



Description des projets

Initiation au Machine Learning

Présenté par : Mously Diaw



Classification

13 propositions de projet



CLF01

Churn score

Vous êtes Data Scientist au sein de "fortuneo banque", une banque en ligne qui propose une gamme variée de services financiers et bancaires pour des personnes ayant peu ou pas du tout d'historique de prêt.

L'entreprise souhaite mettre en œuvre un outil de "churn scoring" pour calculer la probabilité qu'un client quitte la banque, puis classifie le client en 'churn' ou pas. Elle souhaite donc développer un algorithme de classification en s'appuyant sur des sources de données variées (données comportementales, données provenant d'autres institutions financières, etc.).

Objectif : Identifier les clients susceptibles de quitter la banque.

Impact : Réduire le churn en prenant des mesures préventives ciblées pour retenir les clients à haut risque.

Les données: Voici les données dont vous aurez besoin pour réaliser l'algorithme de classification. L'ensemble des données contient des informations sur les clients des banques qui ont quitté la banque ou qui restent clients. La description des attributs est disponible ici

- Lien github contenant:
 - Un notebook de l'analyse exploratoire mis au propre et annoté.
 - Le code (ou un notebook) des différents tests de modèles mis au propre, dans lequel vous identifierez clairement le modèle final choisi.
 - Tous les scripts Python
 - Le code pour les tests unitaires, tracking, déploiement continu, ...
- Un support de présentation pour la soutenance (25 slides maximum)

Pour faciliter votre passage au jury, déposez sur Github tous les livrables du projet. La structure du répertoire github est à définir mais il doit contenir au minimum un dossier appelé “notebooks” contenant les notebooks créés pour répondre à cette mission. A titre illustratif, vous pouvez nommer les fichiers selon l'ordre dans lequel il apparaît, par exemple “churn_01_analyse.ipynb”, “churn_02_model.ipynb”, et ainsi de suite.

- 15 à 20 min de Présentation
 - Présentation de la problématique, de son interprétation et des pistes de recherche envisagées.
 - Présentation du cleaning effectué, du feature engineering et de l'exploration.
 - Présentation des différentes pistes de modélisation effectuées.
 - Présentation du modèle final sélectionné ainsi que des améliorations effectuées.
 - Démo de l'API de prédiction et/ou le dashboard (si besoin)
- 5 à 10 min de questions-réponses