

Examen Data Science – IA

Travaux pratiques

Duree : 2H00

Decouverte du jeu de donnees

- **Le fichier :** *“WA-Fn-UseC-Telco-Customer-Churn.csv” (978 KB)*

What's in the Telco Customer Churn data set?

This data set provides info to help you predict behavior to retain customers. You can analyze all relevant customer data and develop focused customer retention programs.

A telecommunications company is concerned about the number of customers leaving their landline business for cable competitors. They need to understand who is leaving. Imagine that you're an analyst at this company and you have to find out who is leaving and why.

The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

Instructions

Envoyer votre fichier contenant le code python (notebook ou autre fichier de code python) à l'adresse mail suivante :
cmeniche@hotmail.com

LE NOM DU FICHIER DOIT CONTENIR VOTRE NOM

Mettre en sujet du mail : examen data science TRAVAUX PRATIQUE

Examen

Vous devez aider l'entreprise "Blue Telco" à créer un modèle prédictif, pour prédire quel client "risque" de quitter l'entreprise (taux de churn)

Vous disposez d'une "base" historique (le fichier *"WA-Fn-UseC-Telco-Customer-Churn.csv"*), indiquant quels clients ont quitté dans le passé "Blue Telco".

Vous devez :

1. Construire au moins 2 modèles de prédiction différents
2. Ensuite, une fois l'étape 1 finie, Choisir un des 2 modèles que vous aurez construits, et améliorer sa précision de prédiction (accuracy)

Important : Il n'y a pas de correction type (standard) sur le code à fournir. Vous êtes libre de votre démarche, du choix des features, choix des modèles et libre de vos manipulations et code

Le critère d'évaluation ("notation") sera votre démarche de data scientist et vos explications (n'hésitez surtout pas à commenter). Cette évaluation n'est pas une course à la précision, la précision du modèle ne sera pas le critère d'évaluation