



# **IMPLEMENTEZ UN MODELEDE SCORING**


**SAIDALI BACAR ABDALLAH  
02-2024**

**PROJECT-7 OPENCLASSROOMS**





# CONTENT

- 
- |           |   |
|-----------|---|
| <b>01</b> | INTRODUCTION                                  |
| <b>02</b> | ANALYSE EXPLORATOIRE                          |
| <b>03</b> | MODELISATION                                  |
| <b>04</b> | INTERPRETATION                                |
| <b>05</b> | DASHBOARD                                     |
| <b>06</b> | LES LIMITES ET LES<br>AMELIORATIONS POSSIBLES |
| <b>07</b> | CONCLUSION                                    |

# I. INTRODUCTION

## Le contexte



Nous intervenons en tant que data scientist sur un projet proposé par une société financière qui s'appelle "Prêt à dépenser". Elle propose des crédits à des clients ayant peu ou pas d'historique de prêt.

## Les objectifs



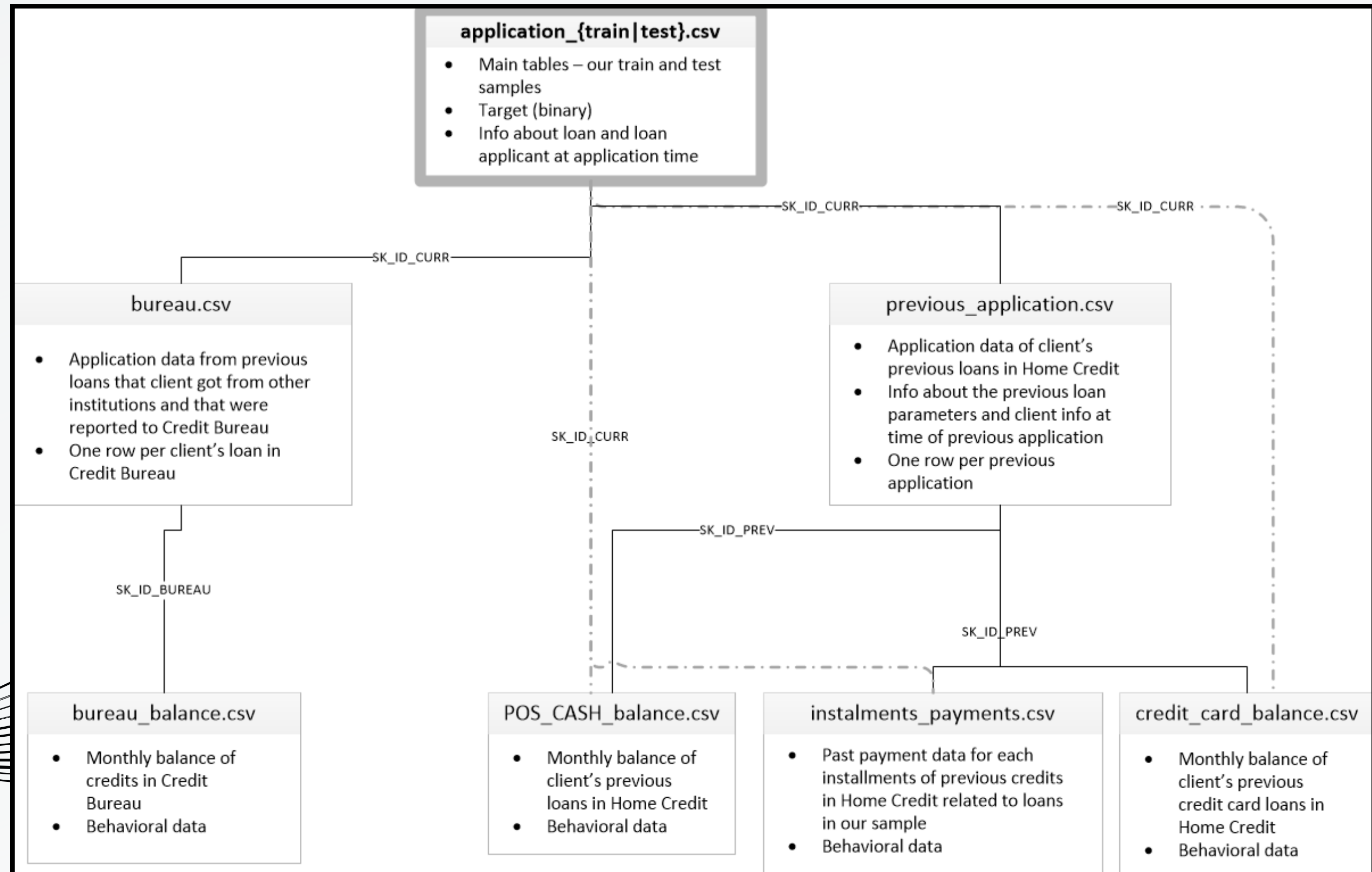
- mettre en place un outil de scoring qui capable de prédire la probabilité qu'un client puisse rembourser son crédit
- mettre en place un dashboard interactif pour faciliter la relation banque - client et gagner en transparence.



# II. ANALYSE EXPLORATOIRE

Les données: 7 fichiers

- type CSV
- environ 200 variables
- une variable "TARGET"



# II. ANALYSE EXPLORATOIRE

## Le process

### Preprocessing

- traitement individuel des fichiers
- création de nouvelles variables
- jointure de tous les fichiers

xx lignes  
xx variables

1.KERNEL  
KAGGLE

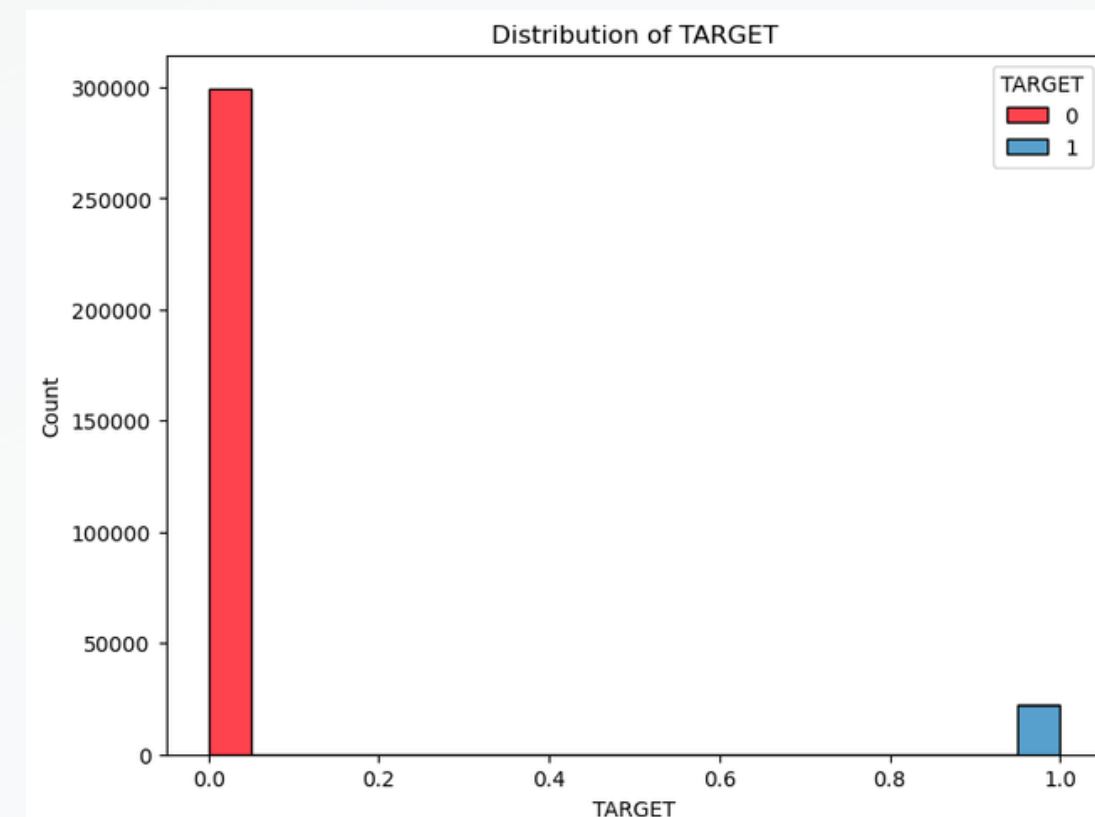
- Traitement des valeurs manquantes
- Suppression des variables corrélés entre elle
- Imputation de valeurs manquantes - méthode iterative

xx lignes  
xx variables

NETTOYAGE DES  
DONNEES

- Analyse univariée
- Analyse bivariée

## ANALYSE





# III. MODELISATION

## La problématique

Classification sur des données qui souffrent d'un grand déséquilibre de classe  
Il faut prendre en compte que les faux négatifs sont 10 fois plus coûteux que les faux positifs.  
On souhaite pas accorder un crédit à un client qui n'a pas la capacité de rembourser.

01

LA FONCTION COUT  
METIER

La fonction cout métier:  
Un faux négatif est 10 fois plus  
coûteux qu'un faux positif  
Implémenter dans la matrice de  
confusion.  
Paramètre: custom\_scorer

02

SPLIT DES  
DONNEES

60% entraînement  
20% validation  
20% test

03

DESEQUILIBRE

Librairie SMOTE  
Création de nouveaux individus à  
partir de la classe minoritaire

04

EVALUATION DES  
MODELES

Mise en place d'un pipeline et  
d'un GridSearch CV qui  
comprend les modèles suivant:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- LightGBM Classifier

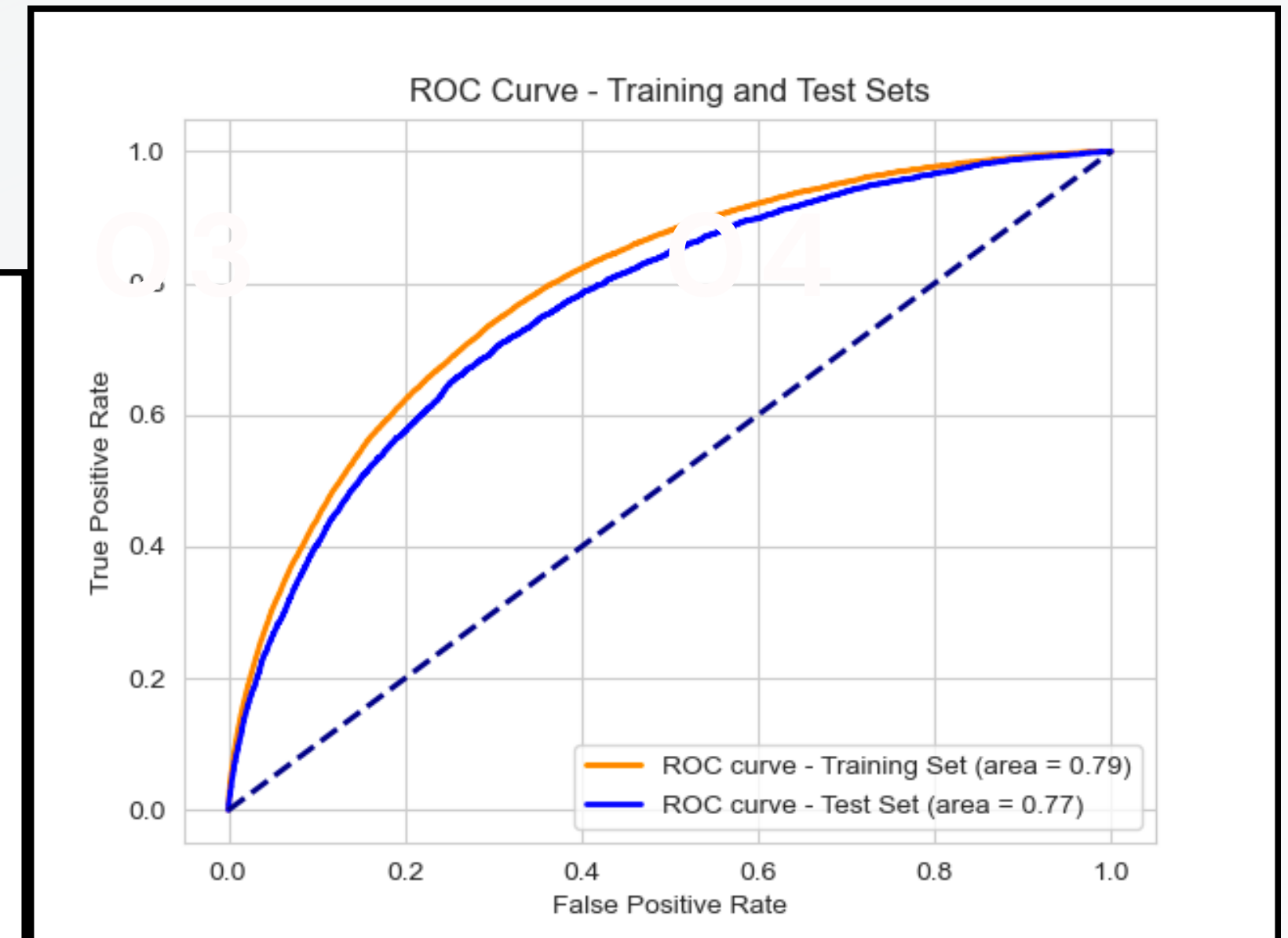
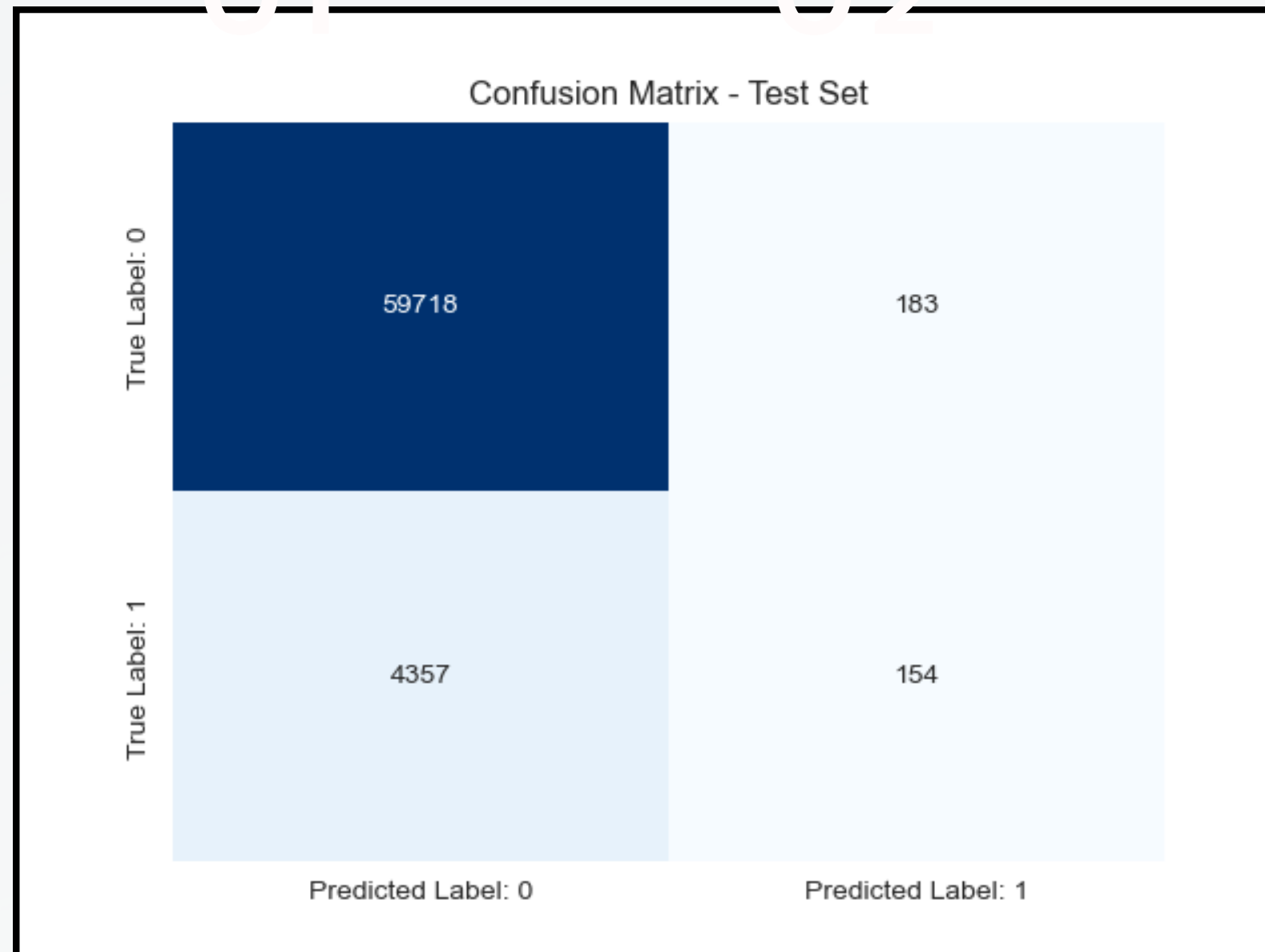
# III. MODELISATION

## La problématique

Classification sur des données qui souffrent d'un grand déséquilibre de classe  
Il faut prendre en compte que les faux négatifs sont 10 fois plus coûteux que les faux positifs.  
On souhaite pas accorder un crédit à un client qui n'a pas la capacité de rembourser.

## Matrice de confusion

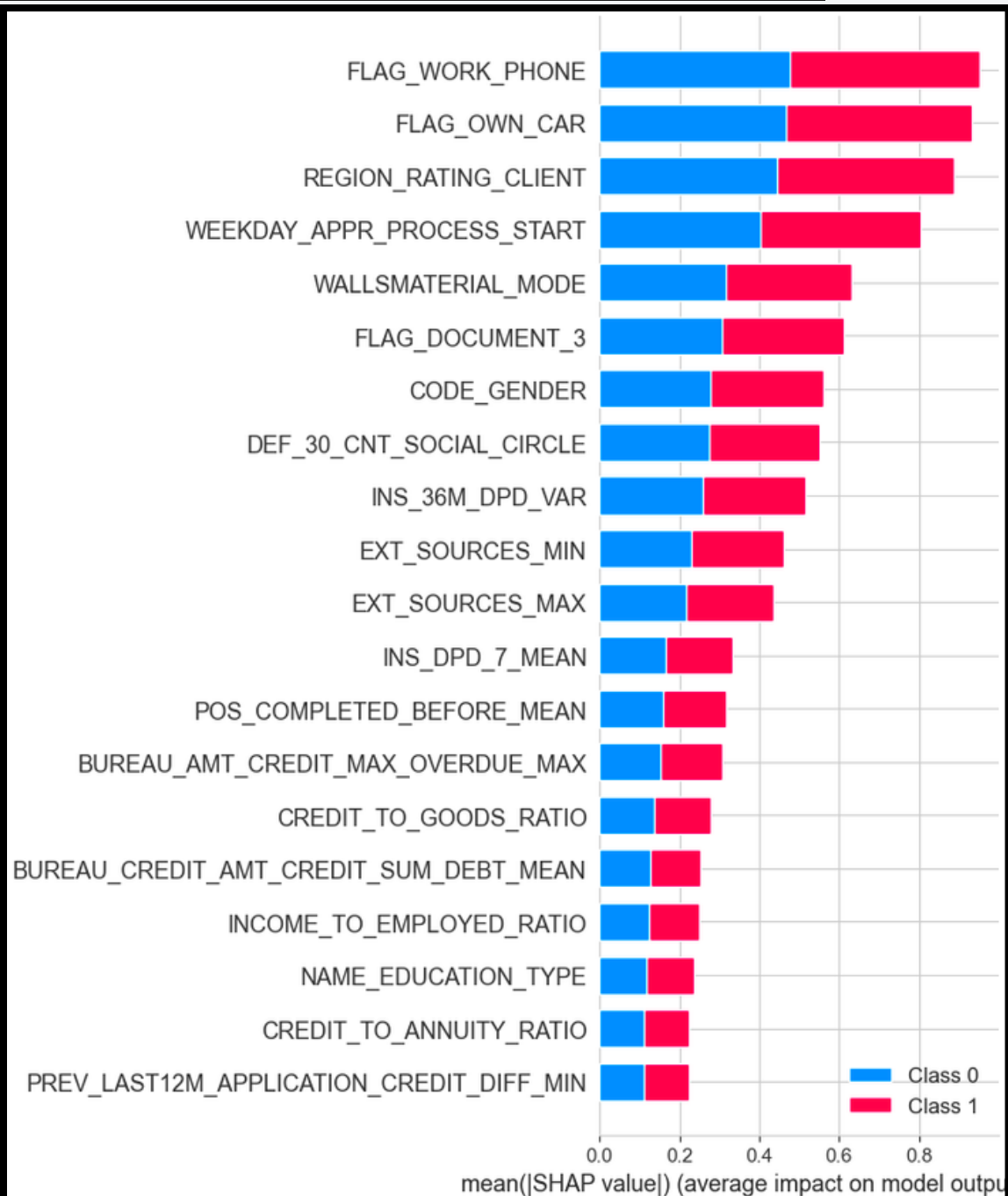
### ROC-AUC



Faire les graphiques  
custom\_scorer  
roc\_auc

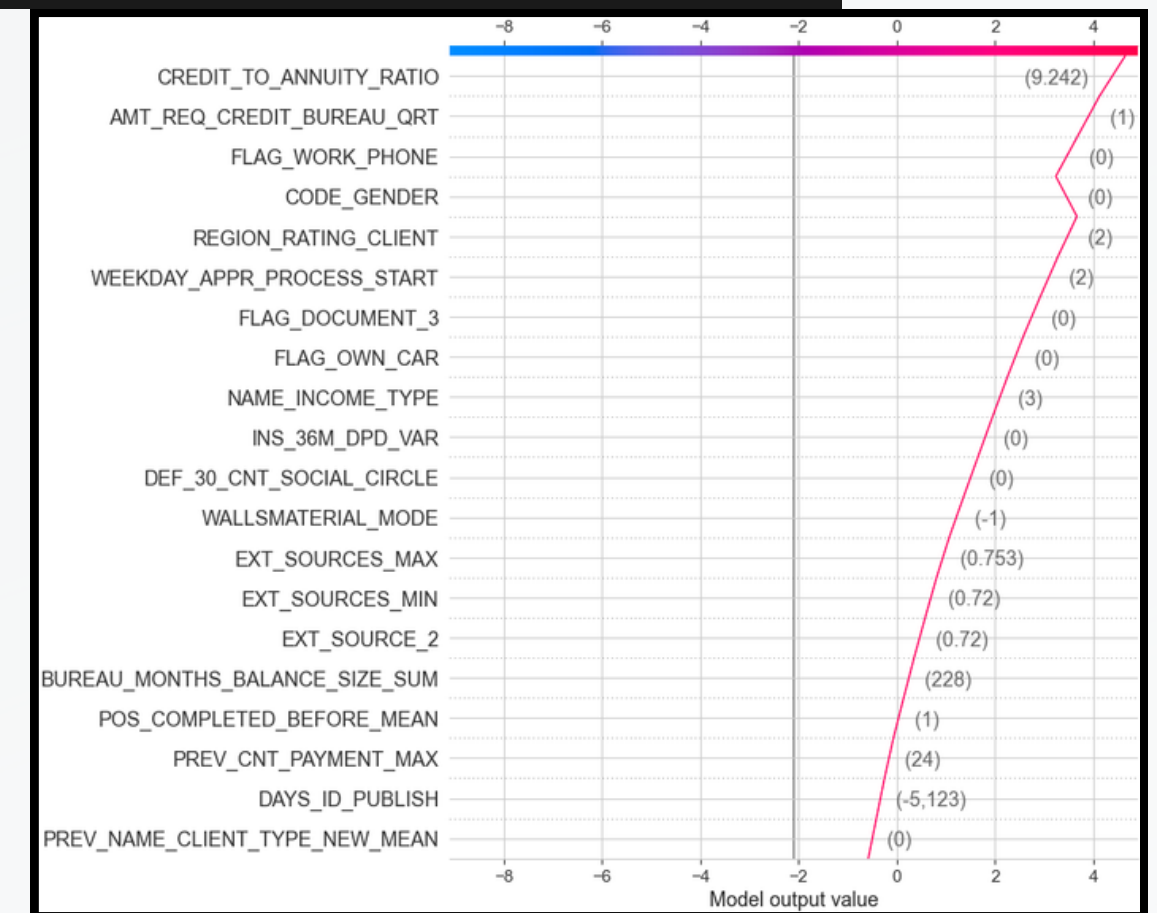
# IV. INTERPRETATION

## SHAP Global

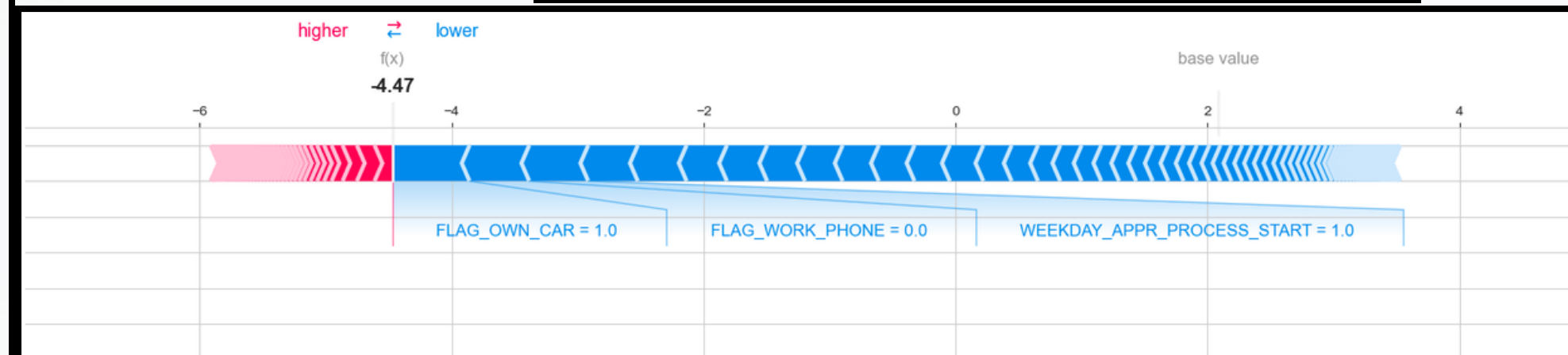


Le principe: SHAP calcule l'importance d'un variable en comparant ce qu'un modèle prédit avec ou sans cette variable. L'ordre dans le lequel le modèle voit les variables peut affecter sa prédiction, d'où le fait de tester dans les variables dans tous les sens.

## SHAP Local



- Les graphiques
- summary plot
  - decision plot
  - force plot

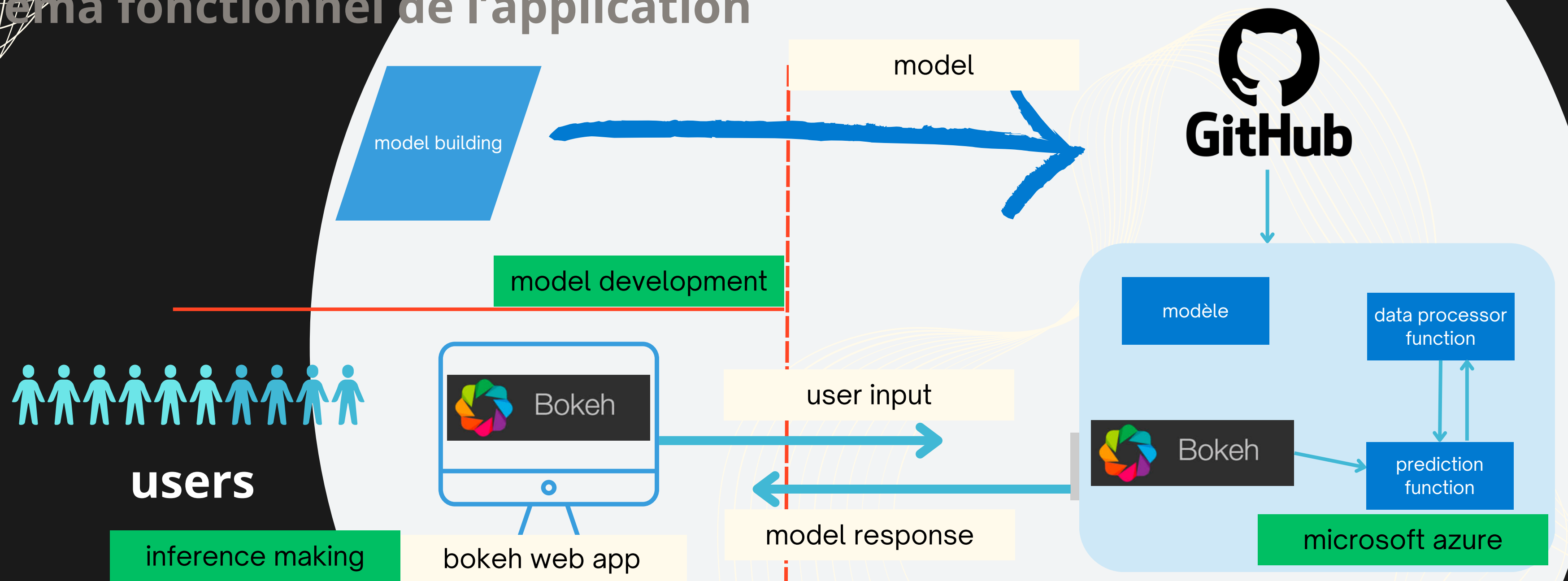




# V. DASHBOARD



## Schéma fonctionnel de l'application



# VI. LES LIMITES ET LES AMÉLIORATIONS POSSIBLES



- *méconnaissance du milieu bancaire – être plus critique sur le pre-processing*
- *métrique d'évaluation et la fonction coût*
- *Améliorer les performances de la modélisation*

*Développer un dashboard avec une partie client et une partie banque*

*Un dashboard dans le lequel on peut simuler plusieurs scénarios de crédit*

# VII. CONCLUSION



UTILISATION ET MODIFICATION D'UN KERNEL KAGGLE

ENTRAINEMENT D'UN MODELE DE SCORING

FONCTION COUT ET EVALUATION

INTERPREPATION

DASHBOARD