

International Data Science Institute -INPHB

Data Science - Sécurité Projet ML

2 Semaines

2025-2026

INSTRUCTIONS

- *Toujours justifier vos résultats*
 - *Documents non autorisés*
-
-

1 Objectif général du projet

Développer une chaîne de traitement complète en Machine Learning pour :

- modéliser le montant du FDE et la probabilité de non-paiement ou de résiliation d'un abonné,
- segmenter les clients en groupes homogènes selon leur comportement,
- sélectionner les variables pertinentes pour expliquer les comportements,
- valider rigoureusement les performances des modèles proposés,
- interpréter les résultats et formuler des recommandations opérationnelles.

2 Jeu de données

- Environ 23 millions de lignes, 46 variables.
- Variables quantitatives : CUBCONS, MONT-FDE, MONT-TTC, DELAI_REGL, etc.
- Variables qualitatives : CATEGORIE, DR, NOUVEAU, RETARD, RESILIE, etc.
- Objectifs à prédire : RETARD, RESILIE, MONT-FDE.

3 Tâches à réaliser

Exploration des données

- Décrire la distribution des variables quantitatives.
- Étudier les relations entre CUBFAC, CUBCONS, MONT-FDE, etc.
- Visualiser les comportements de paiement par zone et par type.
- Vérifier les corrélations.

Régression (étude comparative : linéaire, Ridge, Lasso, PLS, PCR)

- Modéliser MONT-FDE.
- Comparer les modèles : MCO, Ridge, Lasso, Elastic Net, PCR/PLS.
- Utiliser la validation croisée.
- Évaluer les performances : MSE, R^2 , biais/variance.

Classification (logistique, arbres, RF, boosting, k-NN)

- Prédire RETARD ou RESILIE.
- Comparer les approches : logistique, CART, Random Forest, Boosting, k-NN.
- Construire la matrice de confusion.
- Évaluer avec Accuracy, F1, AUC, etc.

Sélection de variables

- Comparer Ridge, Lasso, Random Forest (importance des variables), méthode Out-Of-Bag.
- Analyser les variables explicatives retenues.

4 Livrables attendus

- Rapport de projet (15 pages maximum).
- Présentation (slides).
- Code Python/R propre et commenté.
- Visualisations (graphiques, matrices, etc.).
- Tableau d'évaluation des performances.