

PROJET CLASSIFICATION

Prédire la **fraude** dans les services de
paiement financiers
(**CARTE DE CREDIT**)



Brief

Contexte

Depuis des décennies l'utilisation des cartes de crédit pour réaliser des 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.

Bien entendu, vous allez devoir gérer ce projet avec une méthodologie précise :

- Découpage des tâches
- Planification
- Présentation du POC
- Rédaction de rapports

Vous trouverez plus bas les spécifications techniques.

Cahier des charges

- Etude merise du projet et BDD Relationnelle
 - o Dictionnaire de données
 - o MCD
 - o BDD (MySQL ou SQLite)
- IA
 - o Plusieurs Modèles de classification
 - o Evaluations avec GridSearch
 - o Prédiction avec le meilleur modèle
- Interface
 - o Maquette
 - o Interface Graphique pour la prédiction

Spécifications techniques :

MERISE et BDD Relationnelle

1. Créer un MCD avec looping à partir des données
2. Produire le script SQL du MPD correspondant
3. Créer la BDD correspondante

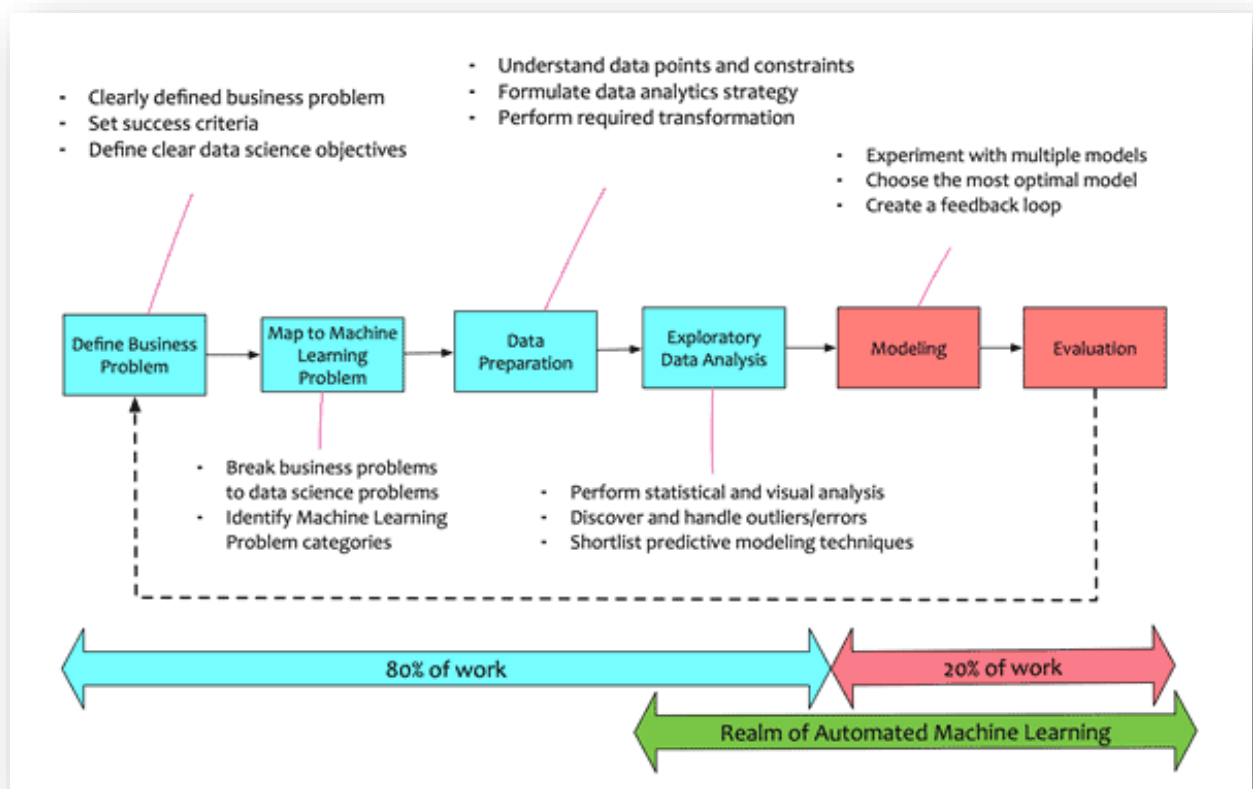
IA

1. Comprendre les données (variables)
2. Importer les données
3. Séparation des données en sous ensemble d'entraînement et un sous ensemble de test.
4. Construction du modèle
5. Entraînement du modèle avec le sous-ensemble d'entraînement
6. Prédications
7. Evaluation du modèle (e.g. matrice de confusion)
8. Résoudre le problème des données déséquilibrées (SMOTE)
9. Recommencer pour un nouveau modèle
10. Comparer entre 3 modèles pour choisir le meilleur le plus adapté à notre DataSet.

Interface

1. Maquette pour les opérateurs
 - a. Page d'accueil (login)
 - b. Page de chargement des données
 - c. Page de prédiction
2. Réalisation de l'interface
 - a. Technologie Python / Flask
 - b. Respecter la maquette définie

N.B. Respecter bien le cycle de vie d'un projet d'IA:



1 partie bonus

- Création une base de données (dans MySQL ou SQLite) pour contenir les données des transactions.
- Résoudre le problème du déséquilibre des données.
- Faire plusieurs modèles pour choisir le meilleur.

Gestion de projet

Vous allez devoir organiser une gestion de projet qui prévoit :

- Des réunions
- Des prises de décisions
- Une planification
- La rédaction d'un rapport
- La préparation d'une soutenance

Attention il sera important de désigner un chef de projet. Il aura pour rôle de s'assurer que les tâches sont bien distribuées et que l'avancement est assuré.

Tips

- Effectuez une réunion de lancement
 - Recontextualiser le projet
 - Prévoyez un plan d'action
 - Découper et planifier vos tâches
- Prévoyez des réunions régulières pour vérifier l'état d'avancement
- Mettez en place un outil pour vérifier l'état d'avancement du projet (Excel, Trello, Gant)

Livrable

Jeudi 09 février 2023

Rapport de projet :

1. Contexte
2. Planification
3. Guide utilisateur
4. Documentation technique
5. Difficultés rencontrées
6. Perspectives d'évolutions
7. Conclusion
8. Bilan de groupe
9. Bilans personnels

Vendredi 10 février 2023

Jour de la soutenance

Interface de détection de fraude :

- Il doit répondre aux spécifications techniques
- Être codé avec de bonnes pratiques (code clair et commenté)

Soutenance de Projet

- 20 minutes maximum de présentation avec Slides
 - Intro
 - Planification / Répartition des tâches
 - Démonstration
 - Difficultés rencontrées
 - Perspectives d'évolution
 - Conclusion
 - Bilan de groupe
 - **Bilan perso**
- 10 minutes de questions

Ressources

- Comprendre les données :

<https://www.kaggle.com/arjunjoshua/predicting-fraud-in-financial-payment-services/data>

- Pour aller plus loin, Learning from Imbalanced Classes:

[Learning from Imbalanced Classes - Silicon Valley Data Science \(svds.com\)](#)

- Choisir le bon modèle :

[ML 11 : MODELE ET HYPERPARAMETRE - YouTube](#)

- SMOTE pour la classification déséquilibrée avec Python:

[Comment traiter les problèmes de classification déséquilibrée en machine learning ? \[The "Imbalanced classification problem"\] | by Blue DME — Tech | Medium](#)

[SMOTE for Imbalanced Classification with Python \(machinelearningmastery.com\)](#)

[SMOTE using Python. Achieving class balance with few lines... | by Dr. Saptarsi Goswami | Towards Data Science](#)

Modalités d'évaluations

Passage par groupe de 4 (ou de 5)

Tous les membres du groupe doivent participer oralement

La soutenance doit être accompagnée d'une présentation visuelle

La soutenance ne doit pas dépasser 20 minutes

La présentation est claire et bien préparée

Les questions ne dépasseront pas 10 minutes

Le jury sera composé du formateur et du responsable pédagogique (il pourra être assisté par une personne extérieure à la formation)

Chaque participant intervient

Le rapport regroupe tous les éléments demandés (il est bien structuré)

La planification montre que chaque membre s'est vu attribuer une tâche

Objectifs visés d'apprentissage

Savoir réaliser une étude de merise

Savoir construire une IHM en respectant les standards de l'UI/UX

Savoir réaliser une application graphique

Savoir mettre en place un algorithme de classification

Savoir réaliser une prédiction avec un modèle de classification

Être capable de réaliser les CRUD avec SQL