

Analyse de données assurantielles

BOUSSENGUI François

Université de Tours - Master Économiste d'entreprise

September 18, 2021



Table des matières

- 1 APRIL Moto
- 2 Outil de veille concurrentielle
- 3 Prévention de l'attrition
- 4 Estimation de la durée de vie

Table des matières

1 APRIL Moto

- Filiale du groupe APRIL
- **Courtier en assurance** : commerçant indépendant jouant le rôle d'intermédiaire dans une opération commerciale entre une compagnie d'assurances et le consommateur final
- **Courtier grossiste en assurance** : courtier en assurance spécialisé sur une ou plusieurs niches de marché
- Conçoit, gère et distribue des produits d'assurances pour les deux-roues



Table des matières

2 Outil de veille concurrentielle

Présentation

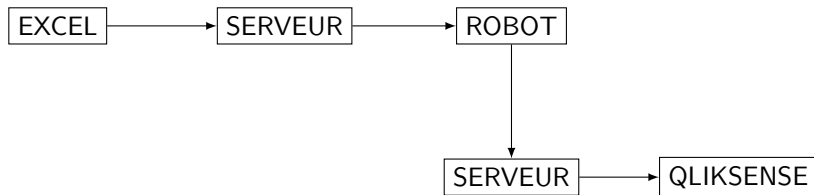
- Robot de web-scraping qui récupère quotidiennement les tarifs d'assurances des principaux concurrents d'APRIL Moto sur des sites internet de comparateurs en fonction des différents scénarios
- 50 scénarios représentatifs des différents cas pertinents du portefeuille clients ont été définis. Ils possèdent les caractéristiques suivantes : **sexe, date de naissance, profession, type de véhicule (Moto, Scooter, Cyclomoteur, NVEI), la marque, le modèle, sa situation et maritale, etc.**

Les objectifs de la mission ont été les suivants :

- S'assurer que le robot passe bien sur les scénarios définis
- Augmenter son taux de succès (qui était d'environ 70 %)
- Améliorer la fluidité dans la remontée et l'analyse des résultats

Fonctionnement

Le processus d'exécution du robot est le suivant :



Résultats

- Création de nouveaux scripts adaptés aux différents scénarios
- Le taux de succès quotidien est actuellement de 91 %
- Une application Qlik Sense est alimentée quotidiennement par les informations récupérées

Limites et maintenance

Limites :

- Les captchas
- Débit internet
- Évolution de la structure des formulaires

Maintenance :

- En consultant la page de suivi technique de l'application Qlik Sense
- En consultant les fichiers émanant de la journalisation

Application QS

Une application Qlik Sense a été créée afin de visualiser les résultats récupérés. Elle possède les feuilles suivantes :

- **Baromètre** : contient des statistiques générales.
- **Baromètre (Avec possibilité de filtrer sur la date)** : cette feuille a exactement la même finalité que la première à la différence près que l'on peut sélectionner une date précise afin d'avoir un comparatif avec les dernières données récupérées.
- **Analyse** : contient des analyses plus détaillées. (évolution de la prime moyenne proposée par les assureurs ; filtrage sur un scénario en particulier)
- **Suivi technique** : contient des informations sur la performance du robot.

Table des matières

3 Prévention de l'attrition

Objectif

- Limiter et anticiper le départ des clients
- Mise en place de techniques d'apprentissages automatique \Rightarrow identifier les personnes ayant un risque de départ plus élevé que d'autres

Les données

87042 contrats actifs ou résiliés sur la période du 01-06-2020 au 01-09-2020 grâce à des requêtes SQL sur les BDD de internes.

Nettoyage et mise en forme de la BDD

- Gestion des données manquantes
- Les données erronées
- Recodage et transformation des certaines variables

⇒ On se retrouve avec une BDD contenant 78019 observations et 38 variables que l'on peut classer en 4 groupes :

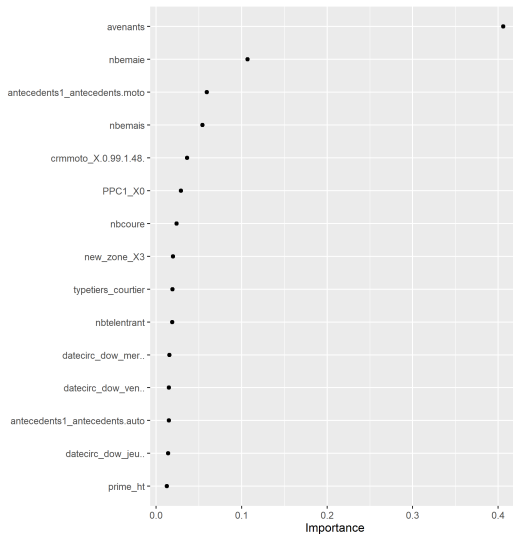
- **Les variables individuelles**
- **Les variables relatives aux caractéristiques du contrat de l'assuré**
- **Les variables relatives aux caractéristiques du véhicule de l'assuré**
- **Les variables relatives au canal d'acquisition**

Modélisation

État du contrat	Effectif	%
Actif	68338	87.6
Résilié	9681	12.4

- Prise en compte du déséquilibre des classes
- Métriques d'évaluation de la performance : **Précision ; Recall**
- Algorithme retenu : **Gradient Tree Boosting**

Résultats



Matrice de confusion

	Actifs (Réalité)	Résiliés (Réalité)	Total
Actifs (Prédictions)	16800	2086	18886
Résiliés (Prédictions)	285	335	620
Total	17085	2421	19506

- **Précision** : 0.54 \Rightarrow Lorsque notre modèle prédit un individu a résilié son contrat, il est correct 54 % du temps.
- **Recall** : 0.13 \Rightarrow Le modèle identifie correctement 13 % des individus dont le contrat est résilié.

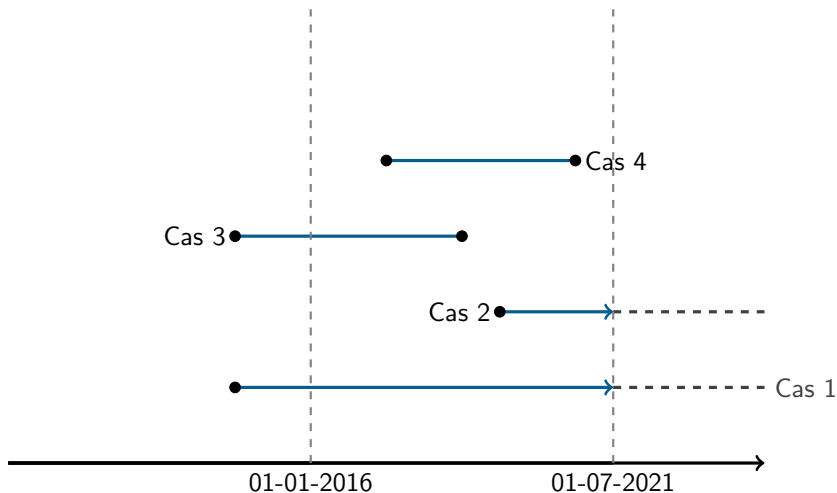
Perspectives d'amélioration

- Revoir le choix des variables à inclure
- Essayer d'autres modélisations

Table des matières

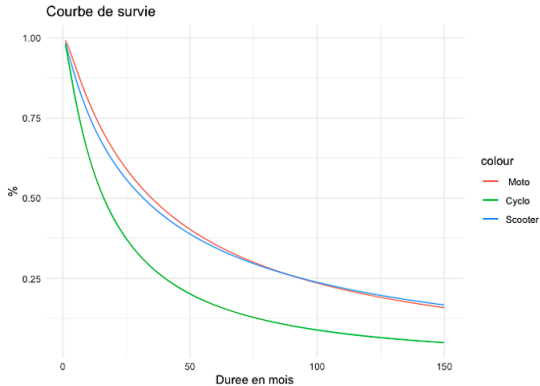
4 Estimation de la durée de vie

Contexte



Résultats

⇒ Estimations réalisées grâce à la **loi log-normale** notée $\text{Log-N}(\mu, \sigma^2)$



- ⇒ Ecart significatif entre la durée de vie observée et la durée de vie estimée du au fait que le motif de résiliation est variable selon l'individu
- ⇒ Prise en compte dans la modélisation des différents motifs de résiliation : **Modèle à risque compétitif**

Conclusion

Mon apport durant ce stage :

- Amélioration de l'outil de veille concurrentielle
- Réalisation d'études ponctuelles
- Détermination de la valeur client

L'apport de ce stage :

- Découverte de nouveaux logiciels : QlikSense, Talend
- Gain en compétence de programmation
- Travail en autonomie et en équipe

MERCI POUR VOTRE ATTENTION !