



Test technique

Objectif

Cet exercice a pour but d'évaluer vos compétences en machine learning. Vous serez évalués selon les critères suivants :

- Statistiques descriptives
- Machine Learning
- Restitution des résultats et explications
- Clarté et qualité du code

Description et énoncé

Le jeu de données qui vous a été transmis contient deux fichiers : train.csv et test.csv. Les données sont composées de 14 features (X_i), catégoriques et numériques. Le fichier train contient les données d'entraînement (32561 lignes). Le but est de prédire la colonne Y en fonction des autres ($X...$).

Le jeu de données est anonymisé, rendant l'interprétation plutôt difficile, mais ce n'est pas le sujet de cet exercice. Nous avons choisi de ne pas avoir de questions spécifiques, afin de laisser le candidat libre d'aborder le problème comme bon lui semble. Vous pouvez utiliser les techniques et algorithmes de votre choix.

La seule contrainte : produire un notebook jupyter qui nous sera transmis par la suite et qui contiendra la description des différentes étapes et opérations, les insights sur le dataset, etc. Vous ajouterez également une évaluation de la performance attendue pour le modèle final obtenu.

Vous pourrez ensuite joindre un fichier avec les prédictions pour les observations contenues dans le fichier de test.csv, ce qui nous permettra d'évaluer la généralisation du modèle (à la Kaggle).