

# Projet Mazars - Centrale

## Tarification de contrats automobiles par des algorithmes de *Machine Learning*

Marie Julienne Diouf  
Imen Said

09 Octobre 2019

### 1 Intitulé

Mazurances, une compagnie souscrivant des polices d'assurances automobile, fait appel à votre expertise pour challenger sa politique de tarification. À cet effet, vous disposez des contrats de son portefeuille constitué d'un peu plus de 100 000 polices. Ce portefeuille est scindé en 2 parties à savoir une partie de *train* constituée des bases **train.contrats** et **train.sinistres** sur laquelle vous pouvez façonner vos modèles et d'une partie de *test* constituée de la base **test.contrats** sur laquelle vous entraînerez vos modèles.

Vous disposez ainsi de 3 fichiers csv contenant chacune une base qui s'articule comme expliqué ci-dessous.

### 2 Description des variables

La base **train.contrats** contient 87 228 polices d'assurance, avec les variables suivantes

- Num\_contrat: Numéro d'affaire
- IMMAT: Immatriculation
- Date\_deb\_situ: Date de début de situation
- ANNEE: Année de situation
- Date\_fin\_situ: Date de fin de situation
- Classe\_age\_situ\_cont: Classe âge du contrat
- Type\_apporteur: Type d'apporteur (agents ou courtiers ou autres)

- creation\_entr: Indicateur de création entreprise (à 2 si création)
- Mode\_gestion: Mode de gestion de la otte
- activité: Activité du client (niveau contrat)
- zone : Zone géographique
- segment : Taille de la flotte
- Fractionnement: Fractionnement de paiement de la prime
- Exposition\_au\_risque: Véhicules année (exposition au risque)
- Age\_du\_véhicule: Classe d'âge du véhicule
- Valeur\_puissance: Combiné puissance valeur du véhicule
- franchise: Classe de franchise de la garantie DTA
- Nombre\_vehicule: Nombre de véhicule
- valeur\_assuree: Valeur assurée
- formule: Formule souscrite
- Catégorie\_ensemble: Catégorie du véhicule
- nombre\_de\_sinistre: Nombre de sinistres

La base **test\_contrats** contient les mêmes variables, mais sans informations relatives au nombre de sinistres. La base comporte 32 772 polices d'assurance.

La base **train\_sinistres** contient des informations relatives à 4 568 sinistres observés :

- Num\_contrat: Numéro d'aire
- CHARGE\_SINISTRE: Charge du sinistre
- IMMAT: Immatriculation
- Date\_Deb\_Situ: Date de début de situation
- Date\_Fin\_Situ: Date de fin de situation
- IDA: 1 si sinistre Indemnisation Direct de l'Assuré
- Responsabilité\_IDA: Taux de responsabilité de l'IDA
- ANNEE\_CLOTURE: Année de clôture du sinistre si le sinistre est clôturé
- charge sinistre: Charge sinistre
- ANNEE: Année de situation