Mamadou Issa NDIAYE | Machine Learning | November 18, 2023

Détection de Fraude par Carte de Crédit

Master 1 en Data Science & IA



Tables de matières

[Introduction 2](#_Toc151458188)

[Anatomie d’une fraude à la carte bancaire 3](#_Toc151458189)

# Introduction

Ce rapport sera divisé en de parties dont elles feront ressortir le travail effectué dessus et bien d’autres aspects permettant de mieux comprendre le projet.

Le Machine Learning étant de donner à un ordinateur ou une machine, la capacité d’apprendre sans programmer de manière explicite.

La détection de fraude par carte de crédit est une composante cruciale dans la sécurité des transactions financières. Avec l'essor des transactions électroniques, la nécessité de détecter et de prévenir la fraude par carte de crédit est devenue une priorité majeure pour les institutions financières et les entreprises.

La régression logistique, une technique de Machine Learning, est largement utilisée pour résoudre ce problème en raison de sa capacité à modéliser et prédire des événements binaires, tels que la fraude ou la non-fraude, en se basant sur des caractéristiques ou des variables explicatives.

Le processus de détection de fraude par carte de crédit en utilisant la régression logistique commence par la collecte de données historiques comportant des informations sur les transactions, telles que le montant de la transaction, la localisation, l'heure de la transaction, le type de commerçant, etc. Ces données servent de base à l'entraînement du modèle.

Ensuite, ces données sont nettoyées et préparées pour l'analyse. Cela peut inclure l'élimination des valeurs aberrantes, la normalisation des données, la gestion des valeurs manquantes, et la création ou l'ingénierie de nouvelles caractéristiques pertinentes pour améliorer la performance du modèle.

Une fois les données préparées, le modèle de régression logistique est entraîné à partir de ces données historiques étiquetées comme frauduleuses ou non. Le modèle cherche à établir une relation entre les caractéristiques des transactions et la probabilité qu'une transaction soit frauduleuse.

Lors de la phase de test, le modèle est évalué sur des données qu'il n'a pas vues auparavant. Il prédit la probabilité de fraude pour chaque nouvelle transaction et applique un seuil pour classer les transactions en frauduleuses ou non. Ce seuil est généralement ajusté pour trouver le bon équilibre entre la sensibilité (capacité à détecter la fraude réelle) et la spécificité (capacité à éviter les fausses alertes).

La régression logistique présente des avantages pour la détection de fraude par carte de crédit, notamment sa capacité à gérer des ensembles de données massifs, sa rapidité d'exécution et sa facilité d'interprétation des résultats. Cependant, il est important de noter qu'elle n'est pas la seule méthode utilisée pour cette tâche, et souvent, une approche combinée de plusieurs techniques de Machine Learning est adoptée pour améliorer la précision et la robustesse de la détection de fraude.

# Anatomie d’une fraude à la carte bancaire et Répartition des fraudes bancaires :

Cette image nous montre comment se fait la fraude par carte bancaire pour comprendre de manière explicite et concise.

