

Master en veille et intelligence compétitive
Spécialité : Gestion

**RAPPORT MINI PROJET 1 ANALYSE DU CHURN
CLIENT BANCAIRE**

Élaboré par

Allagui Chaima - Omra Mejri - Asma Amami - Marwa Werghi - Chaima Haidri

Professeur

Beldi Makrem

Année universitaire 2025/2026



TABLE DES MATIÈRES

LISTE DES FIGURES	iii
INTRODUCTION GÉNÉRALE	1
1 Contexte et problématique	2
1.1 Contexte général	3
1.2 Problématique	3
1.3 Objectifs du projet	3
2 Présentation des données	4
2.1 Source des données	5
2.2 Description des variables	5
2.3 Variable cible	5
3 Méthodologie	6
3.1 Démarche analytique	7
3.2 Préparation des données	7
3.3 Analyse exploratoire	7
4 Modélisation	9
4.1 Choix du modèle	10
4.2 Optimisation des hyperparamètres	10
4.3 Modèle final	10
5 Évaluation des performances et Interprétation des résultats	11
5.1 Indicateurs utilisés	12
5.2 Résultats	12
5.3 Matrice de confusion	12
5.4 Interprétation des résultats	12
5.4.1 Variables les plus importantes	12
5.4.2 Analyse métier	13

6	Recommandations	14
6.1	Recommandations stratégiques	15
6.2	Recommandations opérationnelles	15
6.3	Limites du projet	15
6.4	Perspectives	15
	CONCLUSION GÉNÉRALE	16



LISTE DES FIGURES

3.1	Les étapes d'un projet Data Science	7
3.2	Matrice de corrélation	8



INTRODUCTION GÉNÉRALE

Dans un contexte de forte concurrence bancaire, la fidélisation des clients constitue un enjeu stratégique majeur. Ce projet vise à analyser les facteurs expliquant le churn (attrition client) et à développer un modèle prédictif performant permettant d'identifier les clients à risque. L'étude s'appuie sur un jeu de données réel contenant plus de 10 000 clients bancaires.

Une démarche complète de data science a été suivie : exploration des données, préparation, modélisation et interprétation des résultats. Le modèle Random Forest optimisé par GridSearchCV a atteint une performance élevée avec un AUC de 0,986 sur l'échantillon test. Les résultats montrent que l'activité transactionnelle, le taux d'utilisation du crédit et le comportement d'inactivité sont des facteurs clés du churn.

Des recommandations opérationnelles sont proposées afin d'aider les établissements bancaires à anticiper le départ des clients et à mettre en place des stratégies de rétention ciblées.

Contexte et problématique

Sommaire

1.1	Contexte général	3
1.2	Problématique	3
1.3	Objectifs du projet	3

1.1 Contexte général

Le secteur bancaire connaît une transformation digitale accélérée, marquée par l'émergence de nouvelles banques digitales et fintechs. Dans ce contexte, les clients disposent de nombreuses alternatives, ce qui accroît le risque d'attrition.

Le churn client représente une perte directe de revenus et un coût élevé lié à l'acquisition de nouveaux clients. Identifier précocement les clients susceptibles de quitter la banque devient donc un levier stratégique.

1.2 Problématique

Comment exploiter les données clients afin de :

- Comprendre les facteurs explicatifs du churn bancaire ?
- Prédire efficacement les clients à risque de départ ?
- Aider à la prise de décision via des recommandations basées sur les données ?

1.3 Objectifs du projet

- Analyser les caractéristiques comportementales et socio-démographiques des clients
- Construire un modèle de classification performant
- valuer la performance du modèle à l'aide de métriques adaptées
- Identifier les variables les plus influentes dans le churn
- Proposer des recommandations business exploitables

Présentation des données

Sommaire

2.1	Source des données	5
2.2	Description des variables	5
2.3	Variable cible	5

2.1 Source des données

Le jeu de données utilisé est BankChurners.csv, contenant 10 127 observations et 23 variables décrivant des clients bancaires.

2.2 Description des variables

Les variables se répartissent en plusieurs catégories :

- Variables démographiques : âge, genre, situation familiale, niveau d'éducation
- Variables comportementales : nombre de transactions, montant total des transactions, mois d'inactivité
- Variables financières : limite de crédit, solde revolving, taux d'utilisation.
- Variable cible : Attrition_Flag (client actif ou attrité)

2.3 Variable cible

Une variable binaire Churn a été créée :

- **1** : client attrité
- **0** : client existant

La distribution montre un déséquilibre modéré avec environ 16 % de churners.

Méthodologie

Sommaire

3.1	Démarche analytique	7
3.2	Préparation des données	7
3.3	Analyse exploratoire	7

3.1 Démarche analytique

La méthodologie adoptée suit les étapes standards d'un projet de data science :

- Compréhension du besoin métier
- Exploration et nettoyage des données
- Feature engineering
- Modélisation
- valuation et interprétation
- Recommandations

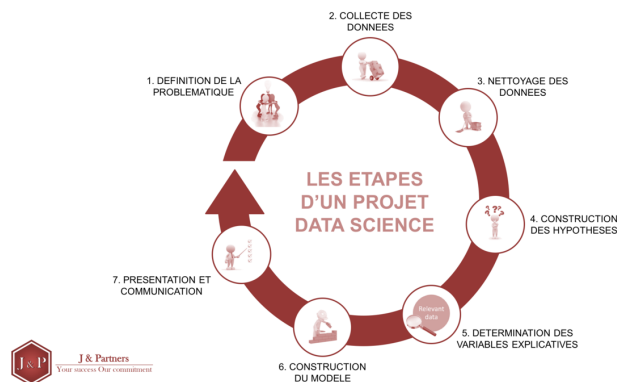


FIGURE 3.1 – Les étapes d'un projet Data Science

3.2 Préparation des données

- Suppression des variables non pertinentes (CLIENTNUM, variables Naive Bayes)
- des variables catégorielles via Label Encoding
- Standardisation des variables numériques avec StandardScaler
- Séparation train/test (70 % / 30 %) avec stratification

3.3 Analyse exploratoire

Des visualisations ont permis d'identifier des différences significatives entre churners et non-churners, notamment au niveau du solde revolving et de l'activité transactionnelle. Une analyse de corrélation a été menée afin de détecter d'éventuelles multicollinéarités.

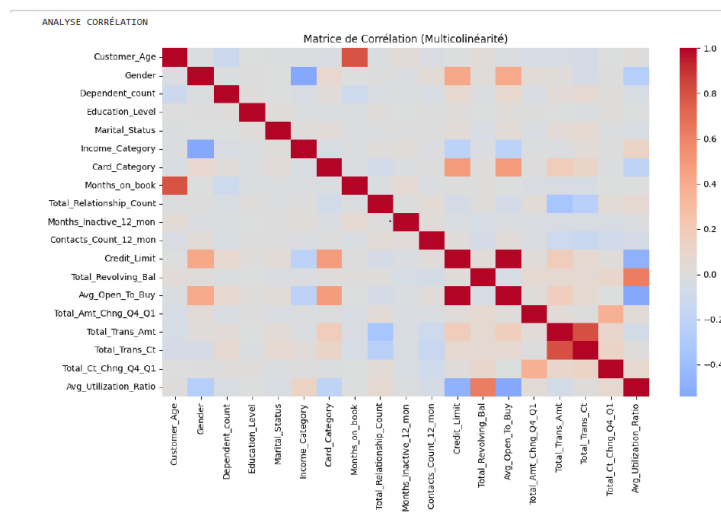


FIGURE 3.2 – Matrice de corrélation

Modélisation

Sommaire

4.1	Choix du modèle	10
4.2	Optimisation des hyperparamètres	10
4.3	Modèle final	10

4.1 Choix du modèle

Le modèle Random Forest Classifier a été retenu pour :

- Sa robustesse face au bruit
- Sa capacité à gérer les relations non linéaires
- Son interprétabilité via l'importance des variables

4.2 Optimisation des hyperparamètres

Une optimisation par GridSearchCV a été réalisée sur :

- Nombre d'arbres
- Profondeur maximale
- Taille minimale des nœuds

La métrique d'optimisation retenue est l'AUC.

4.3 Modèle final

Les meilleurs paramètres obtenus permettent d'atteindre une excellente généralisation sur l'échantillon test.

Évaluation des performances et Interprétation des résultats

Sommaire

5.1	Indicateurs utilisés	12
5.2	Résultats	12
5.3	Matrice de confusion	12
5.4	Interprétation des résultats	12
5.4.1	Variables les plus importantes	12
5.4.2	Analyse métier	13

5.1 Indicateurs utilisés

- Accuracy
- Precision
- Recall
- F1-score
- AUC-ROC
- Matrice de confusion

5.2 Résultats

Le modèle obtient :

- AUC test : 0,986
- Accuracy : 96 % La capacité du modèle à identifier les churners est jugée très satisfaisante.

5.3 Matrice de confusion

La matrice de confusion montre un faible taux de faux négatifs, ce qui est crucial dans un contexte de rétention client.

5.4 Interprétation des résultats

5.4.1 Variables les plus importantes

Les principales variables explicatives du churn sont :

- Total_Trans_Ct
- Total_Trans_Amt
- Avg_Utilization_Ratio
- Total_Revolving_Bal
- Months_Inactive_12_mon

5.4.2 Analyse métier

Les clients à faible activité transactionnelle et à forte utilisation de crédit présentent un risque élevé d'attrition. L'inactivité prolongée constitue également un signal d'alerte fort.

Recommandations

Sommaire

6.1	Recommandations stratégiques	15
6.2	Recommandations opérationnelles	15
6.3	Limites du projet	15
6.4	Perspectives	15

6.1 Recommandations stratégiques

- Mettre en place des alertes churn basées sur le score du modèle
- Segmenter les clients à risque pour des campagnes ciblées
- Proposer des offres personnalisées aux clients inactifs

6.2 Recommandations opérationnelles

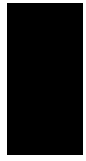
- Renforcer le suivi des clients à forte utilisation du crédit
- Encourager l'usage des services bancaires digitaux
- Mettre en place des programmes de fidélisation

6.3 Limites du projet

- Données statiques (absence de dimension temporelle)
- Déséquilibre des classes
- Absence de variables comportementales digitales

6.4 Perspectives

- Intégration de modèles avancés (XGBoost, LightGBM)
- Analyse temporelle du churn
- Déploiement du modèle via un dashboard décisionnel



CONCLUSION GÉNÉRALE

Ce projet démontre l'apport de l'analytique avancée dans la compréhension et la prédiction du churn bancaire. Les résultats obtenus confirment la pertinence des techniques de machine learning pour soutenir la prise de décision stratégique et améliorer la rétention client.