
DEFAULT PROBABILITY PREDICTION

Professeur

Dr Imane RAHIL

Etudiants

HOUNSAVI Jules Koffi

TARDANE Hafsa



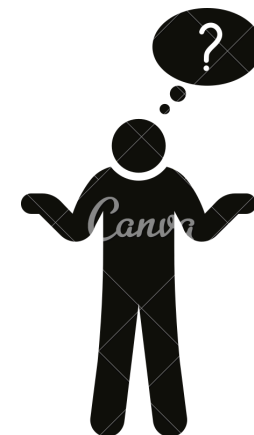
SOMMAIRE

- ① INTRODUCTION
- ② OBJECTIFS
- ③ METHODOLOGIE
- ④ DEFINITIONS
- ⑤ RESULTAT
- ⑥ DEMO
- ⑦ CONCLUSION
- ⑧ Q/A

1-INTRODUCTION

Les défauts de paiement constituent un enjeu crucial pour les institutions financières, entraînant des pertes significatives et compromettant la stabilité économique. Identifier et anticiper ces risques est essentiel pour une gestion proactive et durable des portefeuilles de crédit.

Cela soulève une question clé: comment prédire si un individu est susceptible de faire défaut sur ses obligations financières en s'appuyant sur ses données historiques?



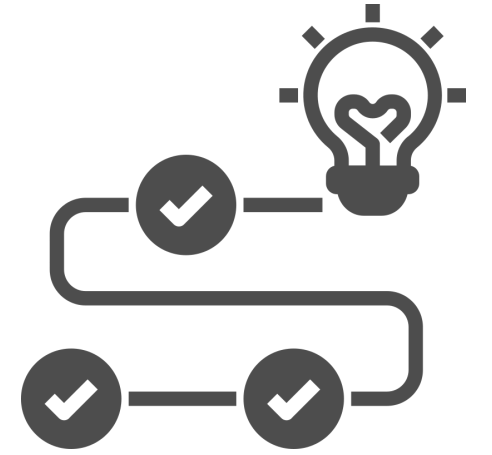
2-OBJECTIFS

Développer un modèle prédictif capable d'estimer la probabilité qu'un individu fasse défaut sur ses obligations financières dans les deux prochaines années, en utilisant des données historiques issues du dataset *Give Me Some Credit* de Kaggle.



4-METHODOLOGIE

Pour atteindre l'objectif principal, plusieurs étapes méthodologiques ont été suivies :



-Explorer les données : analyser les tendances, identifier les anomalies et étudier les relations entre les variables.

-Préparer les données : gérer les valeurs manquantes, les valeurs redondantes et les valeurs aberrantes.

-Tester plusieurs algorithmes : comparer la performance de modèles tels que la régression logistique, Random Forest et XGBoost.

-Évaluer la performance : utiliser des métriques (AUC-ROC, précision, rappel) pour sélectionner le meilleur modèle.

-Développer une application web : créer une interface utilisateur intuitive permettant aux utilisateurs de soumettre des données et de recevoir des prédictions en temps réel.

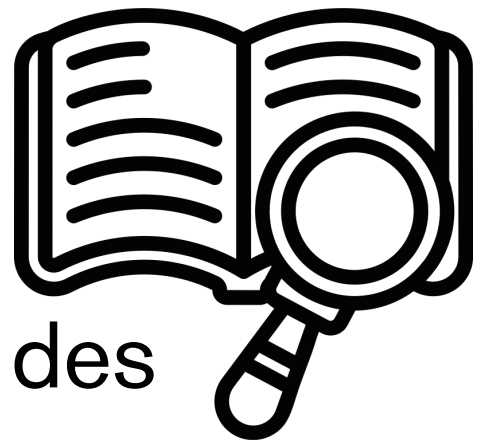
3-DEFINITIONS

Voici quelques définitions des termes techniques que nous verrons au cours de ce projet.

-**Machine Learning**: branche de l'intelligence artificielle permettant aux ordinateurs d'apprendre et de faire des prédictions sans être programmés explicitement.

-**Kaggle**: plateforme en ligne pour data scientists, offrant des datasets, des compétitions et des ressources pour le machine learning.

-**Flask**: framework web léger et minimaliste de Python utilisé pour développer des applications web et des APIs de manière simple et flexible.





-**API** (Application Programming Interface): interface permettant à différentes applications de communiquer entre elles, souvent utilisée pour intégrer des fonctionnalités externes dans un projet.

-**Régression Logistique**: modèle de machine learning utilisé pour des problèmes de classification binaire, prédisant la probabilité qu'une observation appartienne à une classe donnée.

-**Random Forest**: modèle de machine learning qui construit plusieurs arbres de décision pour prédire une valeur, Il combine les résultats de ces arbres pour améliorer la précision et réduire le sur-apprentissage.

-**XGBoost** (Extreme Gradient Boosting): algorithme de machine learning basé sur le gradient boosting, performant pour les tâches de classification et de régression, utilisé pour des données complexes.

5-RÉSULTAT

L'objectif de ce projet étant d'estimer la probabilité qu'un individu soit en défaut de paiement en fonction de ses caractéristiques financières et personnelles, nous avons testé trois modèles de classification: la régression logistique, Random Forest et XGBoost.

Le modèle XGBoost a surclassé les autres en obtenant un score AUC-ROC de **0.8596**, ce qui témoigne de sa meilleure capacité à séparer les classes par rapport aux autres modèles.



Évaluation des performances du modèle XGBoost :

-Accuracy : 0.9364

-Precision : 0.5766

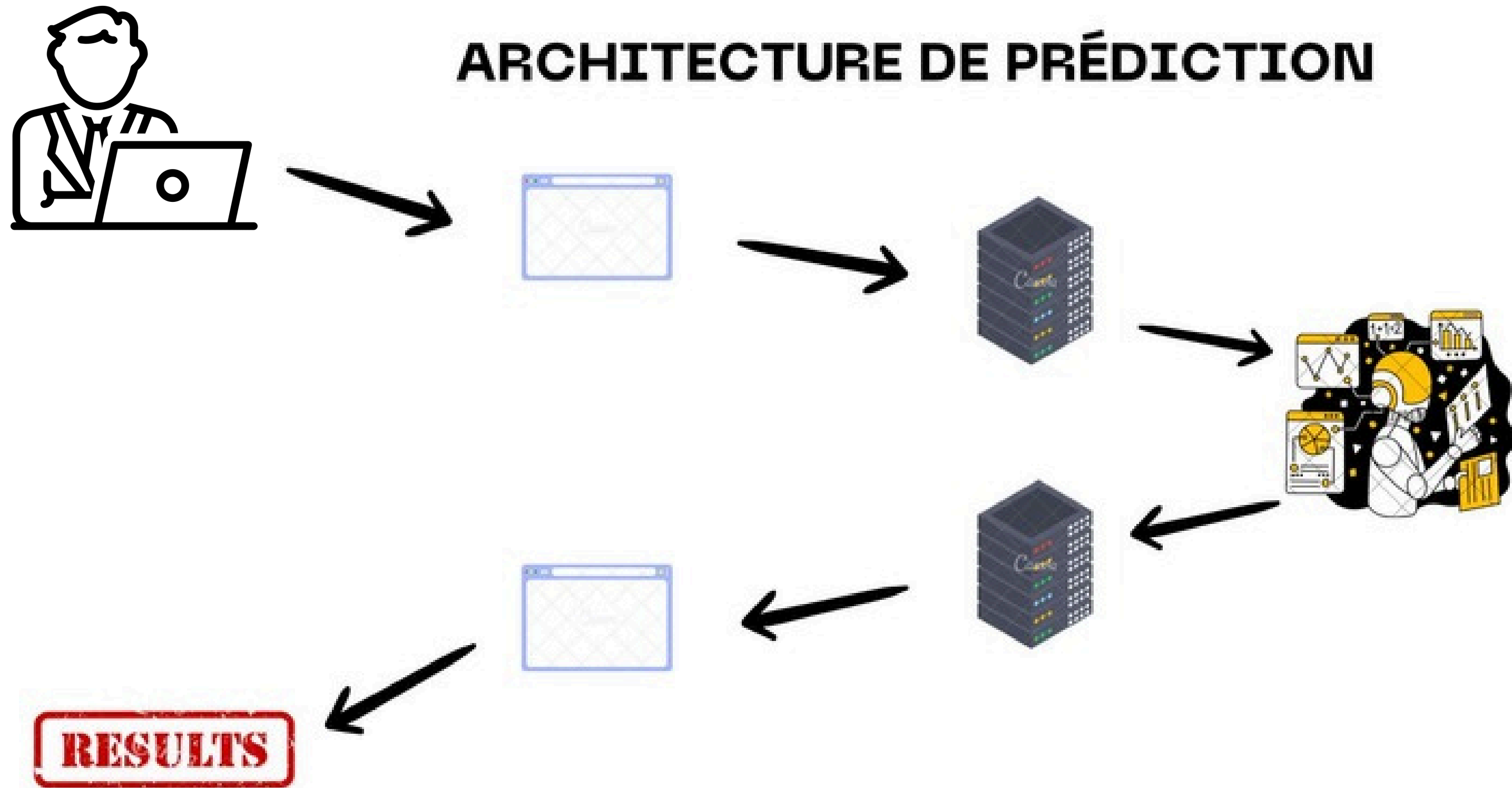
-Recall : 0.1790

-F1 Score : 0.2732

-ROC-AUC Score : 0.8596

Les résultats indiquent que bien que l'accuracy soit élevée, le modèle a rencontré des difficultés à identifier les clients à risque, ce qui est reflété par le faible rappel et le score F1. Toutefois, le modèle XGBoost reste le plus performant parmi ceux testés et sera retenu pour la suite du projet.

6-DEMO



7-CONCLUSION

En conclusion, ce projet a permis de développer un modèle prédictif visant à estimer le risque de défaut de paiement en utilisant des modèles de machine learning. Les résultats montrent qu'un modèle performant pourrait permettre d'identifier plus efficacement les clients susceptibles de faire défaut, aidant ainsi les banques et autres prêteurs à mieux gérer les risques associés à leurs portefeuilles de crédit.

Pour l'avenir, plusieurs pistes d'amélioration sont envisageables, telles que l'intégration de réseaux neuronaux ou l'optimisation des paramètres des modèles existants, afin d'améliorer encore les résultats obtenus.

Merici

A background network diagram consisting of numerous gray circular nodes of varying sizes connected by thin, light gray lines. The nodes are distributed across the entire frame, creating a complex, interconnected web-like structure.

QUESTIONS-RÉPONSES