# Détection de Fraude sur Cartes Bancaires

## 1.Introduction

La fraude par carte bancaire constitue un enjeu critique pour les institutions financières, les fintechs et les systèmes de paiement en ligne, en raison des pertes économiques qu'elle génère et de son impact direct sur la confiance des clients. Avec l'augmentation continue du volume de paiements électroniques, les approches traditionnelles basées sur des règles métier statiques deviennent insuffisantes pour détecter des comportements frauduleux de plus en plus sophistiqués.

Le projet s'appuie sur le dataset « Credit Card Fraud Detection » mis à disposition sur Kaggle par la Machine Learning Group (MLG) de l'ULB en collaboration avec Worldline, qui contient des transactions bancaires réelles réalisées par des porteurs de cartes européens en septembre 2013. L'objectif est de concevoir un modèle de Machine Learning capable de prédire si une transaction est frauduleuse (classe 1) ou légitime (classe 0), en tenant compte d'un fort déséquilibre entre les classes.

### 2. Problématique et objectifs

La problématique centrale est la suivante : comment détecter automatiquement, en temps quasi réel, un très faible nombre de transactions frauduleuses au sein d'un volume massif de transactions normales, tout en limitant les fausses alertes pour ne pas dégrader l'expérience client.

Les objectifs opérationnels sont :

- Maximiser le rappel (recall) sur la classe fraude, afin de réduire au minimum le nombre de fraudes non détectées.
- Limiter les faux positifs, pour éviter de bloquer un trop grand nombre de transactions légitimes et conserver une bonne précision.
- Développer un modèle robuste, stable et suffisamment performant pour être déployé dans un environnement réel (banque, fintech, PSP, etc.).

## 3. Description du dataset

Le dataset comprend 284 807 transactions, dont 492 sont étiquetées comme frauduleuses, soit environ 0,172% du total, ce qui en fait un jeu de données extrêmement déséquilibré. Il comporte 31 variables au total : 30 variables explicatives et une variable cible binaire « Class » (0 = transaction normale, 1 = fraude).

Les principales variables sont :

Time : temps écoulé (en secondes) depuis la première transaction du jeu de données.

Amount : montant de la transaction.

V1 à V28 : variables issues d'une transformation de type PCA appliquée aux données brutes, afin d'anonymiser les informations sensibles tout en conservant la structure statistique.

Class : variable cible indiquant la nature frauduleuse ou non de la transaction.

Contraintes :

Les variables PCA sont anonymisées, ce qui complique l'interprétation économique ou métier des patterns identifiés.

La sévérité du déséquilibre de classes impose le recours à des techniques spécifiques de rééchantillonnage ou de pondération pour éviter qu'un modèle naïf n'apprenne à prédire uniquement la classe majoritaire.

## 4. Prétraitement des données

### 4.1 Nettoyage

Les transactions en double ont été supprimées afin d'éviter de biaiser l'apprentissage, et une vérification des valeurs manquantes a été réalisée, confirmant qu'elles sont absentes ou négligeables dans ce dataset. Ce nettoyage garantit la cohérence statistique des données utilisées pour entraîner les modèles.

### 4.2 Normalisation

Les variables Time et Amount ont été standardisées (centrage-réduction) afin de les mettre sur une échelle comparable et de faciliter l'optimisation des modèles sensibles à la mise à l'échelle, comme la régression logistique ou les méthodes à base de distances. Les variables V1 à V28, déjà issues d'une PCA, sont par construction centrées et réduites et ne nécessitent pas de normalisation supplémentaire.

### 4.3 Gestion du déséquilibre

Plusieurs approches ont été envisagées pour traiter le déséquilibre de classes :

Oversampling (SMOTE) : génération artificielle de nouvelles observations de la classe minoritaire, en interpolant entre des exemples réels proches, ce qui permet de rééquilibrer le dataset tout en conservant l'information existante.

Undersampling : réduction du nombre d'exemples de la classe majoritaire, simple mais potentiellement destructrice d'information utile.

Pondération des classes (class_weight) : affectation d'un poids plus élevé aux exemples frauduleux dans la fonction de perte, très adapté aux modèles linéaires et aux arbres de décision.

Dans ce projet, une combinaison SMOTE + class_weight="balanced" est recommandée afin de renforcer l'apprentissage sur la classe minoritaire tout en limitant le surapprentissage sur des données synthétiques.

## 5. Analyse exploratoire (EDA)

L'analyse exploratoire montre que la variable Amount est fortement concentrée sur de faibles montants, tandis que certaines fraudes apparaissent associées à des fourchettes de montants spécifiques, ce qui peut guider le choix de variables dérivées (binning, log-transformations, etc.). La variable Time révèle des cycles journaliers et des créneaux horaires où la densité de fraudes est plus élevée, suggérant des effets temporels exploitables pour la détection.

La visualisation de la distribution de la variable Class confirme le caractère extrêmement rare des fraudes et justifie l'usage de métriques adaptées telles que la courbe Precision–Recall. L'étude des corrélations met en évidence que certaines composantes (par exemple V12, V14, V17) présentent des distributions sensiblement différentes entre transactions frauduleuses et normales, ce qui indique leur pouvoir discriminant malgré l'absence d'interprétation métier directe.

## 6. Modélisation Machine Learning

Plusieurs familles de modèles supervisés de classification binaire ont été évaluées :

Régression logistique : utilisée comme baseline, elle offre une bonne interprétabilité globale et des performances correctes mais limitées sur des patterns non linéaires.

Random Forest : ensemble d'arbres de décision robustes, capable de modéliser des relations non linéaires et d'intégrer des poids de classes, généralement plus performant qu'un modèle linéaire sur ce type de données.

XGBoost / LightGBM : modèles de gradient boosting sur arbres, très efficaces pour exploiter des patterns rares dans des données déséquilibrées et fréquemment cités comme les meilleurs classifieurs sur des tâches de détection de fraude.

Dans ce contexte, XGBoost ou LightGBM ont généralement offert les meilleurs compromis entre rappel, précision et F1-score, grâce à leur capacité à optimiser directement des fonctions de perte adaptées aux données déséquilibrées et à leur flexibilité de réglage des hyperparamètres.

## 7. Évaluation des modèles

Dans un contexte de détection de fraude, l'accuracy globale n'est pas informative, car un modèle qui prédirait systématiquement « non fraude » serait très exact mais totalement inutile. Les métriques utilisées sont donc :

Recall (sensibilité) sur la classe fraude : métrique prioritaire, car elle mesure la proportion de fraudes correctement détectées.

Precision : proportion de fraudes réellement observées parmi les transactions prédites comme frauduleuses.

F1-score : moyenne harmonique entre précision et rappel, utile pour évaluer l'équilibre entre détection et fausses alertes.

Matrice de confusion : tableau permettant de visualiser le nombre de vrais positifs, faux négatifs, vrais négatifs et faux positifs.

ROC-AUC et surtout Precision–Recall AUC, plus pertinente lorsque la classe positive est rare.

Dans la plupart des études empiriques, XGBoost obtient un F1-score et une AUC supérieurs, tout en maintenant un rappel élevé, alors que la Random Forest se montre solide mais légèrement en retrait, et la régression logistique sert de référence minimale acceptable. Une matrice de confusion satisfaisante se

caractérise par un nombre très faible de fraudes non détectées (FN) et un niveau de faux positifs compatible avec les contraintes opérationnelles.

## 8. Interprétation et discussion

Les principaux atouts du modèle retenu sont :

Une très bonne capacité de détection des transactions frauduleuses grâce à la combinaison de techniques de rééchantillonnage et d'algorithmes d'ensemble performants.

Une robustesse face au déséquilibre des classes et une architecture modulaire (pipeline) facilement intégrable dans un système de scoring temps réel.

Les limites identifiées sont :

Le caractère anonymisé des variables PCA réduit l'explicabilité métier des décisions de classification.

Le dataset couvre une période courte (deux jours de transactions), ce qui ne reflète pas forcément la dynamique à long terme des comportements de fraude.

L'absence de données contextuelles (adresse IP, navigateur, device, géolocalisation, historique client) limite la capacité du modèle à capturer des signaux comportementaux riches.

Des pistes d'amélioration incluent :

L'utilisation d'outils d'explicabilité comme SHAP pour analyser la contribution de chaque variable aux décisions de fraude, en particulier pour les modèles de type XGBoost.

L'exploration de modèles de deep learning, tels que des autoencodeurs (détection d'anomalies) ou des architectures LSTM sur séries temporelles, qui ont montré de bons résultats sur ce dataset dans la littérature.

L'intégration de données temporelles plus longues et de variables comportementales, ainsi que l'ajustement dynamique du seuil de décision en fonction du niveau de risque ou du profil client.

## 9. Conclusion

La détection de fraude sur cartes bancaires est un problème emblématique de classification déséquilibrée, dans lequel la rareté des fraudes et les contraintes métier imposent de privilégier la sensibilité et la

précision sur la classe minoritaire plutôt que l'accuracy globale. En combinant un prétraitement adapté, une EDA ciblée et des modèles avancés tels que XGBoost ou LightGBM, il est possible de construire un système de Machine Learning capable d'identifier une grande proportion de transactions frauduleuses tout en maîtrisant les faux positifs.

Ce travail illustre la capacité des approches data-driven à renforcer la sécurité des paiements, à réduire les pertes financières et à améliorer la confiance des utilisateurs, à condition d'intégrer continuellement de nouvelles données, de surveiller les performances et d'expliquer les décisions aux acteurs métier et aux régulateurs