USE CASE : DETECTION DE FRAUDE PAR CARTE DE CREDIT

Transactions par carte de crédit anonymisées étiquetées comme frauduleuses ou authentiques

CONTEXTE :

La fraude par carte de crédit entraîne chaque année des pertes de milliards de dollars pour une société financière et ses clients. Les attaques et les fraudes augmentent à un rythme très rapide de nos jours. Il est important que les sociétés de cartes de crédit soient en mesure de reconnaître les transactions frauduleuses par carte de crédit afin que les clients ne soient pas facturés pour des articles qu'ils n'ont pas achetés.

Un système capable de classer les transactions en tant que frauduleuses est extrêmement important pour la société de cartes de crédit afin que ces transactions frauduleuses puissent être refusées. Cela permettra non seulement d'économiser de l'argent, mais aussi de donner confiance aux clients. Un autre point important à noter est que le système doit être très efficace pour ne pas classifier les transactions non frauduleuses comme frauduleuses.
Par exemple : si une transaction authentique est refusée, le client pourrait se frustrer et résilier son contrat.

OBJECTIF :

L'objectif de cette recherche est de répondre à la problématique **« Comment utiliser le machine learning pour prédire l'état (frauduleux ou non frauduleux) d'une carte de crédit bancaire ? ».**

CONTENU DU DATASET :

L'ensemble de données contient les transactions effectuées par cartes de crédit en septembre 2013 par des titulaires de cartes européens.
Cet ensemble de données présente les transactions qui se sont produites en deux jours, où nous avons 492 fraudes sur 284 807 transactions. L'ensemble de données est très déséquilibré, la classe positive (fraudes) représente 0,172 % de toutes les transactions.

Il ne contient que des variables d'entrée numériques qui sont le résultat d'une transformation PCA. Malheureusement, en raison de problèmes de confidentialité, nous ne pouvons pas fournir les caractéristiques originales et plus d'informations de base sur les données. Les caractéristiques **V1, V2, … V28** sont les principales composantes obtenues avec PCA, les seules caractéristiques qui n'ont pas été transformées avec PCA sont **'Time'** et **'Amount'**. La fonctionnalité **'Time'** contient les secondes écoulées entre chaque transaction et la première

transaction dans l'ensemble de données. La caractéristique '**Montant**' est le montant de la transaction, cette caractéristique peut être utilisée pour l'apprentissage sensible au coût en fonction de l'exemple. La caractéristique '**Class'** est la variable de réponse (cible) et elle prend la valeur 1 en cas de fraude et 0 sinon.

Compte tenu du rapport de déséquilibre de classe, nous recommandons de mesurer la précision à l'aide de l'aire sous la courbe de rappel de précision (AUPRC). La précision de la matrice de confusion n'est pas significative pour la classification déséquilibrée.

Source : https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

SKILLS :

On choisit un langage de programmation parmi : Python, R, Java, C ou autres.