

FORMATION Mlops

Analyse et compréhension simple du problème Customer Churn

Le **customer churn** correspond au fait qu'un client quitte un service ou résilie son abonnement. Ce phénomène est très fréquent dans les entreprises qui fonctionnent par abonnement, comme les entreprises de télécommunication et les banques.

Dans cette dataset (issue de Kaggle), chaque ligne représente un client et décrit ses informations personnelles, son type de contrat, les services qu'il utilise ainsi que les montants qu'il paie.

L'objectif principal est de prédire si un client va partir (**Churn = Yes**) ou rester (**Churn = No**) à l'aide du machine learning.

1-Objectif du problème

L'objectif est d'aider l'entreprise à anticiper le départ des clients afin de mettre en place des actions de fidélisation comme des réductions ou une amélioration du service.

2-Type de problème

Il s'agit d'un problème de classification binaire en machine learning. Le modèle doit prédire une seule information parmi deux possibilités :

- le client quitte le service (**Churn = Yes**),
- le client reste fidèle à l'entreprise (**Churn = No**).

3-Variable cible

La variable cible est la colonne Churn qui indique si le client part ou non.

4-Variables explicatives

Les variables explicatives décrivent le profil et le comportement du client. Elles peuvent être regroupées en plusieurs catégories :

- **Informations personnelles** : sexe, situation familiale(Partner,Dependents).
- **Informations contractuelles** : type de contrat, ancienneté du client(tenure), méthode de paiement.
- **Services utilisés** : accès à Internet, support technique, services de streaming.
- **Informations financières** : charges mensuelles et montant total payé par le client.

Ces variables permettent de mieux comprendre pourquoi certains clients quittent le service plus souvent que d'autres.

5-Enjeux principaux

Cette dataset présente plusieurs défis importants :

- la présence de nombreuses **variables catégorielles** qui doivent être transformées avant l'entraînement du modèle,
- la nécessité d'un **bon prétraitement des données** pour obtenir de bonnes performances.

Ces éléments doivent être pris en compte lors de la construction du modèle.

6-Importance pour le MLOps

Ce problème est intéressant en MLOps car le comportement des clients peut évoluer dans le temps, ce qui nécessite un suivi du modèle et un réentraînement régulier.