

Mini Document de Cadrage

Plan de Travail Capstone

Projet: Prédiction du départ des clients dans le secteur bancaire à l'aide du Machine Learning

Réalisé par: Riché Fleurinord et Micka Louis

Date début: 10 octobre 2025

Soutenance: 27 octobre 2025

1. Contexte et problématique

Les banques dépensent beaucoup pour acquérir de nouveaux clients, mais fidéliser les clients existants est plus rentable. Identifier les clients susceptibles de partir permet de mettre en place des stratégies de fidélisation efficaces.

Problématique: Quelles caractéristiques influencent le plus la décision d'un client de quitter la banque et peut-on prédire ce départ à l'avance?

2. Objectifs

Objectif principal:

Construire un modèle capable de prédire si un client risque de quitter la banque afin d'aider le département marketing à anticiper les départs.

Objectifs spécifiques:

- Identifier les facteurs déterminants du churn (âge, solde, score, activité, etc.)
- Construire un modèle prédictif performant (Logistic Regression, Random Forest, XGBoost)
- Interpréter les résultats (importance des variables, SHAP values, confusion matrix)
- Proposer des actions concrètes de rétention client

3. Données

- **Source:** Kaggle
- **Nombre de clients:** 10 000
- **Variables:** démographiques (Géographie, Sexe, Âge), financières (Score, Solde, Ancienneté, Nombre de produits, Salaire estimé)
- **Variable cible:** Exited (indique si le client a quitté la banque)

4. Méthodologie

1. Exploration et nettoyage des données (EDA)
2. Feature engineering: encodage, normalisation, nouvelles features
3. Modélisation: Logistic Regression, Random Forest, XGBoost, éventuellement Neural Network
4. Optimisation des modèles: GridSearchCV et cross-validation
5. Interprétation et visualisation: importance des variables, SHAP plots, clusters à risque
6. Déploiement: application Streamlit pour prédiction du churn

5. Livrables attendus

- Notebook Jupyter complet et documenté
- Rapport PDF avec interprétation et recommandations
- Présentation PowerPoint ou Canva
- Présentation Power BI
- Application web Streamlit pour démonstration

6. Planning prévisionnel

Tâche	Début	Fin	Livrable attendu
Lancement et cadrage	12/10/2025	12/10/2025	Mini document de cadrage et repository GitHub initialisé
Exploration des données (EDA)	13/10/2025	14/10/2025	Notebook EDA complet
Préparation et Feature Engineering	15/10/2025	17/10/2025	Notebook de préparation et nouvelles features
Modélisation initiale	18/10/2025	19/10/2025	Rapport sur tests de modèles : Logistic Regression, Random Forest, XGBoost
Optimisation des modèles	20/10/2025	21/10/2025	Notebook du modèle optimisé avec hyperparamètres ajustés
Interprétation et storytelling	22/10/2025	23/10/2025	Ébauche du rapport final avec visualisations et recommandations
Rapport final	24/10/2025	25/10/2025	Rapport complet prêt à soumettre
Préparation de la présentation et répétition	26/10/2025	26/10/2025	Diaporama validé et répétition finale
Soutenance	27/10/2025	27/10/2025	Présentation orale et démonstration du projet

7. Outils et technologies

- **Langage:** Python
- **Bibliothèques:** Pandas, NumPy, Scikit-learn, XGBoost, LightGBM, Matplotlib, Seaborn, SHAP
- **Déploiement:** Streamlit
- **Environnement:** Jupyter Notebook