

Compétition de tarification Auto

Date limite de rendu : Dimanche 8 janvier 2023 – 23h59

Contexte

Vous êtes en charge de l'estimation des primes pures d'une assurance auto pour l'année à venir. Vous avez à votre disposition un historique (base dataTrain) pour 60 000 polices observées pendant au moins 3 mois lors de l'année précédente. Il faut donc construire un modèle de tarification à l'aide des outils d'apprentissage statistique. Ce modèle de tarification devra ensuite être utilisé pour proposer des primes à 20 000 nouveaux assurés (base test).

Livrables

Les livrables pour ce projet sont :

- un rapport écrit sous format pdf ;
- un fichier avec les primes pour chaque nouvel assuré ;
- le code de la prédiction pour le modèle retenu.

Le **rapport écrit** devra détailler la démarche suivie pour la création du modèle de tarification. Il décrira notamment :

- les données utilisées (feature engineering, data management, ...) ;
- la méthodologie pour le choix du modèle (algorithmes testés, méthode de comparaison, ...) ;
- l'estimation du modèle ;
- le modèle retenu (comportement, interprétation, pouvoir de prédiction, ...).

Les **primes** seront sous la forme d'un fichier csv (séparateur virgule) nommé *premium_n.csv* avec deux colonnes : id et premium.

Le **code** (dans le langage que vous souhaitez), prendra en entrée la base de données initiale et permettra de créer le fichier de primes avec le modèle retenu.

Critères de notation

Les critères de notation pour ce projet sont :

- la propreté du rapport ;
- la pertinence de la démarche suivie : détaillez vos réflexions !
- l'originalité de la démarche ;
- la compréhension du comportement du modèle retenu ;
- le résultat de la tarification (cf. ci-dessous).

Afin de comparer les modèles de tarification des groupes, une simulation de marché sera mise en place :

1. des marchés seront créés à partir de 6 groupes tirés aléatoirement ;
2. dans chaque marché, les nouveaux assurés se voient proposer les primes de chaque groupe ;
3. les nouveaux assurés choisissent alors la prime la plus faible ;
4. les groupes obtiennent de leurs assurés ainsi attribués les primes proposées et les sinistres (inconnus au moment de la tarification).

La mesure de comparaison est alors le résultat comptable du groupe, i.e. primes – sinistres sur les assurés du groupe.

Organisation pratique

Le projet est à faire par groupe de 3 étudiants au maximum.

Quand un groupe est constitué, envoyez un mail à pierrick.piette@gmail.com en mettant en cc les membres du groupe, ainsi que le nom et prénom de chaque membre. Vous recevrez un lien pour accéder aux données **propres à votre groupe**.

Rendu effective du projet (i.e. réception de tous les documents) par mail à pierrick.piette@gmail.com

- 48h avant la date limite (i.e. avant vendredi 6 janvier 2023 – 23h59) : + 0,5 point bonus ;
- retard par tranche de 12 heures : - 1 point de pénalité
- au-delà de 48h (i.e. mardi 10 janvier 2023 - 23h59) : pas de correction

Dictionnaire de données

Variable	Définition
gender	Genre du conducteur / conductrice
carType	Type de la voiture
carCategory	Catégorie de la voiture
occupation	Statut professionnel du conducteur
age	Age du conducteur / conductrice
carGroup	Groupe de la voiture
bonus	Bonus-Malus sur la dernière année
carValue	Valeur de la voiture
material	Indicatrice pour la couverture matérielle
region	Région géographique de la résidence
subRegion	Sous-région géographique de la résidence
cityDensity	Densité de la population (hbt / km ²) de la ville de résidence
exposure	Nombre de jour d'observation sur l'année précédente
claimNumber	Nombre de sinistres déclarés
claimValue	Valeur indemnisée au titre des sinistres déclarés