

Complément suite aux retours de l'oral

Pourquoi XGBoost et pas un autre modèle ?

Nous avons choisi **XGBoost** car il est particulièrement performant sur des **données tabulaires** comme celles des transactions (pays, heure, type, montant, etc.).

Il capture efficacement des **relations complexes et non-linéaires**, ce qui est essentiel pour repérer des schémas de fraude.

Il est aussi **rapide**, robuste et facile à mettre en production pour faire du scoring en temps réel.

Enfin, il reste relativement **interprétable** grâce à l'importance des variables, ce qui est important dans un contexte bancaire.

Quelle métrique aurait été la plus intéressante ?

La fraude étant un problème **déséquilibré** (très peu de fraudes comparé aux transactions normales), l'accuracy n'est pas pertinente.

La métrique la plus intéressante est le **Recall** sur la classe fraude, car l'objectif principal est de **détecter un maximum de fraudes**.

Pour garder un bon équilibre avec les faux positifs (transactions normales bloquées)

Après quelques recherches, nous avons remarqué que la métrique **F1-score** est aussi très adaptée.

Celle-ci combine **la précision** et **le recall**, il permet de mesurer la performance globale du modèle en évitant de favoriser uniquement l'un des deux.

Sources :