

Voici le sujet d'évaluation.

Il s'agit d'un challenge data proposé par la compagnie d'assurance Generali en 2019.

<https://challengedata.ens.fr/challenges/19>

Le but est de prédire si un bâtiment fera l'objet d'une réclamation d'assurance pendant une certaine période. La variable cible est :

- 1 si le bâtiment présente au moins un sinistre sur la période assurée.
- 0 si le bâtiment ne présente aucun sinistre sur la période assurée.

En plus d'avoir un modèle prédictif qui prédit au mieux la probabilité de sinistre, il est demandé pour cette évaluation d'expliquer quelles variables ont le plus d'impact sur la variable cible.

Vous serez évalué sur ce que l'on a vu en cours, ne négligez pas la partie analyse et correction du dataset. Ayez une démarche cohérente pour le prétraitement des données. Ne passez pas tout votre temps à paramétrer les modèles finaux.

Certaines variables sont anonymisées, cela n'empêche pas de faire une analyse graphique et de ressortir de l'information de ces derniers.

L'important pour cette évaluation est la démarche et la justification de l'ensemble de vos choix, même si une hypothèse que vous avez pu faire n'obtient pas les résultats escomptés il est important de la présenter et d'avoir une démarche cohérente.

Vous serez évalué sur ce que l'on a vu en cours, ceci n'empêche pas de proposer de nouvelles choses (modèles, métriques, visualisation...) si vous le jugez pertinent, ceci sera évidemment valorisé.

De plus, vous pouvez ajouter des variables externes via le numéro d'Insee pour améliorer les performances du modèle si vous le souhaitez.

Vous trouverez le jeu de données à utiliser en pièce jointe de ce mail.

Pour information, dans le challenge originel, il est demandé d'évaluer les résultats sur un jeu de test fourni, via une métrique non vue en cours (score de gini), il n'est pas demandé pour cette évaluation d'utiliser ce jeu de données et d'utiliser cette métrique.

Vous avez jusqu'au 22 avril inclus pour rendre un rapport qui présente vos résultats et votre démarche, il est demandé également de fournir le code.

Je reste à votre disposition si vous avez des questions. Voici mon mail pro où vous pouvez me contacter (je serais plus réactif sur celle-ci) : gregory.gauthier@businessdecision.com