1. Introduction et Contexte

Dans le contexte où les transactions bancaires numériques se multiplient, les banques font face à une augmentation continue des tentatives de fraude. Les fraudeurs adoptent des stratégies de plus en plus sophistiquées, rendant difficile la détection automatique par des règles fixes. C'est dans ce cadre que IBM pour notre Hackathon a fourni un dataset réel issu d'opérations bancaires effectuées entre 2016 et 2018, nous permettant de développer, d'entraîner et d'évaluer des modèles de détection de fraude.

Notre objectif est de concevoir un modèle d'Intelligence Artificielle capable de prédire si une transaction est frauduleuse (1) ou non frauduleuse (0), afin d'assister les équipes de sécurité bancaire dans la prise de décision en temps réel.

2. Problématique

Les fraudes sur les cartes bancaires représentent des pertes financières importantes pour les banques et leurs clients. La détection doit être rapide pour bloquer les transactions suspectes immédiatement, fiable pour éviter les fausses alertes qui pénalisent les clients et capable de généraliser car les fraudeurs changent continuellement de comportement.

Notre problématique est donc : Comment développer un modèle de Machine Learning capables d'identifier efficacement les transactions frauduleuses.

3. Description du dataset IBM

Les datasets à notre disposition sont les suivants:

transactions_train.csv dans lequel il y a toutes les transactions bancaires (montant, date lieu, carte utilisée, etc) ce sera notre base principale qui nous permettra de détecter des comportements inhabituels: montants anormaux, transactions dans d'autres villes etc.

train_fraud_labels.json qui indique si chaque transaction du train est frauduleurse (1) ou non (0), ce fichier contient la vérité terrain pour faire apprendre le modèle.

cards_data.csv dans lequel nous avons les informations techniques sur les cartes, ce qui nous aide à comprendre si la carte est à risque. ex: carte signalée sur le dark web qui est un indicateur fort de fraude.

user_data.csv qui nous fournit le profil des clients (âge, revenu, localisation, score, crédit, nombre de cartes), ce qui nous permet de comparer la transaction en fonction du profil du client par exemple revenu faible + gros achat = suspect.

mcc_codes.json qui nous fournit une classification des marchands (restauration, jeux en ligne, transport, etc) ce qui nous permet d'analyser les habitudes d'achat et repérer des achats dans des secteurs à risque (ex: casinos est un secteur avec plus de fraudes)

evaluation_features.csv nous donne les transactions sans label, à prédire pour la soumission. ce fichier ne sera pas utilisé pour notre entrainement, seulement pour faire le fichier de prédiction final.

4. Enjeux

Contrairement à de nombreux problèmes de classification, les clients présents dans l'ensemble d'entraînement ne sont pas les mêmes que ceux de l'ensemble d'évaluation
Cela signifie que notre modèle doit :
Généraliser sans mémoriser les comportements spécifiques de certains utilisateurs.

Se baser davantage sur des patterns de transactions, et non sur des identités.
Être résilient face à des comportements nouveaux.

5. Méthodologie