



## Test technique d'Alternant Data Scientist

### Contexte :

Le groupe TF1 dispose d'une plateforme en ligne appelée MYTF1, qui permet aux utilisateurs de regarder gratuitement en streaming les programmes du groupe. Chaque semaine, des millions d'internautes se connectent pour profiter de la variété des programmes disponibles sur MYTF1.

### La mission :

Dans le cadre de cette mission, nous vous fournirons un fichier au format CSV contenant les historiques de visionnage de différents utilisateurs sur la plateforme sur une période d'un mois. Votre rôle consistera à prédire le genre de l'utilisateur (homme ou femme) en fonction des données que vous avez à disposition.

Le rendu attendu est un Jupyter notebook en Python détaillant votre approche.

### Définition des colonnes :

Le fichier CSV comprend les colonnes suivantes :

- "programs" : une liste des identifiants des programmes visionnés par l'utilisateur, dans l'ordre de visionnage.
- "user\_id" : un identifiant unique pour chaque utilisateur.
- "timestamps\_vector" : une liste de timestamps représentant les moments de visionnage de l'utilisateur, triés de manière croissante.
- "device\_vector" : une liste des appareils utilisés par l'utilisateur, dans l'ordre de visionnage.
- "channels\_vector" : une liste des chaînes diffusant les programmes visionnés par l'utilisateur, dans l'ordre de visionnage.
- "program\_genre\_vector" : une liste des genres de programmes visionnés par l'utilisateur, dans l'ordre de visionnage.
- "user\_gender" : C'est notre target. Contient deux valeurs (m pour homme et f pour femme).

Il est important de noter que notre objectif principal n'est pas d'obtenir une solution précise, mais plutôt de tester votre approche pour résoudre un problème d'apprentissage automatique. Si vous avez des questions, n'hésitez pas à me contacter à l'adresse suivante : [rgharbi@tf1.fr](mailto:rgharbi@tf1.fr).

