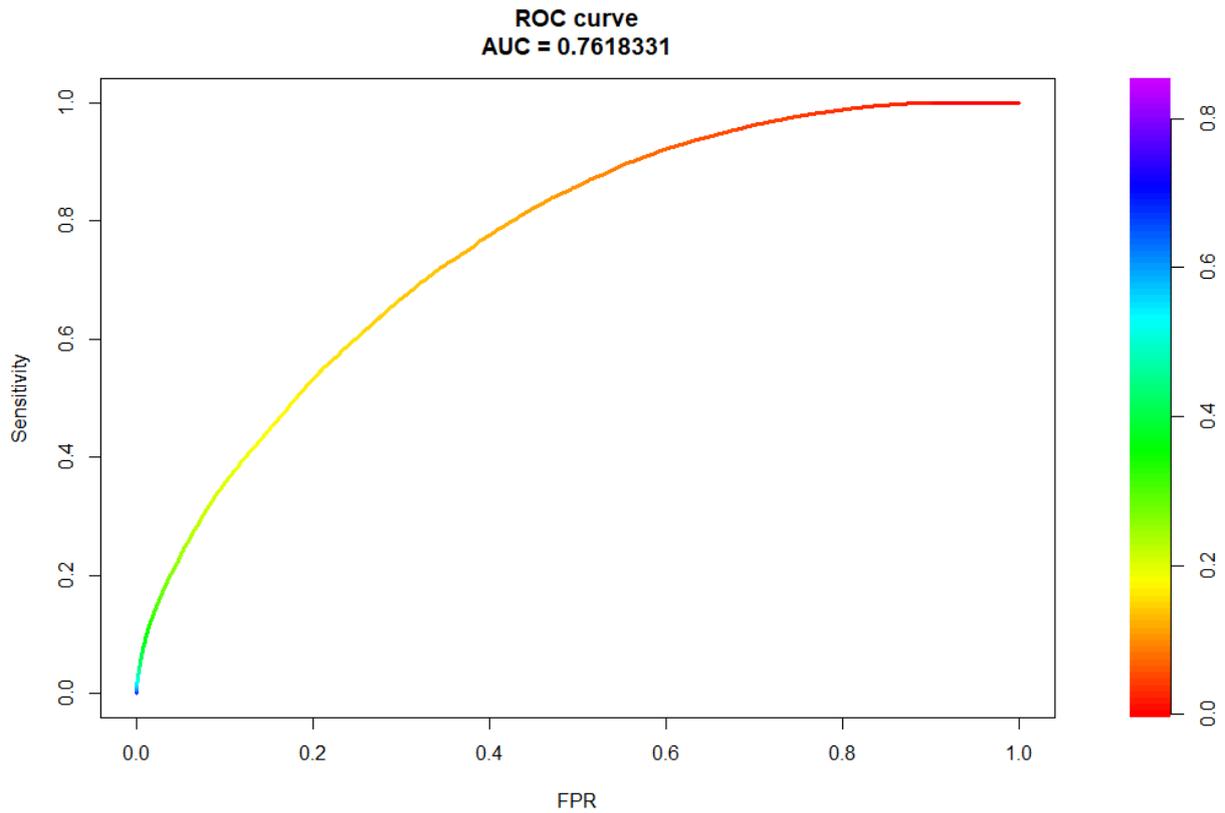


FIGURE 25 : COURBE ROC DU MODELE GRADIENT BOOSTING

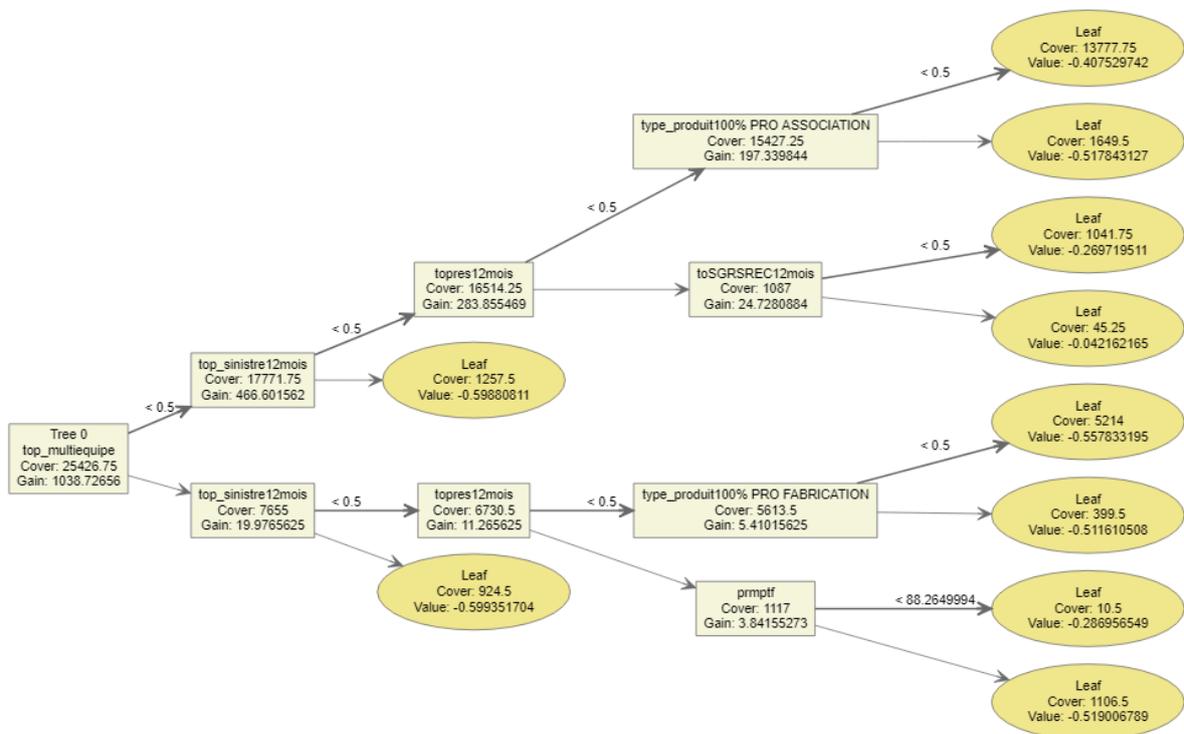


FIGURE 26 : ARBRE REDUIT DU MODELE GRADIENT BOOSTING

L'arbre précédent est une simplification du modèle pour 4 branches. Ce qui permet d'identifier 10 groupes homogènes du portefeuille. L'arbre final est présenté en annexes. Les probabilités de résiliation de chaque groupe sont obtenues en appliquant la fonction logistique à la valeur du **Leaf** obtenue. Plus la valeur du **Leaf** est importante et plus la probabilité de résiliation du groupe est élevée.

Le groupe qui a la probabilité de résiliation la plus grande a les caractéristiques suivantes :

- Client monos équipés,
- Non sinistrés
- Et qui se sont déséquipés dans les douze mois précédant l'échéance de leur contrat MRC.

Le groupe qui résilie le moins est celui des clients sinistrés.

3.1.4. Conclusions sur les 3 modèles :

On remarque que le multi équipement, la sinistralité et le type de produit sont les éléments centraux qui expliquent le plus la résiliation pour les trois modèles utilisés. Le gradient boosting a de meilleures performances en AUC et en erreur de classification que les GLM et les GAM.

3.2. Modèle d'élasticité.

3.2.1. Rappels théoriques

L'élasticité représente la variation du taux de résiliation à une variation de la majoration injectée.

$$Tx(p) = \exp(\beta_0 + \sum_{j=1}^k \beta_j X_j) / [1 + \exp(\beta_0 + \sum_{j=1}^k \beta_j X_j)]$$

En notant β^m les paramètres associés aux variables liées à la majoration (variables notées X^m) et β^s les paramètres des variables n'incluant pas la majoration (variables notées X^s), on peut réécrire $Tx(p)$ comme :

$$Tx(p) = \frac{\exp(m * (\beta^m)X^m + (\beta^s)X^s)}{1 + \exp(m * (\beta^m)X^m + (\beta^s)X^s)}$$

$$Tx(p) = \frac{1}{1 + \exp(-m * (\beta^m)X^m - (\beta^s)X^s)}$$

L'élasticité à la majoration se calcule alors comme :

$$\varepsilon(p) = \frac{\frac{\partial T_x(p)}{T_x(p)}}{\frac{\partial m}{m}}$$

$$\frac{\partial T_x(p)}{\partial m} = \frac{(\beta^m)X^m * \exp(-m * (\beta^m)X^m - (\beta^s)X^s)}{[1 + \exp(-m * (\beta^m)X^m - (\beta^s)X^s)]^2}$$

$$\frac{\partial T_x(p)}{\partial m} = \frac{(\beta^m)X^m * T_x(p)}{[1 + \exp(m * (\beta^m)X^m + (\beta^s)X^s)]}$$

$$\varepsilon(p) = \frac{(\beta^m)X^m * T_x(p)}{[1 + \exp(m * (\beta^m)X^m + (\beta^s)X^s)]} * \frac{m}{T_x(p)}$$

$$\varepsilon(p) = m * (\beta^m)X^m * (1 - T_x(p))$$

On rappelle que m est la majoration (coefficient d'écart au tarif) et que X^m a 1 comme première composante .

Le signe de l'élasticité dépend totalement du signe de $(\beta^m)X^m$.

3.2.2. Application aux données

Pour capter les effets groupés de la majoration croisée aux autres caractéristiques, nous ajoutons les variables croisées dans cette partie. En reprenant le GLM précédant. Nous utilisons le Stepwise dans cette partie également en introduisant initialement tous les croisements de la variable majoration avec les autres caractéristiques. Nous gardons uniquement les croisements significatifs. L'introduction de certaines interactions rend non significative la variable elle-même dans le modèle. Dans ces cas, nous choisissons de garder l'effet croisé plutôt que la variable elle-même si le croisement améliore les performances prédictives du modèle.

Le résumé du modèle sur R est présenté dans la figure suivante.

```
Formula:
top_resil_1_11_mois ~ majo + BCOM + PRMPTF + ANCCclient + anconcontrat +
  MAJO_1 + top_sinistrel2mois + top_multiequipe + top_multigamme +
  top_multisites + topcodeaffaire + topNonProprio + topres12mois +
  topLOLR12mois + toSGRSREC12mois + codereseau + Classement +
  top_mensuel + topres12mois + majo:(ANCCclient + type_produit)

Parametric coefficients:

              Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.534948   0.740857 -10.171 < 2e-16 ***
majo         5.248944   0.690555   7.601 2.94e-14 ***
BCOM20 à 40% 0.123932   0.031354   3.953 7.73e-05 ***
BCOMJusqu'à 20% 0.262776   0.031902   8.237 < 2e-16 ***
BCOMpas de rabais 0.167285   0.072919   2.294 0.021784 *
PRMPTF600-1000€ -0.118057   0.024850  -4.751 2.03e-06 ***
PRMPTF1000-1500€ -0.186930   0.030635  -6.102 1.05e-09 ***
PRMPTF1500-2000€ -0.237945   0.040736  -5.841 5.18e-09 ***
PRMPTF2000€ +  -0.110897   0.032503  -3.412 0.000645 ***
ANCCclient5-100ans 2.542076   1.231050   2.065 0.038926 *
ANCCclient10-20ans 3.818470   1.330147   2.871 0.004095 **
ANCCclient20ans + anconcontrat 5.273863   2.064600   2.554 0.010636 *
MAJO_1plus de 5% -0.021772   0.002698  -8.069 7.09e-16 ***
top_sinistrel2mois 0.029750   0.021446   1.387 0.165374
top_multiequipe -5.883905   0.500185 -11.763 < 2e-16 ***
top_multigamme -1.668335   0.035800 -46.601 < 2e-16 ***
top_multisites 0.276097   0.056466   4.890 1.01e-06 ***
topcodeaffaire -0.231644   0.042599  -5.438 5.39e-08 ***
topNonProprio 0.115805   0.029619   3.910 9.23e-05 ***
topres12mois 0.234914   0.043067   5.455 4.91e-08 ***
topLOLR12mois 0.700540   0.029239  23.959 < 2e-16 ***
toSGRSREC12mois 0.455586   0.023054  19.762 < 2e-16 ***
codereseauCourtier 0.627152   0.045093  13.908 < 2e-16 ***
ClassementStandard -0.182853   0.021677  -8.435 < 2e-16 ***
ClassementSupArierieur 0.043564   0.023574   1.848 0.064612 .
top_mensuel 0.006449   0.023655   0.273 0.785128
majo:ANCCclient5-100ans 0.276878   0.021898  12.644 < 2e-16 ***
majo:ANCCclient10-20ans -2.529346   1.152214  -2.195 0.028149 *
majo:ANCCclient20ans + anconcontrat -3.693063   1.246762  -2.962 0.003055 **
majo:type_produit100% PRO ARTISANS-COMMERCANTS -4.788668   1.939673  -2.469 0.013557 *
majo:type_produit100% PRO ASSOCIATION -0.084622   0.028864  -2.932 0.003371 **
majo:type_produit100% PRO FABRICATION -1.069194   0.047912 -22.316 < 2e-16 ***
majo:type_produit100% PRO SERVICE 0.351502   0.049106   7.158 8.19e-13 ***
majo:type_produit100% PRO SERVICE -0.113323   0.032536  -3.483 0.000496 ***
majo:type_produitANCIENNE GENERATION 0.228572   0.067922   3.365 0.000765 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.0663  Deviance explained = 9.95%
UBRE = -0.35635  Scale est. = 1          n = 127124
```

Les effets croisés significatifs dans notre modèle sont ceux de la majoration avec l'ancienneté client et le type de produit.

Pour chacune de ces variables, les élasticités moyennes estimées sont :

TABEAU 5 : ELASTICITES MOYENNES PAR ANCIENNETE CLIENT

Ancienneté client	Probab Boost	Probab GAM	Elasticités boost	Elasticités GAM	Nombre de contrats	Majo moyenne	Prime moyenne
<5ans	13.6%	13.5%	4.71	4.71	56021	1.07	922
5-10ans	10.2%	10.4%	2.44	2.43	30733	1.07	1028
10-20ans	9.6%	9.5%	1.44	1.44	29670	1.07	1229
20ans +	9.8%	9.8%	0.43	0.43	10700	1.06	1334

TABEAU 6 : ELASTICITES MOYENNES PAR TYPE DE PRODUIT

Type de produit	Probab Boost	Probab GAM	Elasticités boost	Elasticités GAM	Nombre de contrats	Majo moyenne	Prime moyenne	Duration contrat
100% PRO ACPS	10.7%	11.0%	1.84	1.84	30442	1.07	1098	12.1
100% PRO ARTISANS-COMMERCANTS	12.3%	12.4%	3.78	3.78	53033	1.07	1233	4.4
100% PRO ASSOCIATION	6.7%	6.5%	2.36	2.37	9239	1.09	303	6.2
100% PRO FABRICATION	12.7%	12.7%	2.70	2.68	4972	1.09	2766	9.0
100% PRO SERVICE	12.2%	11.8%	3.35	3.37	27232	1.05	591	5.2
ANCIENNE GENERATION	12.4%	13.0%	1.10	1.09	2206	1.06	1156	22.2

Les deux familles d'activités les plus sensibles à la revalorisation sont les Artisans-Commerçants et les Professionnels de Services. L'augmentation de 1% de l'écart au tarif pour ces deux familles augmente les taux de résiliation de 3.78% et 3.37% respectivement pour les Artisans-Commerçants et les Services. Les moins sensibles sont les Pro associations avec un fort taux moyen de revalorisation (9%) et des assiettes de primes plus faibles. Le parallèle peut se faire avec les durations contrats pour chaque produit. Les deux produits les moins sensibles sont les plus anciens avec des assiettes de primes voisines.

TABEAU 7 : ELASTICITES MOYENNES EN FONCTION DE L'INDICATRICE DE SINISTRE 12MOIS

Top 12mois	sinistre	Probab Boost	Probab GAM	Elasticités boost	Elasticités GAM	Nombre de contrats	Majo moyenne
Aucun sinistre		12.6%	12.6%	2.98	2.98	116273	1.07
1 sinistre ou plus		0.1%	0.1%	3.58	3.58	10851	1.07

Le taux de revalorisation moyen des contrats sinistrés est le même que celui des contrats non sinistrés. Mais on remarque que les sinistrés sont plus sensibles à la revalorisation que les non sinistrés bien que le taux de résiliation des contrats sinistrés semble nul.

Nous identifions 27% du portefeuille étudiés comme des profils moins sensibles à la revalorisation. Les caractéristiques associées sont les suivantes :

- Les clients ayant eu au moins 1 sinistre sur les 12 mois précédant l'échéance
- Les détenteurs des produits Pro Associations

- Les contrats ayant plus de 5 ans d'ancienneté avec un rabais initial compris entre 20 et 50%

3.3. Conclusions sur les modèles :

Une première conclusion que nous pouvons tirer sur nos modèles sont que les facteurs les plus impactant sur la sensibilité client à la revalorisation sont :

- La sinistralité qui agit au travers d'un taux de rétention très important et
- Les types de produits et l'ancienneté qui agissent au travers de la majoration sur la résiliation.

Le gradient boosting permet de mieux détecter les résiliations que les GLM et GAM sur nos données. Nous utilisons les probas de rétentions du boosting et les coefficients directeurs des modèles linéaires généralisés pour calculer l'élasticité à la majoration ou au prix des contrats MRC. Les variables les plus importantes qui expliquent la sensibilité des clients à la revalorisation sont très limitées. Même si nos GLM nous permettent d'estimer des effets moyens par produit, l'effet n'est toutefois pas identique au sein des groupes de familles d'activités pour chaque produit. Pour chaque sous-produit et chaque sous-famille d'activité, le risque n'est pas le même et la connaissance du risque par conséquent. Nous décidons ainsi de modéliser séparément chaque produit sous contrainte de la volumétrie. Une autre remarque qu'on pourra faire sur notre modèle est que suivant l'ancienneté, les comportements face à la revalorisation sont différents. Les clients très anciens sont moins sensibles que les affaires nouvelles à la revalorisation. On peut penser aussi à étudier les affaires nouvelles séparément des affaires anciennes vu que le rabais commercial n'aurait à priori pas le même effet suivant que le contrat soit jeune et plus ancien (ces effets croisés sont non significatifs dans nos modèles).

3.4. Modèles par produit.

Dans la suite de ce mémoire nous présentons les modèles simplifiés pour les trois produits les plus sensibles et plus importants en volume :

- Les Pro Artisans Commerçants
- Les Pro ACPS : produits anciens sans affaires nouvelles.
- Les Pro Services

3.4.1. Produit 100% Pro Artisans Commerçants

Ce sous-produit est le plus commercialisé dans la gamme de produits MRC. Il est adressé comme son intitulé l'indique aux Artisans commerçants. Les familles d'activités des clients qu'on peut retrouver sur ce produit sont :

- Bars Restaurants (21% avec une prime moyenne de 1600€)
- Alimentaire (12% avec une prime moyenne de 1300€)
- Bâtiment (11% avec une prime moyenne de 740€)
- Habillement, mode et accessoire (9% avec une prime moyenne de 1100€)

Dans un premier temps, nous introduisons toutes les variables sur la partie Artisans Commerçants de nos données sans interactions. Nous cherchons à savoir si les effets que les variables ont sur la résiliation sont différents du modèle cumulatif des produits présenté précédemment. Les Odds ratios et les variables sélectionnées sont présentés dans le *Tableau 8*. Les variables telles que le chiffre d'affaires et la surface du commerce sont significatives dans ce modèle et l'ancienneté client n'est plus significative dans le modèle.

Ici aussi, on remarque que les niveaux de majorations précédentes ne sont pas significatifs dans le modèle. Les résiliations ne dépendraient pas des niveaux de majorations antérieurs.

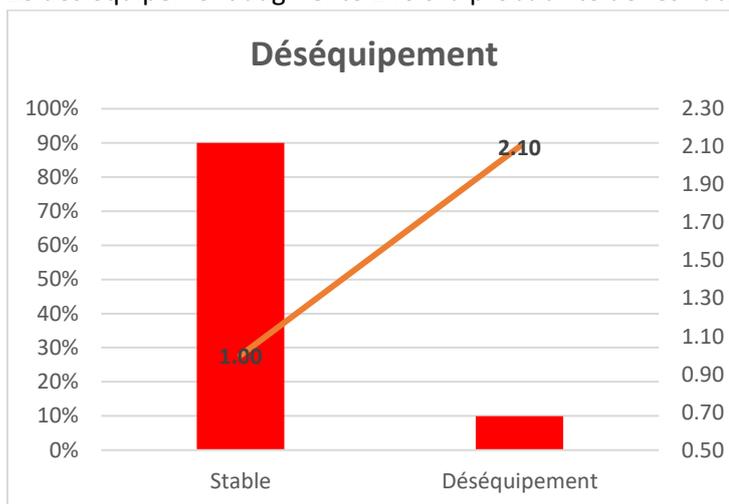
TABLEAU 8 : ODDS RATIOS DU MODELE DE RESILIATION SUR LE PRODUIT PRO ARTISANS COMMERÇANTS

Variables	Estimate	Std. Error	z value	Pr(>)	Significativité	Odds ratio
(Intercept)	-4.39E+00	1.14E+00	-3.846	0.00012	***	0.0
majo	3.10E+00	7.65E-01	4.052	5.07E-05	***	22.2
bcom	8.75E-01	9.70E-02	9.017	< 2e-16	***	2.4
ancontrat	-2.51E-02	3.80E-03	-6.619	3.61E-11	***	1.0
majo_2	-1.30E+00	7.68E-01	-1.691	0.090883	.	0.3
mnt_ca_france	-2.62E-07	5.86E-08	-4.477	7.58E-06	***	1.0
surface_sites	-5.50E-04	1.13E-04	-4.875	1.09E-06	***	1.0
top_sinistre12mois	-5.64E+00	5.75E-01	-9.812	< 2e-16	***	0.0
top_multiequipe	-1.99E+00	6.15E-02	-32.306	< 2e-16	***	0.1
top_multigamme	2.87E-01	9.97E-02	2.879	0.003993	**	1.3
top_multisites	-1.60E-01	5.45E-02	-2.941	0.003272	**	0.9
topcodeaffaire	6.99E-02	3.96E-02	1.767	0.077235	.	1.1
topNonProprio	2.91E-01	5.27E-02	5.525	3.30E-08	***	1.3
topres12mois	7.30E-01	4.36E-02	16.731	< 2e-16	***	2.1

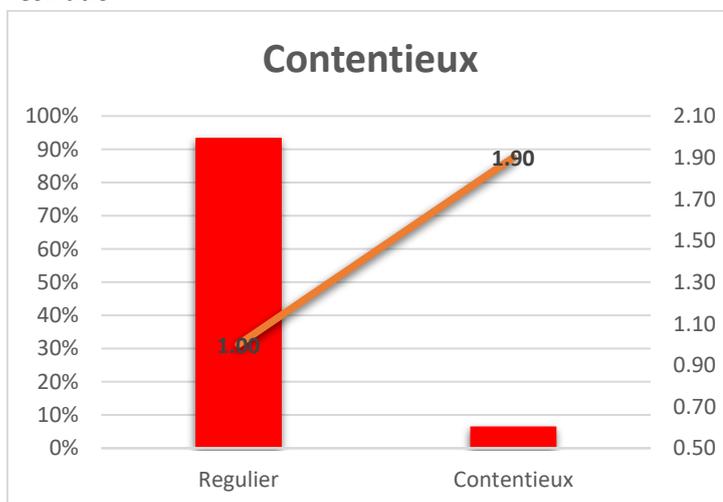
topLOLR12mois	5.57E-01	3.70E-02	15.038	< 2e-16	***	1.7
toSGRSREC12mois	6.19E-01	6.26E-02	9.882	< 2e-16	***	1.9
codereseauCourtier	-2.28E-01	2.97E-02	-7.691	1.46E-14	***	0.8
top_mensuel	2.86E-01	3.32E-02	8.622	< 2e-16	***	1.3
famille_activite_redBÂtiment	1.85E-01	4.93E-02	3.755	0.000173	***	1.2
famille_activite_redBeauté Esthétique	-4.54E-01	5.41E-02	-8.39	< 2e-16	***	0.6
famille_activite_redCommerces et services divers	-3.40E-01	6.74E-02	-5.05	4.43E-07	***	0.7
famille_activite_redFloral et Animaux	-1.58E-01	9.52E-02	-1.661	0.096811	.	0.9
famille_activite_redHabilleme mode et accessoires	-1.65E-01	5.15E-02	-3.197	0.001387	**	0.8
famille_activite_redInformatique Electromenager Téléphonie	-3.24E-01	9.45E-02	-3.431	0.000602	***	0.7
famille_activite_redPapier Impression	-2.84E-01	1.16E-01	-2.438	0.014763	*	0.8
famille_activite_redPlastique	4.35E-01	1.85E-01	2.35	0.018789	*	1.5
famille_activite_redSanté	-4.53E-01	7.77E-02	-5.829	5.57E-09	***	0.6

Toutes choses égales par ailleurs,

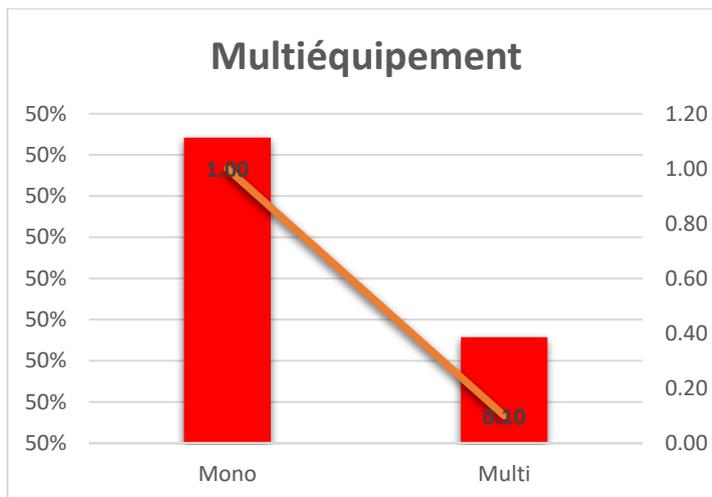
- Le dés équipement augmente 2 fois la probabilité de résiliation



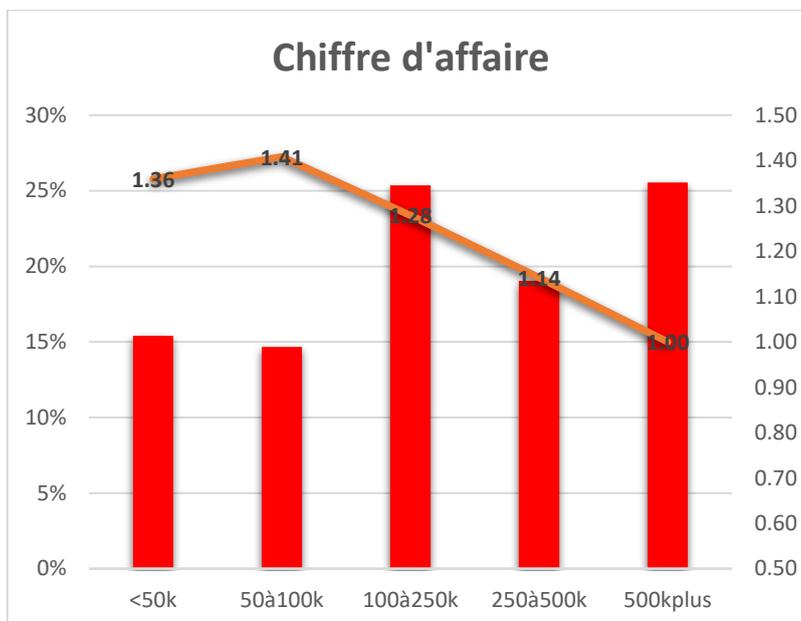
- Les contentieux sur un des contrats du client augmente de deux fois environ la probabilité de résiliation.



- Les clients multi équipés et les clients sinistrés ont de très forts taux de rétentions.



- Le taux de résiliation baisse en fonction du chiffre d'affaires. Le découpage en tranche du chiffre d'affaires permet d'avoir des Odds ratios par tranche de prime.



L'élasticité moyenne du portefeuille est de **3.1**, moins que la valeur estimée dans le modèle de sur tous les produits (**3.8**). Ce qui signifierait qu'une augmentation de l'écart au tarif de 1% augmenterait le taux de résiliation des Produits Artisans Commerçants pris à part de 3.1%.

Nous introduisons ensuite les interactions de la majoration avec les autres variables explicatives pour estimer des élasticités segmentées.

```

Call:
glm(formula = top_resil_1_11_mois ~ majo + top_sinistre12mois +
  ANConcontrat + toSGRSREC12mois + top_sinistre12mois + toSGRSREC12mois +
  mnt_ca_france + surface_sites + top_multiequipe + top_multigamme +
  top_multisites + topcodeaffaire + topNonProprio + topres12mois +
  topLOLR12mois + codereseau + top_mensuel + majo:(PRMPTF +
  top_sinistre12mois + toSGRSREC12mois + famille_activite_red),
  family = binomial, data = AC)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5281  -0.6070  -0.4169  -0.1671   3.7594

Coefficients:
(Intercept)                -6.066e+00  8.621e-01  -7.036  1.98e-12 ***
majo                      4.036e+00  8.015e-01   5.035  4.77e-07 ***
top_sinistre12mois        -3.929e+01  1.397e+01  -2.813  0.004912 **
ANConcontrat5-10ans      -1.972e-01  5.078e-02  -3.882  0.000103 ***
ANConcontrat10ans +
toSGRSREC12mois          8.735e+00  2.658e+00   3.286  0.001018 **
mnt_ca_france              -2.561e-07  6.641e-08  -3.857  0.000115 ***
surface_sites             -5.129e-04  1.228e-04  -4.177  2.95e-05 ***
top_multiequipe          -1.999e+00  6.152e-02  -32.501  < 2e-16 ***
top_multigamme            2.877e-01  9.971e-02   2.886  0.003907 **
top_multisites           -1.613e-01  5.456e-02  -2.957  0.003110 **
topcodeaffaire            8.427e-02  3.847e-02   2.190  0.028510 *
topNonProprio             3.175e-01  5.278e-02   6.015  1.80e-09 ***
topres12mois              7.311e-01  4.365e-02  16.750  < 2e-16 ***
topLOLR12mois            5.698e-01  3.702e-02  15.394  < 2e-16 ***
codereseauCourtier      -1.688e-01  2.904e-02  -5.811  6.21e-09 ***
top_mensuel               3.102e-01  3.333e-02   9.309  < 2e-16 ***
majo:PRMPTF600-1000€     -8.857e-02  3.487e-02  -2.540  0.011080 *
majo:PRMPTF1000-1500€   -1.728e-01  4.431e-02  -3.900  9.62e-05 ***
majo:PRMPTF1500-2000€  -2.080e-01  5.970e-02  -3.485  0.000493 ***
majo:PRMPTF2000€ +
majo:top_sinistre12mois  -1.034e-01  6.151e-02  -1.680  0.092864 .
majo:toSGRSREC12mois     3.108e+01  1.272e+01   2.444  0.014539 *
majo:famille_activite_redBActiment
majo:famille_activite_redBeautÃ© HygiÃ©ne EsthÃ©tique
majo:famille_activite_redCommerces et services divers
majo:famille_activite_redFloral et Animaux
majo:famille_activite_redHabillement mode et accessoire
majo:famille_activite_redInformatique ElectromÃ©nager TA@lÃ©phonie
majo:famille_activite_redPapier Impression
majo:famille_activite_redPlastique
majo:famille_activite_redSantÃ©
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 39740  on 53032  degrees of freedom
Residual deviance: 34583  on 53001  degrees of freedom
AIC: 34647

```

Les nouveaux Odds ratios sont présentés dans le tableau suivant :

TABEAU 9 : ODDS RATIOS DU MODELE DE RESILIATION SUR LE PRODUIT PRO ARTISANS COMMERÇANTS

Variables	Estimate	Std. Error	z value	P-value	Significativité	Odds Ratio
(Intercept)	-6.070	0.86	-7.04	0.00	***	0.0
majo	4.040	0.8	5.04	0.00	***	56.8
top_sinistre12mois	-39.290	13.97	-2.81	0.00	**	0.0
ANConcontrat5-10ans	-0.200	0.05	-3.88	0.00	***	0.8
ANConcontrat10ans +	-0.300	0.06	-5.35	0.00	***	0.7
toSGRSREC12mois	8.730	2.66	3.29	0.00	**	6185.7
mnt_ca_france	0.000	0	-3.86	0.00	***	1.0
surface_sites	0.000	0	-4.18	0.00	***	1.0
top_multiequipe	-2.000	0.06	-32.5	0.00	***	0.1
top_multigamme	0.290	0.1	2.89	0.00	**	1.3
top_multisites	-0.160	0.05	-2.96	0.00	**	0.9
topcodeaffaire	0.080	0.04	2.19	0.03	*	1.1
topNonProprio	0.320	0.05	6.02	0.00	***	1.4
topres12mois	0.730	0.04	16.75	0.00	***	2.1
topLOLR12mois	0.570	0.04	15.39	0.00	***	1.8
codereseauCourtier	-0.170	0.03	-5.81	0.00	***	0.8
top_mensuel	0.310	0.03	9.31	0.00	***	1.4
majo:PRMPTF600-1000€	-0.090	0.03	-2.54	0.01	*	0.9
majo:PRMPTF1000-1500€	-0.170	0.04	-3.9	0.00	***	0.8
majo:PRMPTF1500-2000€	-0.210	0.06	-3.48	0.00	***	0.8

majo:PRMPTF2000€ +	-0.100	0.06	-1.68	0.09	.	0.9
majo:top_sinistre12mois	31.080	12.72	2.44	0.01	*	Inf
majo:toSGRSREC12mois	-7.590	2.49	-3.05	0.00	**	0.0
majo:famille_activite_redBatiment	0.160	0.05	3.21	0.00	**	1.2
majo:famille_activite_redBeauté Esthétique	Hygiène -0.450	0.05	-8.86	0.00	***	0.6
majo:famille_activite_redCommerces et services divers	-0.330	0.06	-5.22	0.00	***	0.7
majo:famille_activite_redFloral et Animaux	-0.210	0.09	-2.34	0.02	*	0.8
majo:famille_activite_redHabillement mode et accessoire	-0.170	0.05	-3.43	0.00	***	0.8
majo:famille_activite_redInformatique Electromenager Téléphonie	-0.310	0.09	-3.43	0.00	***	0.7
majo:famille_activite_redPapier Impression	-0.250	0.11	-2.22	0.03	*	0.8
majo:famille_activite_redPlastique	0.340	0.17	2	0.05	*	1.4
majo:famille_activite_redSanté	-0.440	0.07	-6.11	0.00	***	0.6

Les valeurs calculées des élasticités moyennes par variables importantes sont présentées dans les tableaux suivants :

Familles d'activités	Taux de résiliation	Elasticité moyenne	Nombre de contrats	Majo moyenne	Prime moyenne
Autres	13.1%	7.9	29436	1.07	1455
Bâtiment	13.0%	5.2	5403	1.06	741
Beauté					
Hygiène					
Esthétique	12.0%	6.0	4100	1.08	630
Commerces et services divers	10.1%	7.6	2948	1.06	894
Floral et Animaux	9.6%	7.1	1454	1.08	852
Habillement mode et accessoire	12.9%	6.1	4572	1.06	1084
Informatique					
Electroménager					
Téléphonie	10.7%	6.6	1344	1.06	1423
Papier Impression	9.9%	6.4	961	1.04	1149
Plastique	13.6%	7.2	294	1.08	1486
Santé	8.5%	7.0	2521	1.07	1458

L'élasticité moyenne calculée sur le portefeuille AC est égale à 7 ce qui signifie une augmentation de 7% du taux de résiliation des Artisans commerçants pour 1% d'augmentation de l'écart au tarif. Les familles d'activités les plus sensibles sont les commerces spécialisés dans le Plastique, les commerces et services divers, et ceux catégorisés *Autres* dont l'Alimentaire et les Bars restaurants. Pour les familles *Autres*, l'augmentation de l'écart au tarif de 1% augmente le taux de résiliation de 7.9% (le taux de résiliation devient 14.14%).

Prime	Proba moyenne	Elasticité moyenne	Nombre de contrats	Majo moyenne	Prime moyenne
jusqu'à 600€	14.3%	5.2	18291	1.07	383
600-1000€	13.2%	6.8	12811	1.07	783
1000-1500€	11.9%	7.6	8660	1.07	1225
1500-2000€	10.5%	8.7	4830	1.07	1729
2000€ +	8.7%	10.7	8441	1.07	3482

Le taux de résiliation est décroissant de la prime, mais l'élasticité à la majoration est croissante en fonction de la prime. La majoration moyenne est identique pour toutes les tranches de la prime, mais l'élasticité des grandes primes reste 2 fois plus importante comparée à celle des plus petites primes. Ce qui est logique vu que 7% de revalorisation n'accroît pas de la même manière une assiette de prime de 200€ que celle de 600€.

Ancienneté contrat	Proba moyenne	Elasticité moyenne	Nombre de contrats	Majo moyenne	Prime moyenne
<5ans	13.2%	7.0	42101	1.07	1131
5-10ans	10.2%	7.6	5513	1.07	1511
10ans +	8.1%	8.0	5419	1.07	1739

La sensibilité est décroissante de l'ancienneté du contrat. On retrouve également l'effet de la prime sur l'élasticité précédente vu que les contrats anciens ont des primes plus importantes que les contrats plus jeunes.

Topres12mois	Proba moyenne	Elasticité moyenne	Nombre de contrats	Majo moyenne	Prime moyenne
0	11.9%	7.3	47720	1.07	1224
1	16.8%	6.5	5313	1.07	1311

Les clients s'étant déséquipés dans les 12 mois précédant l'échéance résilient plus mais ne sont pas plus sensible que les clients qui ne se sont pas déséquipés.

Top_sinistre12mois	Probas moyenne	Elasticité moyenne	Nombre de contrats	Majo moyenne	Prime moyenne
0	14.0%	3.4	47040	1.07	1154
1	0.05%	37.2	5993	1.07	1851

On remarque également que les contrats sinistrés ont une très grande sensibilité, jusqu'à 10 fois plus sensible que les contrats non sinistrés même si leur taux de résiliation est quasi nul.

A partir de notre modèle, on peut établir le portrait des contrats les plus sensibles chez les Artisans Commerçants sont Les fortes primes, les contrats plus anciens et les familles d'activité telles que les Bars restaurants, l'Alimentaires, les commerces divers.

Les profils les moins sensibles sont les clients sinistrés.

3.4.2. Produit 100% Pro Services

Les familles d'activités des clients qu'on peut retrouver sur ce produit sont :

- Bureaux Seuls (60% avec une prime moyenne de 550€)
- Activités médicales et Paramédicales (12% avec une prime moyenne de 540€)
- Services aux particuliers (6% avec une prime moyenne de 650€).

L'étude de ce portefeuille permet d'intégrer plus de variables telles que le nombre d'employés qui est significatif dans le modèle (Voir *Tableau 11*). La majoration est significative dans le modèle et on remarque que la valeur de l'élasticité moyenne estimée est de **4.83** ce qui signifie que l'augmentation de 1% de l'écart au tarif augmenterait de 4.83% le taux de résiliation du portefeuille Pro Services.

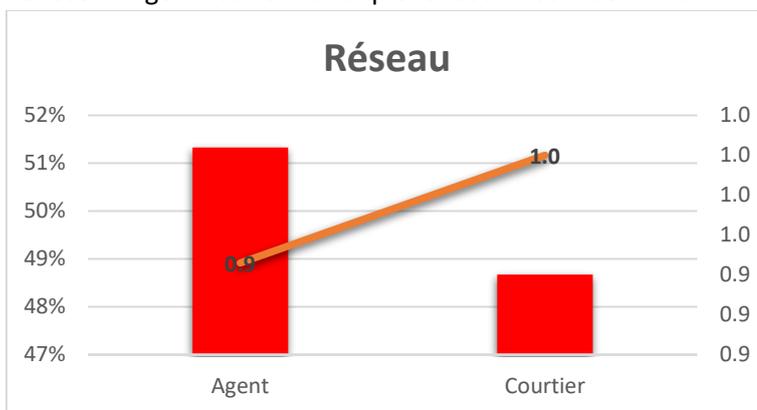
TABLEAU 10 : ODDS RATIOS DU MODELE DE RESILIATION SUR LE PRODUIT PRO SERVICES

Variables	Estimate	Std.Err.	z	P> z
const	-7.21	0.12	-19.4	0.00
majo	4.83	0.01	4.2	0.00
codereseau_Agent	-0.06	0.03	-1.7	0.08
bcom_50% +	-0.65	0.19	-3.4	0.00
bcom_36-50%	-0.16	0.06	-2.5	0.01
bcom_10-20%	0.04	0.04	0.9	0.37
bcom_e_0-10%	0.14	0.05	2.7	0.01
bcom_pas de rabais	0.18	0.05	3.6	0.00
top_sinistre12mois	-2.71	0.19	-13.9	0.00
topmultiequipe	-0.11	0.04	-3.1	0.00
topmultigamme	-0.17	0.05	-3.7	0.00
topcodeaffaire	0.28	0.05	5.6	0.00
topsous12mois	-0.21	0.11	-1.9	0.06
topres12mois	0.42	0.05	8.5	0.00
top_interaction	-0.20	0.13	-1.5	0.12
topLO12mois	0.37	0.03	10.6	0.00
toSGRSREC_12mois	0.43	0.05	8.5	0.00
top_multisites	-0.13	0.05	-2.5	0.01
PRMPTF_250-500	-0.15	0.04	-3.9	0.00
PRMPTF_500-1000	-0.36	0.05	-7.4	0.00
PRMPTF_plus1000	-0.46	0.07	-6.5	0.00
ca_100k-250k	-0.16	0.04	-3.7	0.00
ca_250k500k	-0.12	0.05	-2.5	0.01
ca_50k100k	-0.01	0.04	-0.1	0.90
ca_500k+	-0.05	0.05	-0.9	0.35
surface_100-300m ²	-0.11	0.04	-2.7	0.01
surface_20-50m ²	0.08	0.04	1.8	0.07
surface_mois de 20m ²	0.26	0.05	4.9	0.00
surface_300m ² +	-0.24	0.08	-2.9	0.00

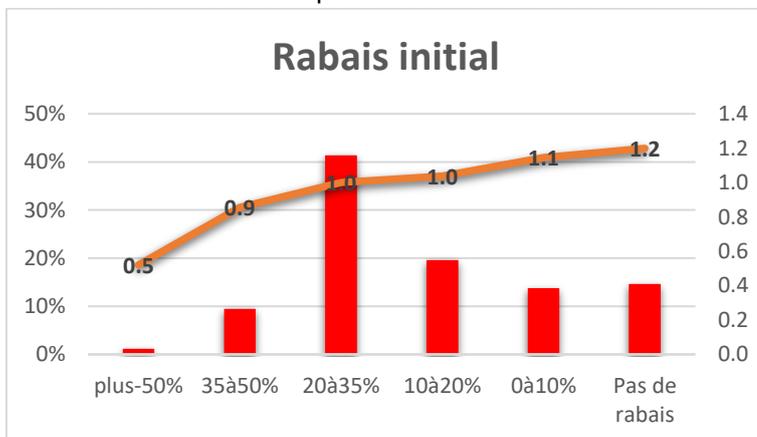
effectif_2-5 salaries	0.10	0.04	2.7	0.01
effectif_5salaries +	0.04	0.05	0.8	0.41
zone_Paris	0.10	0.03	2.9	0.00
topMensuel	0.18	0.03	5.1	0.00
famille_activite_Bonne	-0.53	0.18	-2.9	0.00
famille_activite_Moyenne	-0.27	0.06	-5.0	0.00
Ancclient_1-2ans	0.19	0.05	3.8	0.00
Ancclient_2-5ans	0.21	0.04	6.0	0.00
Ancclient_ moins 1an	0.37	0.08	4.5	0.00

Toutes choses égales par ailleurs :

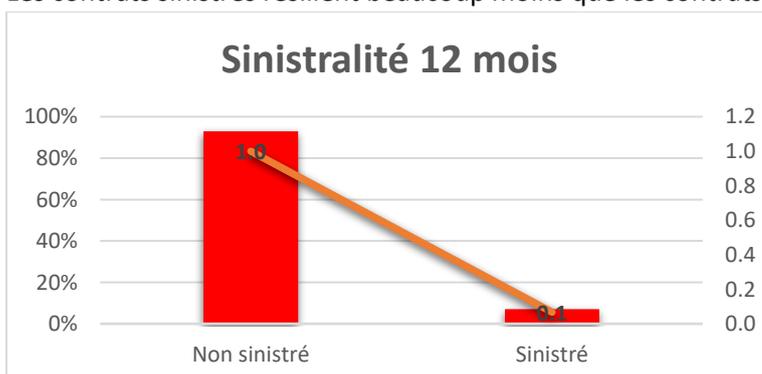
- Le réseau Agent résilie moins que le réseau Courtier mais la différence n'est pas significative.



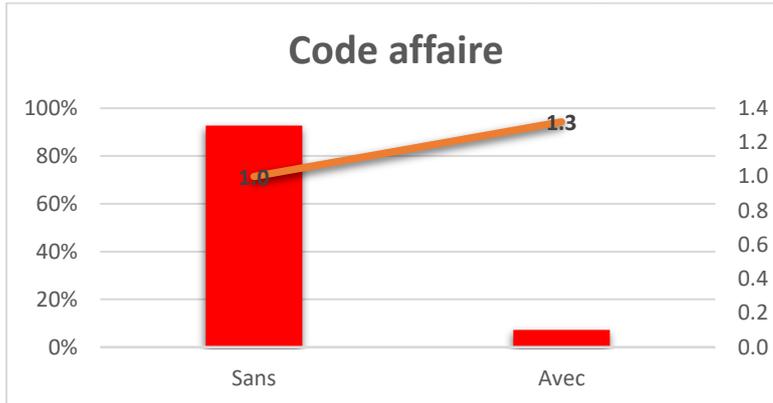
- Plus le rabais initial est important et moins les clients résilient



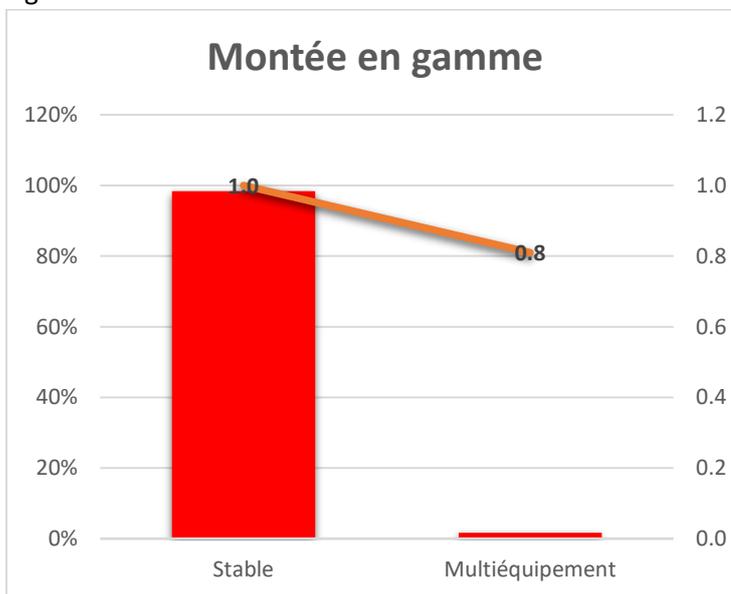
- Les contrats sinistrés résilient beaucoup moins que les contrats non sinistrés



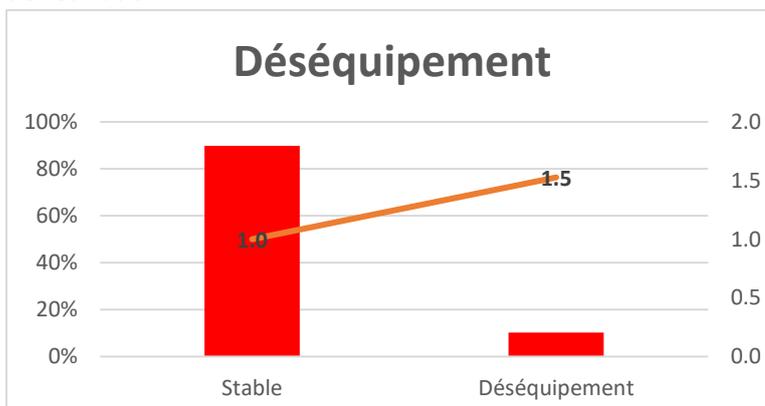
- Les contrats avec des code affaires résilient plus que les contrats sans code affaire.



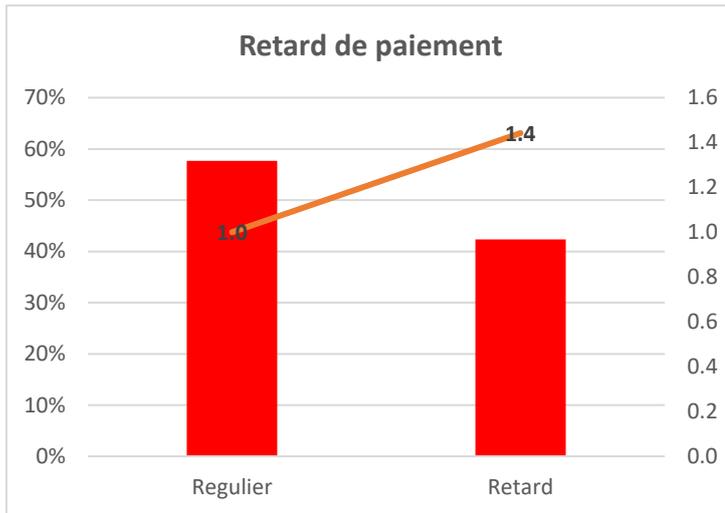
- Le fait de monter en gamme sur les mois précédant l'échéance réduirait de 20% la probabilité de résiliation mais cela concerne uniquement 2% du portefeuille Pro Service et l'effet n'est pas significatif.



- Le fait de se déséquiper sur les mois précédant l'échéance augmenterait de 50% la probabilité de résiliation



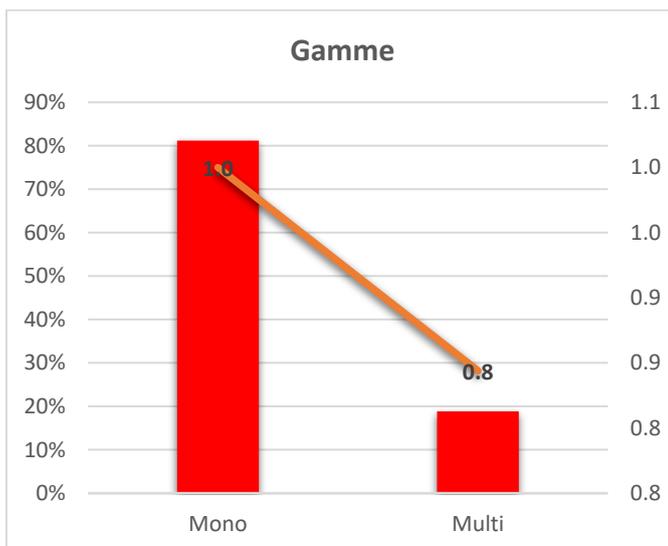
- Les retards de paiement et les contentieux augmenterait de 44 % à 53% la probabilité de résiliation sur le portefeuille Pro Services



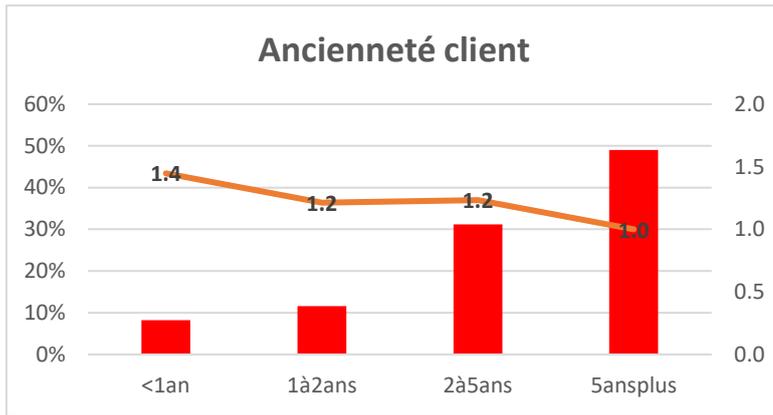
- Les clients multi équipés résilient 10% moins que les clients monos équipés en particulier, les clients qui détiennent des gammes différentes de type IARD, VIE, Epargne-Retraite résilient encore moins que les clients monos équipés ou multi équipés d'une même gamme de produits.



-

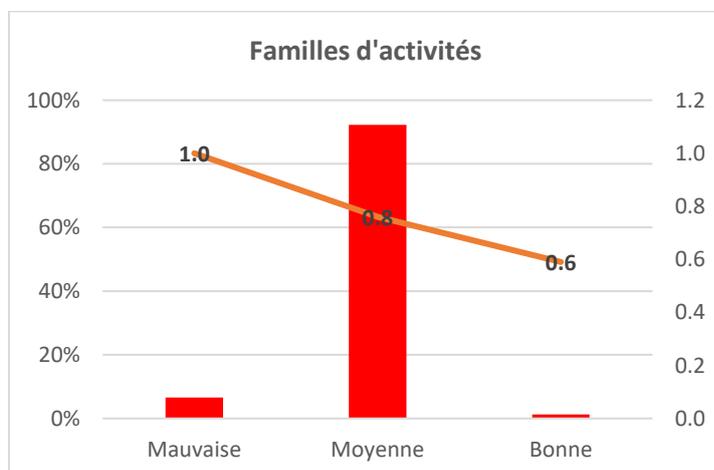


- Les affaires nouvelles résilient plus que les affaires anciennes.



L'introduction des familles d'activités dans un premier temps ne s'est pas révélée significative. Nous avons donc catégorisé les familles d'activité en 3 catégories : bonne, moyenne et mauvaise suivant l'historique de résiliation dans chaque famille d'activité.

Classe	famille_activite	Taux de résil	Effectif moyen
Bonne	Soins aux animaux	5.4%	3.0%
	Sport	6.3%	0.2%
Mauvaise	Activités de Bien-être	15.8%	1.5%
	Bureaux Etudes - Contrôles	14.6%	0.1%
	Conseil en entreprise	12.9%	4.0%
	Événementiels, Spectacles	13.4%	1.0%
Moyenne	Activités Médicales - Paramédicales	7.6%	12.3%
	Bureaux seuls	12.0%	60.5%
	Comptabilité - Finance	7.3%	2.5%
	Expertise	10.6%	0.2%
	Juridique	8.4%	3.2%
	Métiers de l'audiovisuel et de l'écriture	12.0%	1.3%
	SSII	10.4%	3.1%
	Services au particulier et/ou professionnel et entreprise	11.6%	5.0%
	Services à la personne	10.7%	2.1%



Toutes choses égales par ailleurs, les familles d'activités classées moyennes et bonnes résilient 20% et 40% moins que les familles d'activités catégorisées mauvaises.

L'introduction des interactions permet de calculer les élasticité par variables.

- Les clients du portefeuille Pro Service qui ont eu au moins un sinistre dans les 12 mois précédant l'échéance de leur contrat sont presque insensible à la revalorisation. Dans les modèles précédents (sur le Portefeuille complet et sur les Artisans Commerçants), les clients sinistrés avaient une forte sensibilité à la revalorisation, mais leur taux de résiliation étant presque nul, l'accroissement du tarif ne modifiait pas la taille de leur portefeuille. Mais pour les produits Pro Services, non seulement le taux de résiliation des clients sinistrés est faible mais aussi leur élasticité calculée est très faible également.

top_sinistre12mois	Elasticité moyenne	Proba de résiliation	Taux observé	Majoration moyenne	Prime moyenne	Nb Contrats
Non sinistré	4.04	11.9%	12.0%	1.05	586	93%
Sinistré	0.06	0.7%	0.8%	1.05	1043	7%

- Plus la prime est importante et moins les clients résilient, de même l'élasticité décroît en fonction de la prime. Ce qui pourrait en partie expliquer la faible élasticité des clients sinistrés vu qu'ils ont une prime moyenne plus grande que celle des clients non sinistrés.

Prime	Elasticité moyenne	Proba de résiliation	Taux observé	Majoration moyenne	Prime moyenne	Nb Contrats
Moins de 250€	10.41	15.1%	15.3%	1.05	172	26%
250 – 500 €	3.71	11.8%	11.6%	1.05	362	34%
500 – 1000€	0.16	8.7%	8.9%	1.05	697	24%
1000€ +	0.02	6.6%	6.7%	1.05	1857	15%

- Les *Bureaux Seuls* représentent 60% des familles d'activités du portefeuille Pro Services. Comparés aux autres familles d'activités, les Bureaux Seuls résilient plus. Mais leur élasticité est plus faible que celle des autres familles d'activités.

La classification des familles d'activités suivant leur taux observé permet de voir que les familles catégorisées mauvaises qui ont un taux de résiliation observé au-delà de la moyenne ont une très forte sensibilité à la revalorisation même si leur pourcentage dans le portefeuille est faible.

Groupe famille	Elasticité moyenne	Proba de résiliation	Taux observé	Majoration moyenne	Prime moyenne	Nb Contrats
Autres	5.1	10.6%	9.7%	4.51	683	37%
Bureaux seuls	3.0	11.5%	12.0%	5.39	579	63%

Famille d'activités	Elasticité moyenne	Proba de résiliation	Taux observé	Majoration moyenne	Prime moyenne	Nb Contrats
Bonne	-1.8	5.5%	5.6%	1.05	1531	1%
Mauvaise	11.2	13.6%	13.6%	1.04	630	7%
Moyenne	3.3	11.0%	11.1%	1.05	605	92%

En somme les Produits Pro Services sont des produits très sensibles par rapport aux autres produits bien que leur prime moyenne soit plus faible que celles des Artisans Commerçants.

3.4.3. Produit 100% Pro ACPS

Sur ce produit, les familles d'activités sont très mal renseignées par les intermédiaires (Agent et Courtier). Toutefois, les secteurs qui caractérisent ce portefeuille sont :

- Les Bureaux Seuls
- Bars Restaurants
- Bâtiment
- Alimentaire

De même, les informations sur les variables de risques sont absentes ce qui limite les variables utilisables dans la modélisation de la résiliation de ce produit.

Un premier GLM sans introductions des interactions montre que sur ce produit, très peu de variables sont significatives.

TABLEAU 11 : ODDS RATIOS DU MODELE DE RESILIATION SUR LE PRODUIT ACPS

Variables	Estimate	Std. Error	z value	Pr(>)	Significativité
(Intercept)	-5.11E+00	1.07E+00	-4.769	1.85E-06	***
majo	3.07E+00	9.99E-01	3.075	0.002106	**
prmpft	4.41E-05	1.57E-05	2.815	0.004879	**
ancclient2	3.11E-02	4.67E-03	6.676	2.45E-11	***
ancontrat	-3.95E-02	5.47E-03	-7.213	5.49E-13	***
top_sinistre12mois	-1.65E+01	1.27E+02	-0.129	0.896974	
top_multiequipe	-1.69E+00	7.17E-02	-23.586	< 2e-16	***
top_multigamme	3.93E-01	1.06E-01	3.701	0.000215	***
topres12mois	6.23E-01	6.26E-02	9.956	< 2e-16	***
topLOLR12mois	3.80E-01	4.75E-02	8.014	1.11E-15	***
toSGRSREC12mois	6.12E-01	1.30E-01	4.724	2.31E-06	***
codereseauCourtier	-2.29E-01	4.21E-02	-5.424	5.82E-08	***
top_mensuel	2.26E-01	4.41E-02	5.125	2.97E-07	***

La variable famille d'activité n'est pas significative dans ce modèle ce qui signifierait une absence de disparité de résiliation entre les familles d'activités composant le portefeuilles ACPS, ce qui n'est pas le cas en termes de risque comme on a pu le remarquer dans l'étude des produits Pro Artisans-Commerçants et Pro Service. Le produits ACPS n'est autre que les ces deux derniers en version ancienne.

On remarque que la prime a un effet positif sur la résiliation et que l'indicatrice de sinistre n'est pas significative dans ce modèle comparé aux modèles antérieurs.

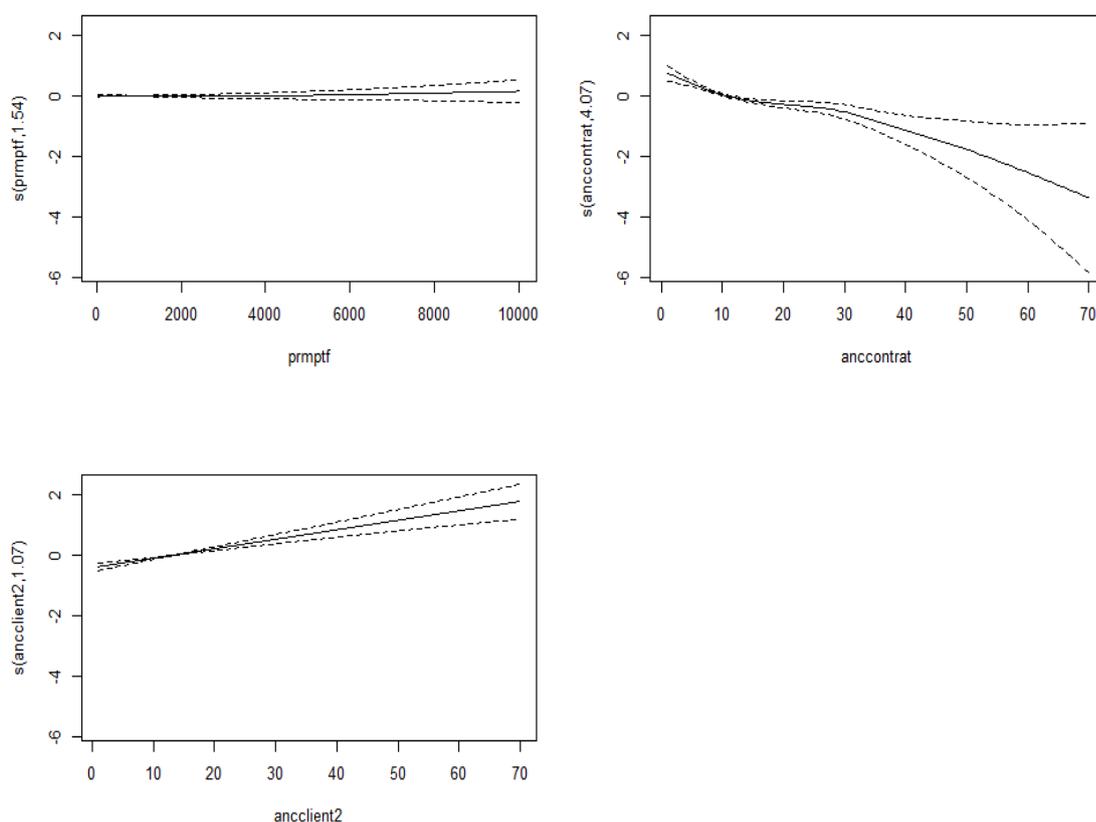


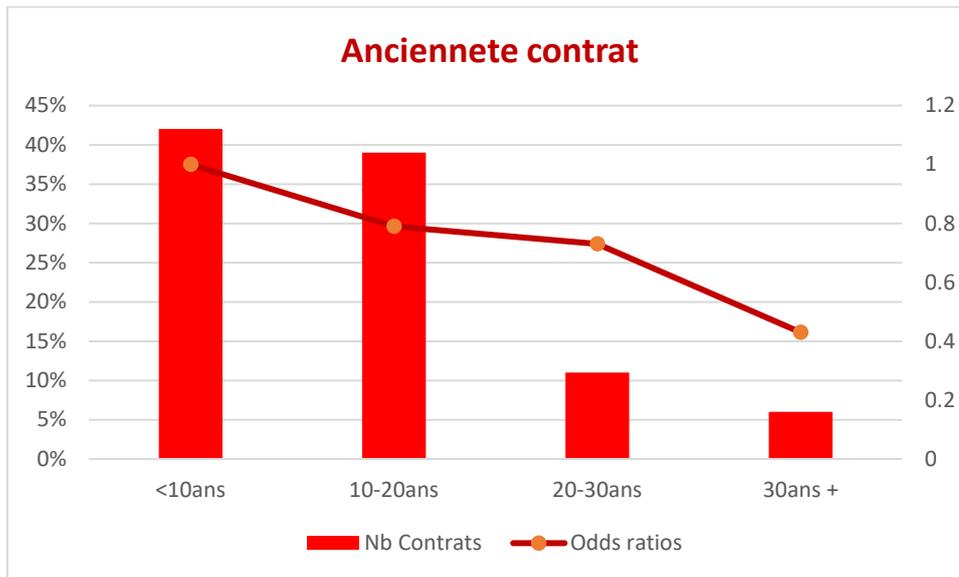
FIGURE 27 : TAUX DE RESILIAN EN FONCTION DES VARIABLES CONTINUES (MODELE ACP)

L'utilisation des modèles GAM suggère un découpage de l'ancienneté contrat en 3 tranches. L'intégration des tranches d'ancienneté contrat donne les odds ratios présentés dans le tableau suivant :

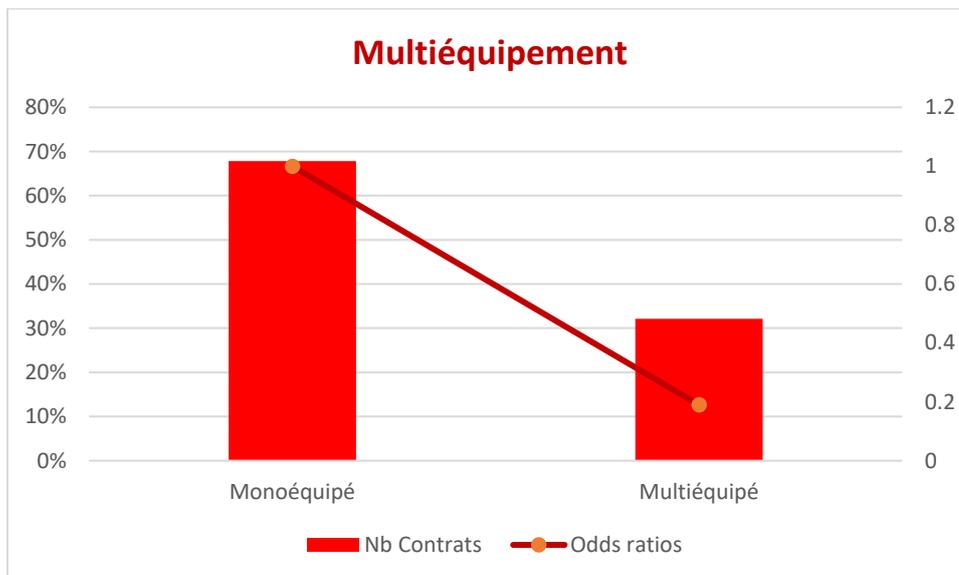
Variables	Odds ratio	2.50%	97.50%	Pr(>z)	Significativité
(Intercept)	0	0	0.04	3.49E-07	***
majo	20.81	2.97	145.76	0.002239	**
ancclient2	1.02	1.01	1.03	3.04E-07	***
ANCcontrat10-20ans	0.79	0.72	0.87	6.80E-07	***
ANCcontrat20-30ans	0.73	0.6	0.88	0.001437	**
ANCcontrat30ans +	0.43	0.3	0.64	2.36E-05	***
toSGRSREC12mois	1.86	1.45	2.39	1.28E-06	***
top_multiequipe	0.19	0.17	0.22	< 2e-16	***
top_multigamme	1.5	1.22	1.84	0.000136	***
topres12mois	1.99	1.76	2.25	< 2e-16	***
topLOLR12mois	1.43	1.31	1.57	1.85E-14	***
top_mensuel	1.29	1.19	1.41	3.10E-09	***

Toutes choses égales par ailleurs :

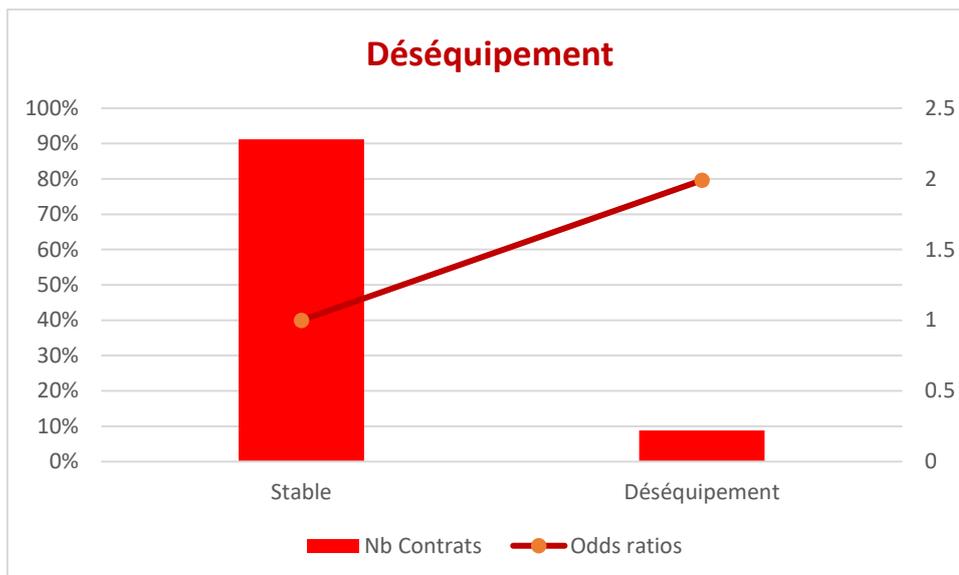
- La résiliation décroît en fonction de l'ancienneté contrat



- Les clients multi équipés résilient 80% moins que les clients monos équipés détenteurs de produit ACPS



- Le dés équipement augmente 2 fois la probabilité de résiliation des Produits APCS



En introduisant les variables croisées, aucune interaction avec la majoration ne s'est pas révélée significative sur les Produits ACPS. Bien que les familles d'activités soient très différentes en termes de profils de risques, nous n'avons pas remarqué d'effets croisés significatifs. La variation de l'élasticité dépend de la probabilité de rétention estimée et du niveau de majoration moyenne.

TABLEAU 12 : ELASTICITES MOYENNE PAR FAMILLE D'ACTIVITE DU PRODUIT ACPS

Familles d'activité	Proba résiliation	de Elasticité moyenne	Majoration moyenne	Prime moyenne
Juridique	3%	3.08	1.04	391.93
Enseignement	4%	3.10	1.06	3580.2667
Conseil en entreprise	5%	3.04	1.05	1229.4
Bricolage	6%	3.05	1.04	2636.58
Bâtiment	9%	2.99	1.05	1105.8833
Commerces et services divers	9%	2.80	1.05	1459.506
Papier Impression	9%	2.71	1.04	959.894
Aménagement décoration	9%	2.89	1.06	1037.1457
Plastique	10%	2.96	1.08	5161.68
Alimentaire	10%	2.94	1.08	2049.1116
Autres	11%	2.85	1.07	1093.5666
Activités Médicales - Paramédicales	11%	2.79	1.04	444.906
Bureaux seuls	11%	2.84	1.06	945.6125
Beauté Hygiène Esthétique	11%	2.88	1.07	695.8088
Autos-cycles bateaux aéronautique	12%	2.90	1.07	2108.458
Santé	12%	2.81	1.07	1819.66
Floral et Animaux	12%	2.85	1.08	850.5
Bois	12%	2.86	1.05	114.44
Loisirs musique	12%	2.79	1.07	3902.73
Bar restaurant	12%	2.83	1.08	2349.2236
Habillement mode et accessoire	12%	2.81	1.06	1186.3267
Informatique Electroménager Téléphonie	14%	2.79	1.06	806.2978
Hôtel hébergement	17%	2.78	1.07	3384.8675
Electronique Electricité	19%	2.60	1.06	1108.84

Applications possibles de l'étude

L'identification des profils clients les plus ou moins sensibles permet de construire des scénarios de majoration pour chaque client. Pour chaque taux de revalorisation injecté, nous calculons les probabilités de résiliation associées aux divers scénarios d'augmentation du taux de revalorisation.

Nous rappelons que le taux de revalorisation injecté provient des Bureaux d'études techniques chargés de la tarification et que le taux de revalorisation transmis au client est celui qui prend en compte la sensibilité du client, sa valeur client, et le choix final de l'intermédiaire de défendre son client (c'est d'ailleurs une limite de nos modèles de ne pas pouvoir intégrer les défenses antérieures des intermédiaires dans nos modèles). Ainsi, l'étude permet de mettre en place des scores de fragilité à l'attention des intermédiaires afin de les sensibiliser sur le risque de départ des clients.

Dans nos modèles, la variable ancienneté étant centrale. Le croisement des probabilités par groupes et par ancienneté permet d'avoir une loi de chute pour chaque profil de client et chaque scénario d'évolution de la revalorisation. Cette loi de chute permettra de mettre à jour la valeur client. Les valeurs clients mises à jour permettront suivant les nouvelles valeurs obtenues de choisir les clients à protéger contre les forts taux de revalorisations.

Conclusions

Dans ce mémoire, nous avons étudié les résiliations et l'élasticité prix du produit MRC de GENERALI. Nous avons organisé notre étude en deux étapes.

Une première qui consiste à déterminer les probabilités de rétentions des contrats MRC, nous utilisons 3 modèles qui sont les modèles linéaires généralisés à cause de leur facilité d'implémentation, les modèles additifs généralisés pour améliorer le découpage des variables continues et estimer des effets groupés plutôt que linéaires et les gradient boosting pour améliorer la détection des résiliations. Le gradient boosting permet d'avoir de meilleures performances prédictives que les GLM et les GAM sur le modèle des résiliations.

Une deuxième étape dans laquelle nous utilisons les meilleurs modèles paramétriques et leurs coefficients des variables croisées avec la majoration pour estimer les élasticités groupées.

On a remarqué dans l'étude que les clients multi équipés qui représentent plus de 30% du portefeuille résilient moins et que plus ils sont anciens et moins ils sont sensibles aux revalorisations. De même, les contrats sinistrés résilient très peu mais avec une sensibilité très forte à la revalorisation. Et le taux quasiment nul des résiliations des contrats sinistrés fait qu'au final, le groupe des sinistrés n'est pas très sensible à la majoration. Les clients qui sont déjà entrés en contentieux avec l'entreprise sur les 12 mois précédant l'échéance représentent des profils très fragiles sur lesquels il faut mener des actions pour réduire leur départ bien que leur effectif soit très peu élevé. Pour les produits, les Pro Associations qui sont très peu diversifiés sont les profils qui résilient le moins. L'étude du portefeuille au complet montre que les deux Produits les plus récents que sont les Pro Artisans Commerçants et les Pro Services sont les plus sensibles et sont ceux qui résilient le plus. Afin d'étudier les disparités au sein de chaque famille de risques, nous avons modéliser les Produits de façon séparée et introduit les familles d'activités pour chaque Produit afin de segmenter d'avantage les élasticités calculées. Cette séparation a permis de comparer les familles d'activités entre elles et de déterminer les familles de risques les plus sensibles à la revalorisation pour chaque produit. Il serait intéressant d'étudier séparément les affaires nouvelles aussi, les clients multi équipés également afin d'améliorer la segmentation des groupes.

Aussi, dans nos modèles, la majoration n'est pas la variable la plus importante ce qui signifie que l'effet de la revalorisation n'explique que très partiellement les fuites de contrats. Les valeurs d'élasticités faibles montrent que l'impact de la revalorisation sur les résiliations reste faible et qu'il faudrait explorer d'autres pistes telles que le changement de risques (avenants).

Mais bien que les valeurs d'élasticités mesurées par nos modèles restent faibles, elles permettent d'identifier les profils sensibles à la revalorisation. L'identification des divers profils permet de mettre en place des scénarios de majorations en partant des majorations pré-termes des bureaux d'études techniques et de sensibiliser les Agents et Courtiers sur le risque de résiliation des clients. Aussi, les modèles de résiliations permettent de déterminer les nouvelles lois de chutes pour chaque scénario de revalorisation et de mettre à jour la valeur client qui est centrale dans le choix de la protection du client contre les forts taux de revalorisation.

Bibliographie

1. **Mémoire de Yacine ABOUTAYBI** : Multi-risques commerce : Refonte de la valeur client multi-risques commerce - Etude de la résiliation – 2021
2. **Mémoire de Massile MOURAH** : Multi-risques Habitation : Modélisation de l'élasticité aux prix en assurance MRH – 2017
3. Wood, S. N. (2006). **Generalized additive models**: an introduction with R. Chapman and Hall/CRC.
4. Brownlee, J. (2020). *Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning*. Machine Learning Mastery.
5. Brownlee, J. (2016). *XGBoost With python: Gradient boosted trees with XGBoost and scikit-learn*. Machine Learning Mastery.
6. **Mémoire de Christophe Dutang**: Assurance non vie : Regression models of price elasticity in non-life insurance – 2011
7. **Mémoire de Mehdi BOUEDDINE** : Assurance Automobile : Analyse de l'impact d'une variation du tarif sur le comportement des assurés lors de l'acte de souscription et de résiliation – 2013
8. **Mémoire de SOURISSEAU J** : Modélisation du taux de transformation et de l'élasticité prix en affaire nouvelle pour l'assurance automobile – 2012
9. **Mémoire de Oussama CHIH** : Sensibilité du taux de résiliation au prix en assurance MRH occupant et simulation du portefeuille – 2011
10. **GEIGLE T., VIGNAL G. [2016]**, "Modélisation de la valeur du client entreprise et utilisations opérationnelles", Mémoire, 2016.
11. **GUILLOT A. [2015]**, "Apprentissage statistique en tarification non-vie : quel avantage opérationnel ?", mémoire d'actuaire.
12. **Corrado, Gianluca, and Andrea Passerini**. "Scikit Learn: Machine Learning in Python."
13. **Vianney Perchet**. « Cours de Fondements mathématiques du Machine Learning ».In : (2020-2021).
14. **Zou, Q., Xie, S., Lin, Z., Wu, M., & Ju, Y.** (2016). Finding the best classification threshold in imbalanced classification. *Big Data Research*, 5, 2-8.

15. **Saito, T., & Rehmsmeier, M.** (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3), e0118432.
16. **BALLINA R.** [2014], "Méthodes d'apprentissage appliquées à la tarification non-vie", mémoire.
17. **DENIL M. et al.** [2014], "Narrowing the Gap : Random Forests In Theory and In Practice", article
18. **Charpentier and C. Dutang.** L'actuariat avec R. 2012.

Annexes

1. Performances des modèles

- Modèle 100% Pro Artisans Commerçants

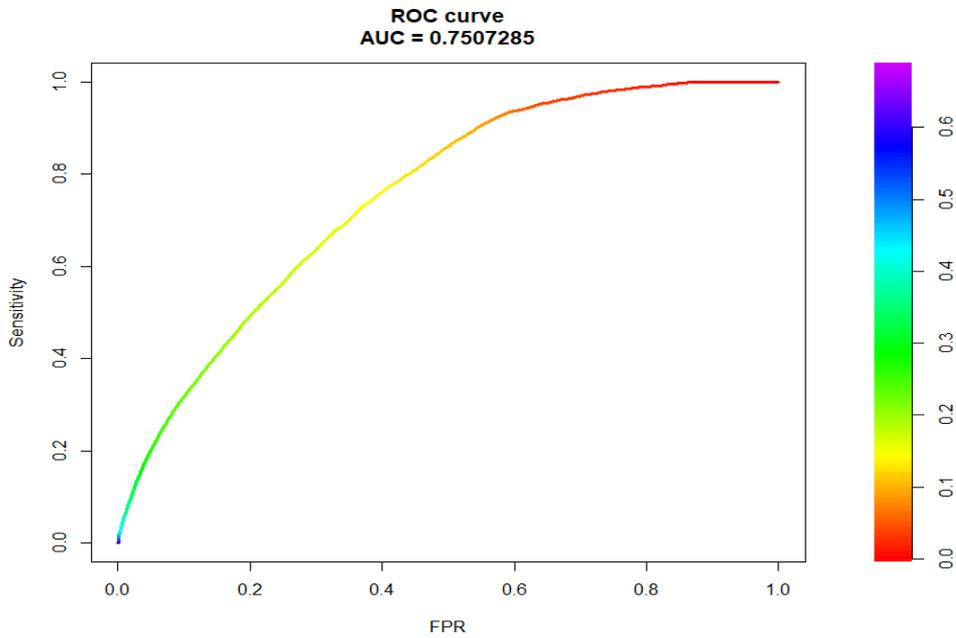


FIGURE 28 : COURBE ROC SUR DONNEES D'ENTRAINEMENT – MODELE PRO ARTISANS COMMERÇANTS

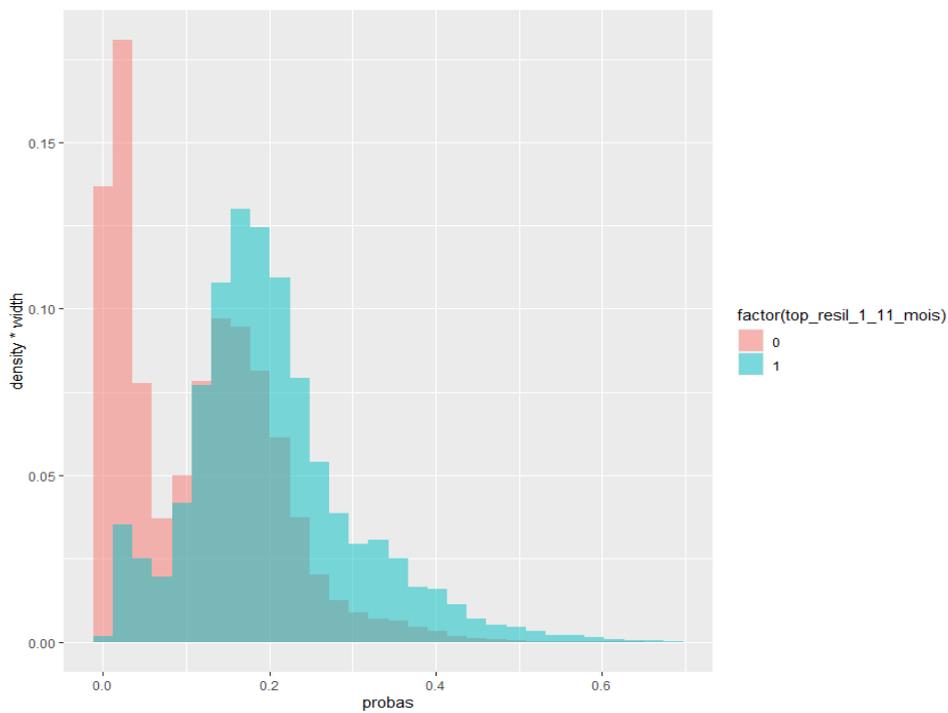


FIGURE 29 : COMPARAISON DES PROBABILITES DE RESILIATION POUR LES CONTRATS RESILIES ET RETENUS - MODELE PRO ARTISANS COMMERÇANTS

- Modèle 100% Pro Service

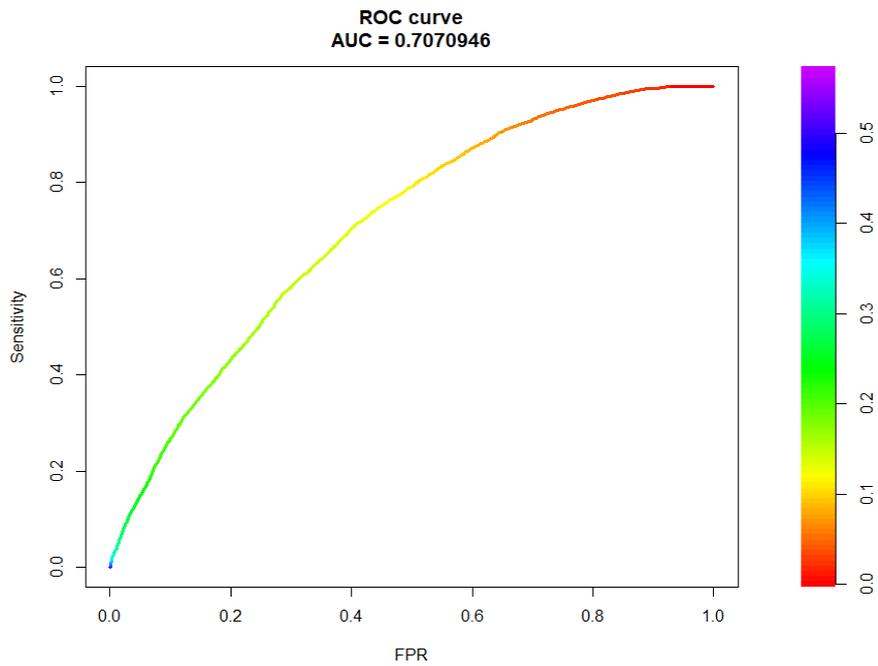


FIGURE 30 : COURBE ROC SUR DONNEES D'ENTRAINEMENT - MODELE PRO SERVICES

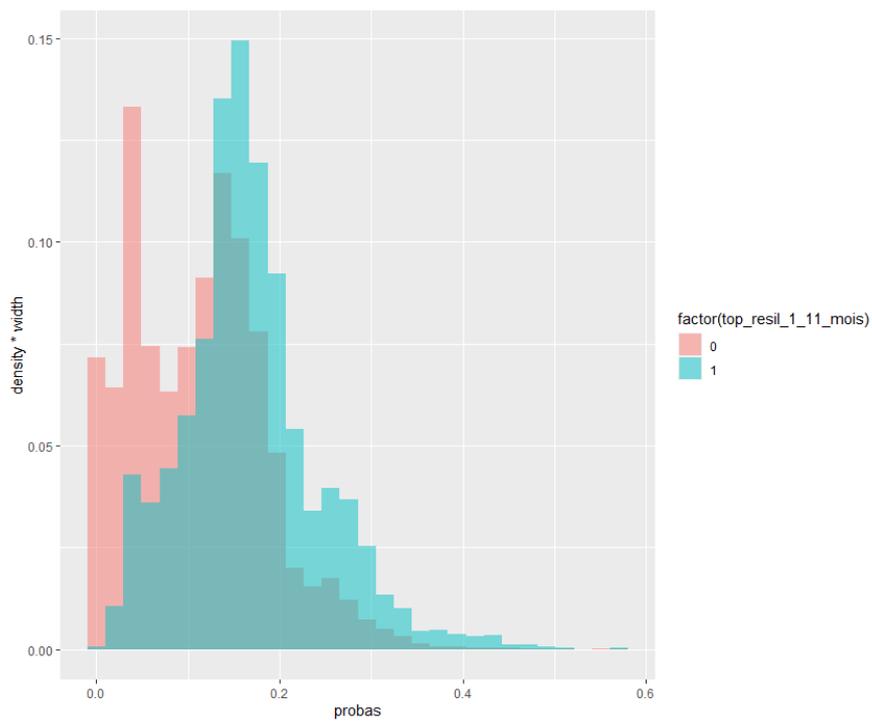


FIGURE 31 : COMPARAISON DES PROBABILITES DE RESILIATION POUR LES CONTRATS RESILIES ET RETENUS - MODELE PRO SERVICES

- Modèle 100% Pro ACPS

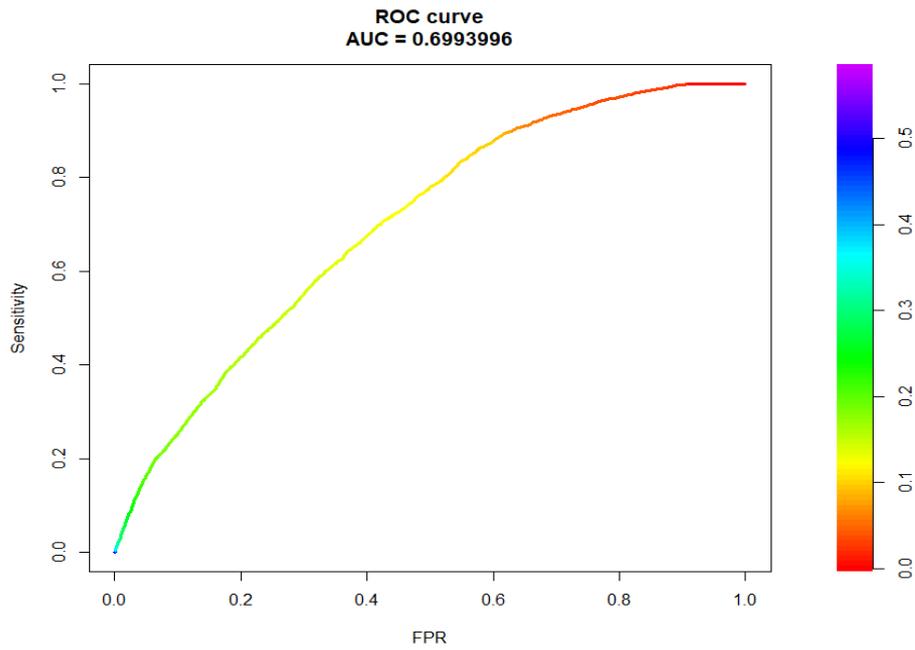


FIGURE 32 : COURBE ROC SUR DONNEES D'EVALUATION – MODELE ACPS

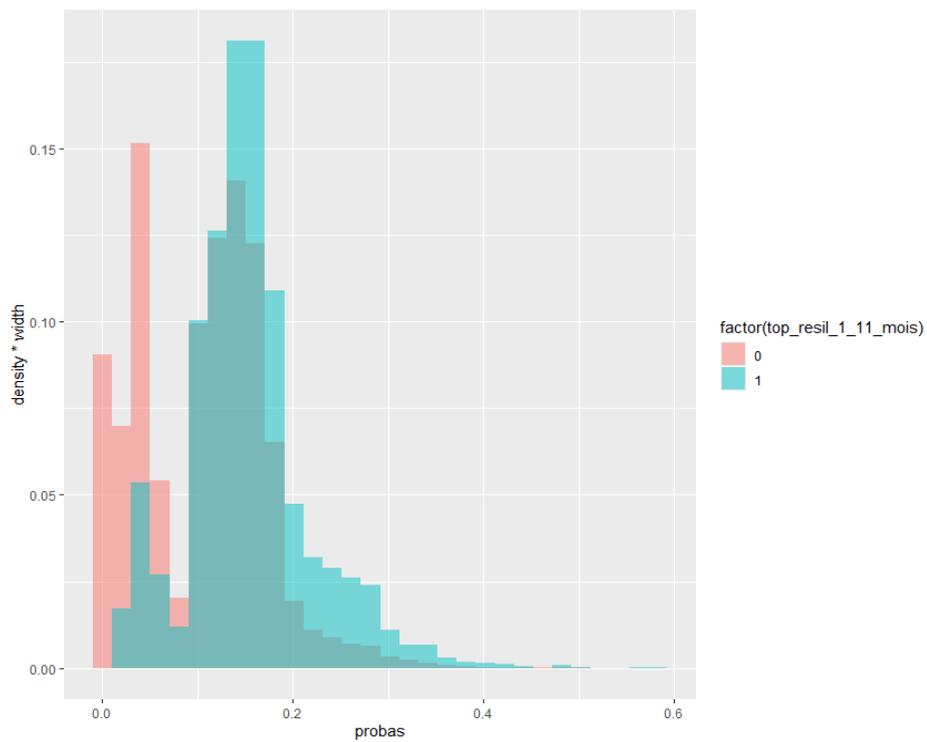


FIGURE 33 : COMPARAISON DES PROBABILITES DE RESILIATION POUR LES CONTRATS RESILIES ET RETENUS - MODELE ACPS

3. Regroupement des familles d'activité.

Nous utilisons l'algorithme **KMEANS** pour regrouper de façon optimale les différentes familles d'activité suivant l'historique de leur taux de résiliation et le nombre de contrats aussi. Le critère de coude nous permet de déterminer le nombre optimal de classes à choisir. Dans cet exemple, la figure suivante donne un coude de 3. Donc nous regroupons la famille d'activités pour le produit AC en 3 groupes : les bonnes, mauvaises et moyennes



Les familles d'activité composant le produit AC sont listés dans le tableau suivant.

Classe	famille_activite	Taux de résil moyen	Nb contrats
Bonne	Bricolage	8.9%	1603
	Commerces et services divers	9.6%	8483
	Enseignement	7.2%	1086
	Floral et Animaux	8.0%	4307
	Hôtel hébergement	8.5%	7599
	Loisirs musique	9.3%	5134
	Métal	9.1%	2349
	NR	9.2%	87
	Pierre	8.5%	176
	Santé	6.9%	7394
Mauvaise	Bar restaurant	13.1%	32260
	Habillement mode et accessoire	12.9%	12598
	Informatique Electroménager Téléphonie	12.5%	3700
Moyenne	Alimentaire	12.1%	17324
	Aménagement décoration	11.4%	6195

Autos cycles bateaux aéronautique	9.9%	5932
Beauté Hygiène Esthétique	11.7%	11468
Bois	10.3%	1685
Bâtiment	11.7%	15853
Electronique électricité	9.8%	2264
Papier Impression	10.1%	2735
Plastique	9.7%	915
Verre	9.8%	285

Liste de figures

Figure 1 : Les différents produits MRC de GENERALI	7
Figure 2 : Répartition des produits d'entrée des Professionnels chez GENERALI	8
Figure 3 : Nombre de contrats et taux de résiliation par année	14
Figure 4 : Nombre de contrats et taux de résiliation par réseau	14
Figure 5 : Nombre de contrats et taux de résiliation par Classement du réseau.....	15
Figure 6 : Nombre de contrats et taux de résiliation par type de produit	15
Figure 7 : Nombre de contrats et taux de résiliation par nombre de contrats du client	16
Figure 8 : Nombre de contrats et taux de résiliation par statut de sinistralité du contrat	17
Figure 9 : Nombre de contrats et taux de résiliation par tranche de la prime	17
Figure 10 : Nombre de contrats et taux de résiliation en fonction de la prime.....	18
Figure 11 : Nombre de contrats et taux de résiliation en fonction de la majoration	19
Figure 12 : Nombre de contrats et taux de résiliation en fonction de la majoration cumulée sur 3 ans	20
Figure 13 : Corrélations des variables quantitatives.....	22
Figure 14 : Corrélations des variables catégorielles	23
Figure 15 : Exemple de Courbe ROC.....	29
Figure 16 : Classement des variables par importance dans le modèle GLM	31
Figure 17 : Courbe ROC du modèle GLM sur données d'évaluation	31
Figure 18 : Odds Ratios du modèle GLM sur tous les produits	32
Figure 19 : Exemple de représentation d'une fonction lisse à l'aide d'une spline cubique.	36
Figure 20 : Taux de résiliation en fonction des variables continues (GAM).....	39
Figure 21 : Courbe ROC du modèle GAM sur données d'évaluation	40
Figure 22 : Comparaison des probabilités de résiliation pour les contrats résiliés et retenus	42
Figure 23 : Exemple d'arbre de décision	43
Figure 24 : Diagramme d'un gradient boosting basé sur les arbres de décision	44
Figure 25 : Courbe ROC du modèle Gradient Boosting	47
Figure 26 : Arbre réduit du modèle Gradient Boosting.....	47
Figure 27 : Taux de résiliation en fonction des variables continues (modèle ACPS)	67
Figure 28 : Courbe ROC sur données d'entraînement – Modèle Pro Artisans Commerçants	74
Figure 29 : Comparaison des probabilités de résiliation pour les contrats résiliés et retenus - Modèle Pro Artisans Commerçants.....	74
Figure 30 : Courbe ROC sur données d'entraînement - Modèle Pro Services	75
Figure 31 : Comparaison des probabilités de résiliation pour les contrats résiliés et retenus - Modèle Pro Services	75
Figure 32 : Courbe ROC sur données d'évaluation – Modèle ACPS	76
Figure 33 : Comparaison des probabilités de résiliation pour les contrats résiliés et retenus - Modèle ACPS.....	76

Liste de tableaux

Tableau 1 : Classification des clients suivant leur valeur actuelle.....	5
Tableau 2 : Evolution du multi équipement des clients MRC	16
Tableau 3 : Correspondance des variables.....	21
Tableau 4 : Odds ratios du modèle GAM sur tous les produits.....	41
Tableau 5: Elasticités moyennes par ancienneté client	51
Tableau 6 : Elasticités moyennes par type de produit	51
Tableau 7 : Elasticités moyennes en fonction de l'indicatrice de sinistre 12mois	51
Tableau 8 : Odds ratios du modèle de résiliation sur le produit Pro Artisans Commerçants	53
Tableau 9 : Odds ratios du modèle de résiliation sur le produit Pro Artisans Commerçants	56
Tableau 10 : Odds ratios du modèle de résiliation sur le produit Pro Services	59
Tableau 11 : Odds ratios du modèle de résiliation sur le produit ACPS.....	66
Tableau 12 : Elasticités moyenne par famille d'activité du produit ACPS.....	69

Note de synthèse

Cadre de l'Etude

La multirisque commerce MRC de GENERALI est une assurance destinée aux professionnels et entreprises qui souhaitent protéger leurs activités. Pour GENERALI France, la MRC représente uniquement 10% du Chiffre d'affaires sur le IARD mais ce qui rend intéressant le produit MRC c'est qu'il possède un profil de clients qui a un fort taux de multi équipement, jusqu' à 34% en 2022. La plupart de ces professionnels commencent par les produits MRC avant de se doter des produits tels que l'Auto, la MRH, l'Épargne. Pour ces types de clients, il est intéressant d'étudier à la fois la rentabilité d'un point de vue produit que d'un point de vue client puisque même si à court terme ils sont moins rentables, le potentiel de multi équipement rend intéressant ces clients sur le long terme. Un indicateur clé permet de centrer la rentabilité sur une vision client : il s'agit de la valeur client. La valeur client est la somme de la valeur actuelle et potentielle du client. La valeur actuelle correspond à la rentabilité actuelle du client. La valeur potentielle du client incorpore le potentiel de multi équipement du client et donc des valeurs que le client pourrait rapporter. La valeur client permet de piloter le parcours client chez GENERALI depuis la souscription.

Lors du renouvellement des contrats, les taux de revalorisations sont indexés sur l'inflation qui est à un niveau élevé. Pour anticiper les résiliations de suites des revalorisations, il est important d'étudier les sensibilités clients à la revalorisation de leurs primes. Aussi, plus le taux de revalorisation est important, et plus le taux de rétention est faible. Ce qui risque de dégrader à la fois la valeur actuelle avec un taux de rétention faible et la valeur potentielle avec une durée courte. Dans ce mémoire, nous modélisons les résiliations autour de la majoration pour analyser les sensibilités à la majoration.

Données d'étude

Pour analyser les résiliations dues à la majoration, nous définissons la variable suivante :

- **Top_resil_1_11mois** une variable indicatrice qui prend la valeur 1 si le client résilie son contrat un mois avant son échéance après avoir pris connaissance de la nouvelle majoration et au plus tard onze mois après l'échéance. Nous considérons donc que l'impact de la revalorisation se fait sur la fenêtre temporelle comprise entre deux revalorisations.

Les variables explicatives potentielles sont regroupées en 5 catégories :

- Les données de l'intermédiaire : réseau, classement du réseau, rabais commercial à la souscription, code affaire.
- Les données du contrat : ancienneté contrat, produit de couverture (**Pro Artisans Commerçants, Pro Fabrication, Pro Services, Pro Associations, Pro ACPS**), la prime actuelle, nombre de sinistres dans les 12 mois précédant l'échéance, les majorations passées.
- Les données du client : ancienneté client, nombre de contrats détenus par le client, nombre de famille de produits détenus, évolution du portefeuille client dans les 12 mois précédents l'échéance (montée en équipement, portefeuille stable ou dés équipement).
- Les données sur le risque : Le nombre de sites couverts, La surface du commerce en m², La qualité de l'occupant, Famille d'activité du commerce.
- Les données d'interactions : Il s'agit des différentes données récoltées durant le parcours du client à travers les échanges entre l'assuré et l'assureur

Nous disposons de deux bases de données : 2019 et 2020. Nous utilisons les données de 2019 pour entraîner les modèles et les données de 2020 pour tester leurs performances.

Modélisation

Nous utilisons 3 algorithmes de classification pour notre modélisation :

- Les Modèles linéaires généralisés
- Les modèles additifs généralisés
- Les gradients boosting à base de arbres de décisions.

Afin de pouvoir calculer implicitement les élasticités en utilisant le meilleur de chaque algorithme, nous divisons le problème en deux :

- Estimation des probas de rétentions.
- Estimations des coefficients directeurs de l'élasticités avec les GLM et GAM.

Les résultats des GLM et des GAM sont voisins avec un AUC de 72%. Les variables les plus importantes sont :

- Le multi équipement : les clients multi équipés résilient moins que les clients monos équipés toutes choses égales par ailleurs.
- La sinistralité 12 mois : les clients sinistrés ne résilient presque pas, toutes choses égales par ailleurs.
- Le type de produit : les Pro Artisans Commerçants et les Pro Services résilient le plus et les Pro Associations résilient le moins, toutes choses égales par ailleurs.

Les gradients boosting donnent un AUC de 76% avec de meilleures estimations des taux de rétentions.

Le calcul des élasticités par variables importantes permet d'avoir les résultats suivants :

Type de produit	Probas Boost	Probas GAM	Elasticités boost	Elasticités GAM	Nombre de contrats	Majo moyenne	Prime moyenne	Duration contrat
100% PRO ACPS	10.7%	11.0%	1.84	1.84	30442	1.07	1098	12.1
100% PRO ARTISANS-COMMERCANTS	12.3%	12.4%	3.78	3.78	53033	1.07	1233	4.4
100% PRO ASSOCIATION	6.7%	6.5%	2.36	2.37	9239	1.09	303	6.2
100% PRO FABRICATION	12.7%	12.7%	2.70	2.68	4972	1.09	2766	9.0
100% PRO SERVICE	12.2%	11.8%	3.35	3.37	27232	1.05	591	5.2
ANCIENNE GENERATION	12.4%	13.0%	1.10	1.09	2206	1.06	1156	22.2

Ancienneté client	Probas Boost	Probas GAM	Elasticités boost	Elasticités GAM	Nombre de contrats	Majo moyenne	Prime moyenne
<5ans	13.6%	13.5%	4.71	4.71	56021	1.07	922
5-10ans	10.2%	10.4%	2.44	2.43	30733	1.07	1028
10-20ans	9.6%	9.5%	1.44	1.44	29670	1.07	1229
20ans +	9.8%	9.8%	0.43	0.43	10700	1.06	1334

Top sinistre 12mois	Probas Boost	Probas GAM	Elasticités boost	Elasticités GAM	Nombre de contrats	Majo moyenne
Aucun sinistre	12.6%	12.6%	2.98	2.98	116273	1.07
1 sinistre ou plus	0.1%	0.1%	3.58	3.58	10851	1.07

On remarque que les variables qui influencent le plus l'élasticité sont :

- Le type de produit : Les produits les plus sensibles sont les Pro Artisans Commerçants et les Pro Services. Le Produit le moins sensible est le Pro Association.
- L'ancienneté client : Plus les clients sont anciens, moins ils résilient et moins ils sont sensibles à la revalorisation.
- Sinistralité 12 mois : Les contrats sinistrés sur les 12 mois précédant l'échéance ne résilient presque pas mais ont une sensibilité plus élevée que les clients non sinistrés.

L'analyse des sensibilités permet de déterminer les contrats les moins sensibles qui ont les caractéristiques suivantes :

- Les détenteurs des produits Pro Associations
- Les contrats ayant plus de 5 ans d'ancienneté avec un rabais initial compris entre 20 et 50%
- Les clients ayant eu au moins 1 sinistre sur les 12 mois précédant l'échéance

Pour étudier les disparités pour chaque sous-produit, nous étudions séparément les 3 Produits les plus importants des gammes MRC qui sont les Pro Artisans Commerçants, les Pro Services et les Pro ACPS.

Ces modèles montrent que pour :

- Les Pro Artisans Commerçants : les contrats les plus sensibles les fortes primes, les contrats plus anciens et les familles d'activité telles que les Bars restaurants, l'Alimentaires, les commerces divers. Les profils les moins sensibles en revanche sont les clients sinistrés
- Les Pro Services : les contrats les plus sensibles sont les contrats de petites primes, les familles d'activités autres que les *Bureaux Seuls*.
- Les Pro ACPS : La sensibilité dépend uniquement des taux de rétentions.

Les valeurs d'élasticités et les probas de résiliation permettent d'identifier les profils sensibles à la revalorisation et de mettre en place des scénarios de majorations, de sensibiliser les Agents et Courtiers sur le risque de résiliation des clients. Le croisement des probabilités par groupes de sensibilité et par ancienneté permet d'avoir une loi de chute pour chaque profil de client et chaque scénario d'évolution de la revalorisation. Cette loi de chute permettra de mettre à jour la valeur client. Les valeurs clients mises à jour permettront suivant les nouvelles valeurs obtenues de choisir les clients à protéger contre les forts taux de revalorisations.

Executive Summary

Framework of the study

GENERALI's commercial multi-risk MRC is an insurance product designed for professionals and companies who wish to protect their activities. For GENERALI France, the MRC represents only 10% of the P&C business, but what makes the MRC product interesting is that its client profile has a high multi-equipment rate, up to 34% in 2022. Most of these professionals start with MRC products before adding products such as Auto, MRH, Savings. For these types of clients, it is interesting to study both the profitability from both a product and a client point of view, since even if they are less profitable in the short term, the multi equipment potential makes these clients interesting in the long term. A key indicator allows us to focus profitability on a customer view: customer value. The customer value is the sum of the current and potential value of the customer. The current value corresponds to the current profitability of the customer. The potential value of the customer incorporates the potential for multi-equipment of the customer and therefore the values that the customer could bring back. The customer value makes it possible to pilot the customer experience at GENERALI starting with the underwriting.

During the renewal of contracts, the revaluation rates are indexed to inflation, which is at a high level. In order to anticipate cancellations due to revaluations, it is important to study the sensitivity of customers to the premium revaluation. Also, the higher the premium increase the lower the retention rate. This risk worsening both the present value with a low retention rate and the potential value with a short duration. In this memoire, we model terminations around the markup to analyze sensitivities to the markup.

Study Data

To analyze the terminations due to the markup, we define the following variable:

- **Top_resil_1_11month** an indicator variable that takes the value 1 if the customer cancels his contract one month before its expiration after having been informed of the new surcharge and at the latest 11 months after the expiration. We therefore consider that the impact of the revaluation is carried out over the time window between two revaluations.

The potential explanatory variables are grouped into 5 groups:

- Intermediary data: network, network ranking, commercial discount at underwriting, business code.
- Contract data: contract seniority, coverage product (*Pro Artisans Commerçants, Pro Fabrication, Pro Services, Pro Associations, Pro ACPS*), current premium, number of claims in the 12 months before expiration, past surcharges.
- Client data: client seniority, number of contracts held by the client, number of different lines of insurance held, changes in the client portfolio in the 12 months before expiry (increase in assets, stable portfolio or reduction in assets).
- Risk data: Number of sites covered, size of the business in m², quality of the owner, business activity family.

- Interaction data: This is the various data collected during the customer's journey through the exchanges between the insured and the insurer.

We have two databases: 2019 and 2020. We use the 2019 data to train the models and the 2020 data to test the models' performance.

Modeling

We use 3 classification algorithms for our modeling:

- Generalized linear models
- Generalized additive models
- Gradient boosting based on decision trees.

In order to implicitly compute the elasticities using the best of each algorithm, we divide the problem into two:

- Estimation of the retention probabilities.
- Estimation of the directing coefficients of the elasticities with GLM and GAM.

The results of GLM and GAM are close with an AUC of 72%. The most important variables are:

- Multi-equipment: multi-equipped customers terminate less than single-equipped customers, all other things being equal.
- 12-month claims experience: customers with claims experience almost never cancel, all other things being equal.
- Type of product: Pro Artisans Commerçants and Pro Services cancel the most and Pro Associations cancel the least, all other things being equal.

The boosting gradients give an AUC of 76% with better estimates of retention rates.

Calculating the elasticities by important variables yields the following results:

Type de produit	Probab Boost	Probab GAM	Elasticities boost	Elasticities GAM	Number of contracts	Avg Markup	Avg premium	Contract duration
100% PRO ACPS	10.7%	11.0%	1.84	1.84	30442	1.07	1098	12.1
100% PRO ARTISANS-COMMERCANTS	12.3%	12.4%	3.78	3.78	53033	1.07	1233	4.4
100% PRO ASSOCIATION	6.7%	6.5%	2.36	2.37	9239	1.09	303	6.2
100% PRO FABRICATION	12.7%	12.7%	2.70	2.68	4972	1.09	2766	9.0
100% PRO SERVICE	12.2%	11.8%	3.35	3.37	27232	1.05	591	5.2
ANCIENNE GENERATION	12.4%	13.0%	1.10	1.09	2206	1.06	1156	22.2

Client seniority	Probab Boost	Probab GAM	Elasticities boost	Elasticities GAM	Number of contracts	Avg Markup	Avg premium
<5Y	13.6%	13.5%	4.71	4.71	56021	1.07	922
5-10Y	10.2%	10.4%	2.44	2.43	30733	1.07	1028
10-20Y	9.6%	9.5%	1.44	1.44	29670	1.07	1229
20Y +	9.8%	9.8%	0.43	0.43	10700	1.06	1334

Top sinistre 12months	Probas Boost	Probas GAM	Elasticities boost	Elasticities GAM	Number of contracts	of Major Markup
No claim	12.6%	12.6%	2.98	2.98	116273	1.07
1 claim or more	0.1%	0.1%	3.58	3.58	10851	1.07

We note that the variables that impact elasticity the most are:

- Product type: The most sensitive products are Pro Artisans Commerçants and Pro Services. The least sensitive product is the Pro Association.
- Customer seniority: The older the clients are, the less they cancel and the less sensitive they are to the increase.
- 12-month claims experience: Policies with claims in the 12 months before maturity hardly ever terminate but are more sensitive than those without claims.

The analysis of sensitivities allows us to identify the least sensitive policies, which have the following characteristics

- Holders of Pro Associations products
- Contracts that are more than 5 years old with an initial discount of between 20% and 50%.
- Clients who have had at least 1 claim in the 12 months before maturity

To study the disparities for each sub-product, we study separately the 3 most important products of the MRC ranges which are Pro Artisans Commerçants, Pro Services and Pro ACPS.

These models show that for:

- Pro Artisans Commerçants: the most sensitive policies are the high premiums, the older policies and the families of activity such as Bars and Restaurants, Food, and various businesses. The least sensitive profiles, on the other hand, are customers who have suffered a loss.
- Pro Services: the most sensitive contracts are those with small premiums, and families of activities other than *Offices*.
- Pro ACPS: The sensitivity depends solely on the retention rates.

Elasticity values and lapse rates allow us to identify profiles that are sensitive to revaluation and to set up increase scenarios, to make Agents and Brokers aware of the risk of customer cancellations. The probabilities by sensitivity group and by seniority are crossed a lapse rate for each client profile and each scenario of increase in value. These lapse rate will allow to update the customer value. The updated client values will allow us to choose the clients to protect against high resets according to the new values obtained.