

Prédiction de cross-selling en assurance

Date limite de rendu : Dimanche 10 juillet 2022 – 23h59

Contexte

Une compagnie d'assurances santé souhaite mettre à profit sa base de données clients afin de diversifier ses activités. Pour ce faire, elle contacte ses nouveaux assurés pour leur proposer de souscrire à une assurance auto. Un premier test a déjà été mis en place sur quelques dizaines de milliers d'assurés. Durant cette campagne marketing, des données ont été récupérées et structurées pour savoir si la personne était effectivement intéressée par une assurance auto, ou non.

L'équipe actuariat doit maintenant créer un modèle de prédiction à partir de cette base pour déterminer si la personne contactée sera intéressée par la souscription d'une assurance auto.

Livrables

Les livrables pour ce projet sont :

- un rapport écrit ;
- un fichier avec les prédictions de la base test ;
- le code de la prédiction pour le modèle retenu.

Le **rapport écrit** devra détailler la démarche suivie pour la création du modèle. 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 **prédictions** seront sous la forme d'un fichier csv (nommé `groupe_n_predictions.csv`) avec deux colonnes : id et Response. Ce fichier servira à estimer le pouvoir de prédiction du modèle sur la base test.

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

Critères de notation

Les critères de notation pour ce projet sont :

- la **propreté** du rapport et du code ;
- la pertinence de la démarche suivie (détaillez vos réflexions !) et son **originalité** ;
- la compréhension du comportement du modèle retenu ;
- le pouvoir de prédiction du modèle et l'écart de ce pouvoir par rapport à ce qui est annoncé dans le rapport.

Pour le pouvoir de prédiction, le critère quantitatif appliqué sera le F1-score :

$$F_1 = 2 \times \frac{\text{précision} \times \text{rappel}}{\text{précision} + \text{rappel}}$$

avec $\text{précision} = VP / (VP + FP)$ et $\text{rappel} = VP / (VP + FN)$

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 8 juillet 2022 – 23h59) : + 0,5 point bonus
- Retard par tranche de 12 heures : - 1 point de pénalité
- au-delà de 48h (i.e. mardi 12 juillet 2022 - 23h59) : pas de correction

Dictionnaire de données

Variable	Definition
Gender	Gender of the customer
Age	Age of the customer
Driving_License	0 : Customer does not have DL 1 : Customer already has DL
Region_Code	Unique code for the region of the customer
Previously_Insured	1 : Customer already has Vehicle Insurance, 0 : Customer doesn't have Vehicle Insurance
Vehicle_Age	Age of the Vehicle
Vehicle_Damage	1 : Customer got his/her vehicle damaged in the past. 0 : Customer didn't get his/her vehicle damaged in the past.
Annual_Premium	The amount customer needs to pay as premium in the year
Policy_Sales_Channel	Anonymized Code for the channel of outreaching to the customer e.g. Different Agents, Over Mail, Over Phone, In Person, etc.
Vintage	Number of Days, Customer has been associated with the company
Response	1 : Customer is interested 0 : Customer is not interested