

COMPTE RENDU

ikram azizi

January 2026

1 Introduction

Analyse Prédictive des Campagnes de Marketing Bancaire Projet Machine Learning January 1, 2026

2 Contexte du projet

Dans le secteur bancaire, les campagnes de marketing direct représentent un coût important. Contacter tous les clients sans distinction entraîne une faible rentabilité. Ce projet vise à optimiser le ciblage des clients afin d'augmenter le taux de souscription aux dépôts à terme tout en réduisant les coûts marketing.

3 Objectif

L'objectif principal est de construire un modèle de **Machine Learning** capable de prédire la probabilité qu'un client souscrive à un dépôt à terme, en s'appuyant sur des données historiques issues de campagnes de marketing direct.

4 Dataset

[leftmargin=1.5cm]

- **Nom :** Bank Marketing Dataset
- **Taille :** Environ 45 000 observations
- **Type de problème :** Classification binaire
- **Variable cible :** Souscription à un dépôt à terme (yes / no)

4.1 Principales variables

- Informations client : âge, emploi, situation familiale, niveau d'éducation
- Données bancaires : solde du compte, prêts en cours
- Données de campagne : nombre de contacts, campagnes précédentes, résultats antérieurs

5 Pipeline de traitement

Le pipeline de traitement des données comprend les étapes suivantes :

1. Nettoyage des données et gestion des valeurs manquantes (*unknown*)
2. Encodage des variables catégorielles via le One-Hot Encoding
3. Suppression de la variable *duration* afin d'éviter le **data leakage**
4. Gestion du déséquilibre des classes avec la pondération des classes
5. Séparation des données en ensembles d'entraînement et de test

6 Modèle utilisé

Le modèle choisi pour ce projet est un **Random Forest Classifier**. Il est intégré dans un pipeline Scikit-learn incluant :

- Le prétraitement des données
- L'entraînement du modèle
- La prédiction sur les données de test

7 Évaluation du modèle

Les performances du modèle sont évaluées à l'aide de métriques adaptées aux jeux de données déséquilibrés :

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

8 Résultats clés

Les résultats montrent que le modèle permet :

- Un meilleur ciblage des clients à forte probabilité de souscription
- Une réduction des contacts marketing inutiles
- Une amélioration du retour sur investissement

9 Perspectives

Plusieurs améliorations peuvent être envisagées :

- Ajustement du seuil de décision selon les objectifs business
- Test d'autres modèles comme XGBoost ou la régression logistique
- Déploiement du modèle dans un environnement réel

10 Technologies utilisées

- Python
- Pandas, NumPy
- Scikit-learn
- Matplotlib, Seaborn

Projet réalisé dans un objectif d'analyse prédictive et d'optimisation marketing.