

# PROJET CLASSIFICATION

**Auteurs :**

**Bart, Philippe, Thibaut, Salah, Hadjer**

**23/02/2024**

# A. Préambule

## A.1. Contexte

Depuis des décennies l'utilisation des cartes de crédit pour réaliser de achats à augmenter de manière exponentielle. Le premier facteur fut la part de marché de plus en plus grande pour les achats en ligne. Ces derniers temps, ce fut une nouvelle accélération pour cause de pandémie. Les choses ont été rapides même trop rapide.

Vous travaillez au sein d'une société spécialisée dans la sécurité des systèmes bancaires. Vous disposez d'un simulateur de paiement Mobile Money. L'étude de cas de simulation de paiement Mobile Money est basé sur une entreprise réelle qui a développé une implémentation d'argent mobile qui offre aux utilisateurs de téléphones mobiles la possibilité de transférer de l'argent entre eux en utilisant leur téléphone comme un porte-monnaie électronique.

Une nouvelle vient d'arriver et les données sont dorénavant suffisantes pour mettre en place un système de prédiction des fraudes à la carte bancaire. La tâche à accomplir est de développer une approche qui détecte les activités suspectes qui sont révélatrices de fraude.

## A.2. Planification

1. Planification de la réunion de lancement ;
2. Rédaction d'une proposition de projet ;
3. Recontextualisation du projet : *A partir de notre base de données (fichier `credit_card_fraud.csv`), développer un modèle de prédiction qui permet de déterminer si une transaction est une fraude ou non ;*
4. Réalisation d'un MCD/MPD ;
5. Construction de la base de données ;
6. Dataviz
7. Choix, construction et évaluation de quatre modèles de classification : `XGBoost`, `AdaBoost`, `KNN`, `RandomForest` ;
8. Construction et réalisation de la maquette.

# Chapitre 1

## Documentation Technique

### 1.1 Base de données

#### 1.1.1 Nettoyage du fichier source

Le fichier ressource qui nous avaient été fournis (`credit_card_fraud.csv`) avait eu besoin d'un petit nettoyage :

- ◇ Les guillemets ont été supprimés ;
- ◇ Les virgules ont été remplacées par des points, pour que les nombres soient correctement interprétés par la `DataFrame` ;
- ◇ Les points-virgules ont été remplacés simplement par des virgules.
- ◇ Les attributs de types numériques, ont été convertis en `float` ou en `int`.
- ◇ Les valeurs des champs `nameOrig` et `nameDest`, furent converties en numérique avec le script :

```
1 def replace_first_letter(value):
2     if value.startswith('C'):
3         return '1' + value[1:]
4     elif value.startswith('M'):
5         return '2' + value[1:]
6     else:
7         return value
```

#### 1.1.2 MCD

La construction de la base de données a commencé par la réalisation du MCD, et par le passage au MPD avec le logiciel *Looping*. Une fois celui-ci terminé, nous avons utilisé le module `mysql.connector` pour établir une connexion, et pour la construction de notre base de données depuis Python.

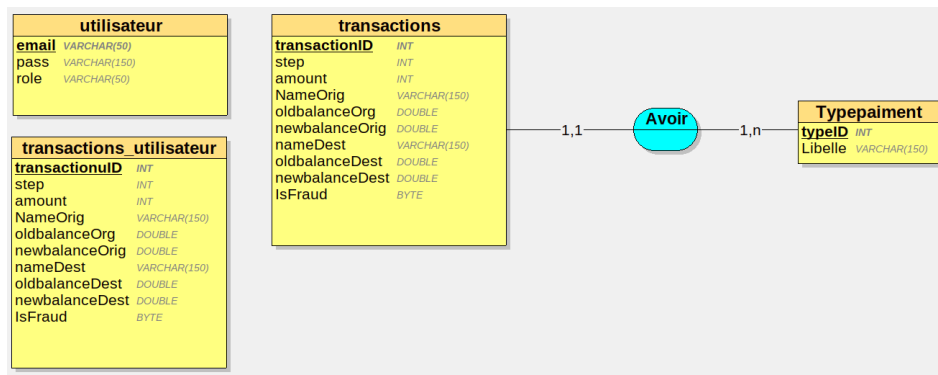


FIGURE 1.1 – MCD à partir du fichier credit\_card\_fraud.csv