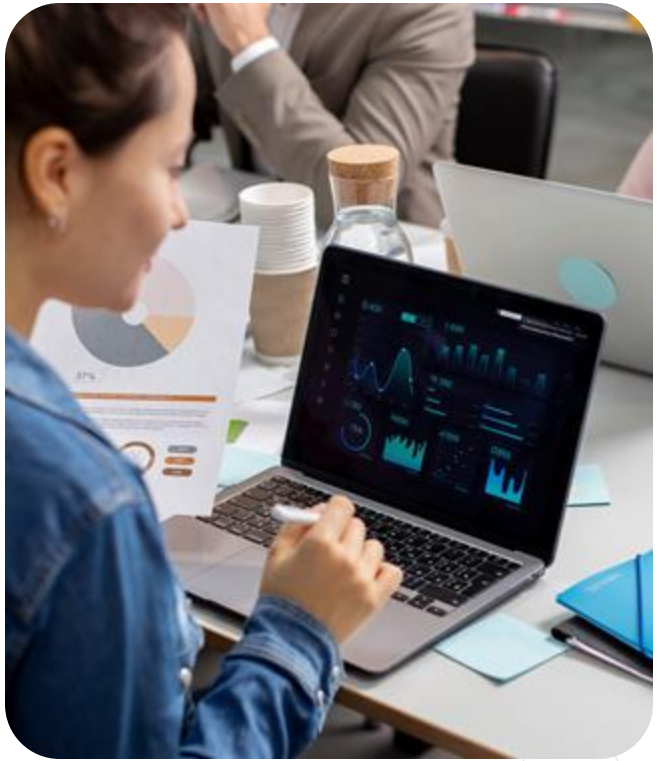
Projet 7:Implémentez un modèle de scoring

Kaoutar El Mardi jury: Daouda Thioye

Table of contents

Introduction

"Prêt à dépenser", une société financière spécialisée dans les crédits à la consommation pour des personnes ayant peu ou pas du tout d'historique de prêt Cette présentation porte sur un projet de développement d'un outil de "scoring crédit" que nous avons réalisé pour notre entreprise.



01

Problématique et jeu de Données

Problématique

Jeu de données

```
data1 = pd.read_csv("application_train.csv")
data2=pd.read_csv("application_test.csv")
data3=pd.read_csv("bureau_balance.csv")
data4=pd.read_csv("bureau.csv")
data5=pd.read_csv("HomeCredit_columns_description.csv",encoding='ISO-8859-1')
data6=pd.read_csv("installments_payments.csv")
data7=pd.read_csv("POS_CASH_balance.csv")
data8=pd.read_csv("previous_application.csv")
data9=pd.read_csv("sample_submission.csv")
```

```
(356251, 797)
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 356251 entries, 0 to 356254
Columns: 797 entries, SK_ID_CURR to CC_COUNT
dtypes: bool(133), float64(606), int64(42), object(16)
memory usage: 1.8+ GB
```

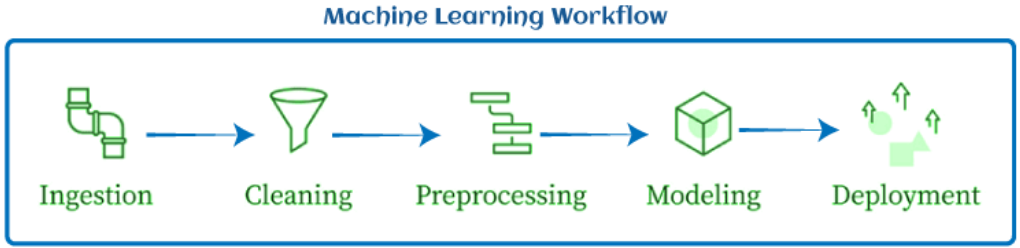
SK_ID_CURR	Identifiant unique pour chaque client dans le jeu de données.
TARGET	Variable cible indiquant si le client a fait défaut sur le prêt (1 = défaut, 0 = non défaut).
CODE_GENDER	Code représentant le genre du client (0 = féminin, 1 = masculin, par exemple).
FLAG_OWN_CAR	Indicateur binaire indiquant si le client possède une voiture (1 = oui, 0 = non).
FLAG_OWN_REALTY	Indicateur binaire indiquant si le client possède une propriété (1 = oui, 0 = non).
CNT_CHILDREN	Nombre d'enfants à charge du client.
AMT_INCOME_TOTAL	Revenu total annuel du client.
AMT_CREDIT	Montant total du crédit demandé par le client.
AMT_ANNUITY	Montant annuel de l'annuité que le client doit payer pour le crédit.
NAME_CONTRACT_TYPE	Type de contrat de prêt (Cash loans, Revolving loans).
FLAG_MOBIL	Indicateur de possession de téléphone mobile.
FLAG_EMP_PHONE	Indicateur si le client a un téléphone professionnel.
FLAG_WORK_PHONE	Indicateur si le client a un téléphone au travail.
FLAG_CONT_MOBILE	Indicateur si le client a un mobile à contact permanent.
FLAG_PHONE	Indicateur si le client a un téléphone fixe.
FLAG_EMAIL	Indicateur si le client a un email.
OCCUPATION_TYPE	Type de profession du client.
CNT_FAM_MEMBERS	Nombre de membres dans la famille du client.
REGION_RATING_CLIENT	Note de la région où vit le client.
WEEKDAY_APPR_PROCESS_START	Jour de la semaine où la demande de prêt a été initiée.
HOUR_APPR_PROCESS_START	Heure de la journée où la demande de prêt a été initiée.
ORGANIZATION_TYPE	Type d'organisation où le client travaille.
EXT_SOURCE_1	Score externe de la source 1.
EXT_SOURCE_2	Score externe de la source 2.
EXT_SOURCE_3	Score externe de la source 3.
APARTMENTS_AVG	Moyenne des valeurs pour les appartements associés au client.
BASEMENTAREA_AVG	Moyenne des valeurs pour les sous-sols associés au client.

02

Modélisation et MLFlow UI

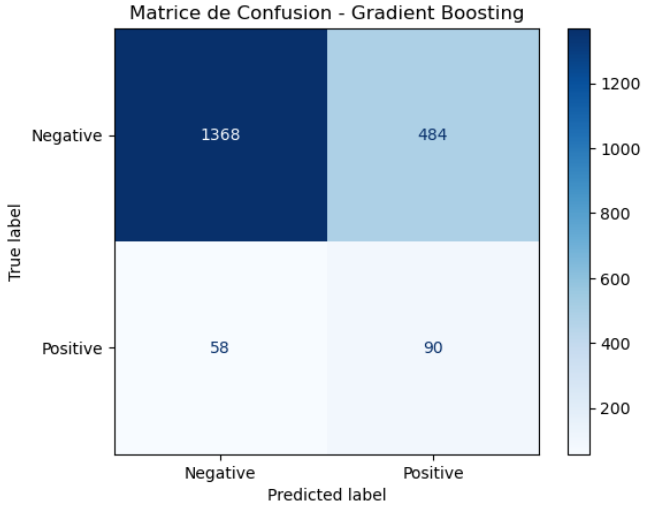
Description du pipeline de modélisation

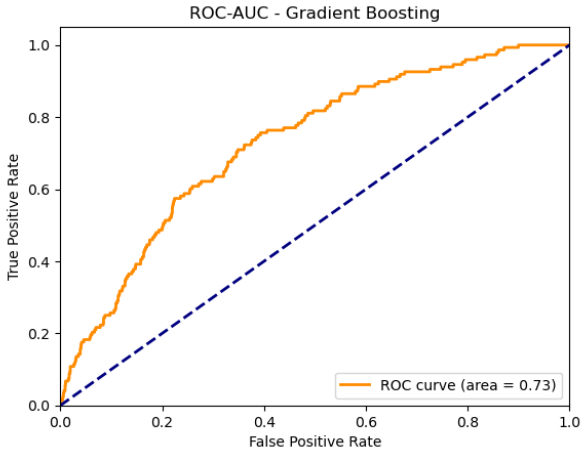
- préparation des données
- Création du Pipeline de Modélisation
- Recherche d'Hyperparamètres
- Évaluation du Modèle
- Suivi et Enregistrement
- Déploiement du Modèle Intégration via API



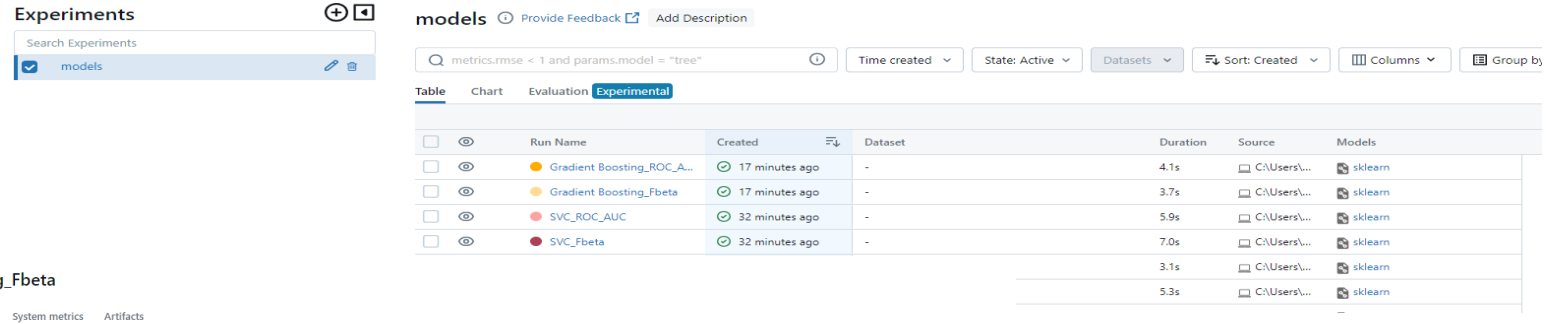
Algorithmes Resultats

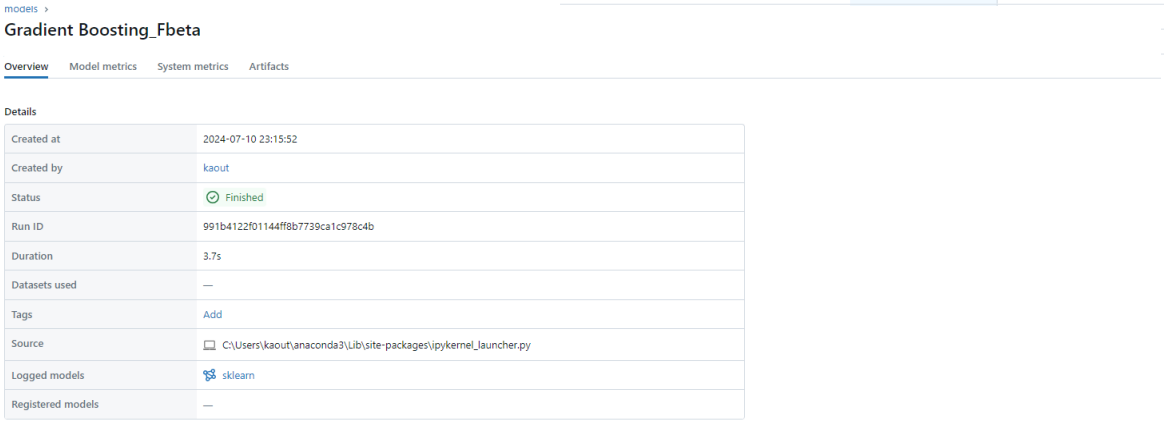
	F-beta (train)	F-beta (test)	ROC AUC (train)	ROC AUC (test)	F-beta (cross-validation)	ROC AUC (cross-validation)	Model	Best Params
0	0.676367	0.662848	0.685176	0.672596	0.674904	0.736434	Logistic Regression	{'classifier__C': 1.230477933008369, 'classifi...
1	0.675421	0.661981	0.684216	0.671787	0.669426	0.734063	SVC	{'classifier__kernel': 'linear', 'classifier__...
2	0.683419	0.665802	0.690444	0.673385	0.659149	0.720698	Gradient Boosting	{'classifier__learning_rate': 0.01626287140946...

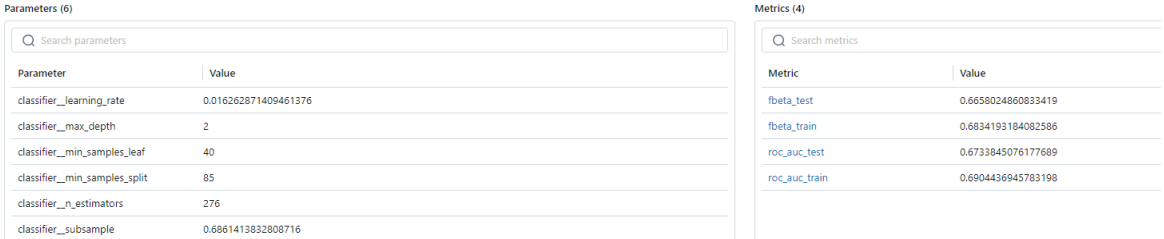




Visualisation du Tracking via MLFlow UI



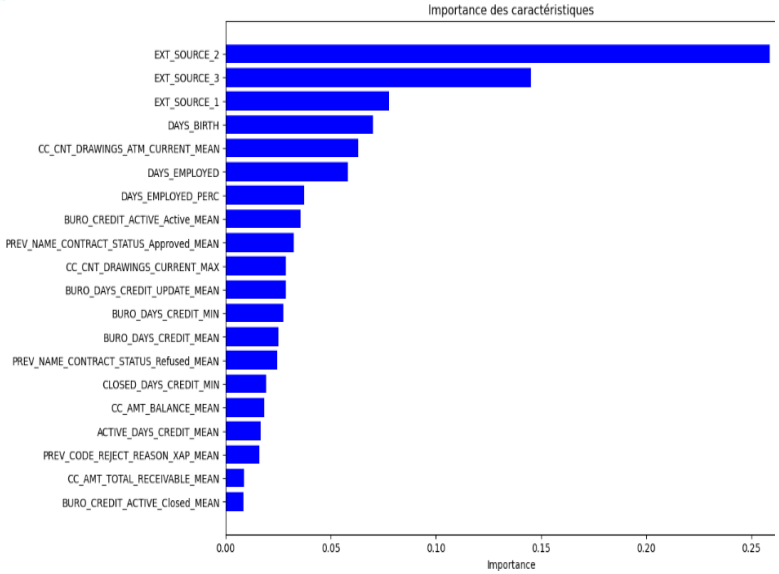


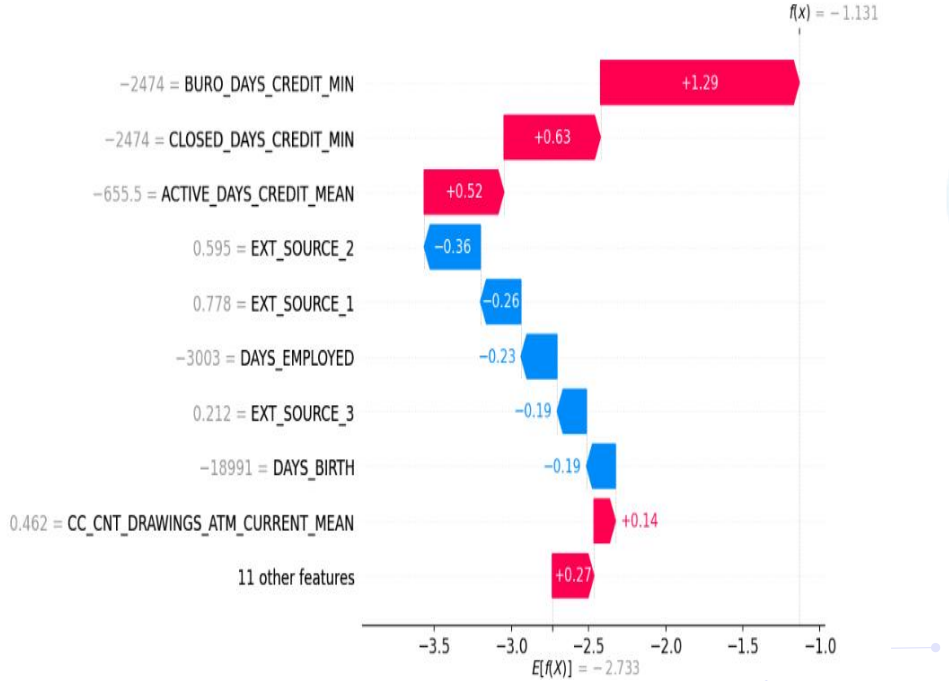


03

Feature Importance Globale et Locale Data Drift

Global et Local





Data drift

Dataset Drift

Dataset Drift is NOT detected. Dataset drift detection threshold is 0.5

121	9	0.0744
Columns	Drifted Columns	Share of Drifted Columns

Data Drift Summary

Drift is detected for 7.438% of columns (9 out of 121).

Search

Column	Type	Reference Distribution	Current Distribution	Data Drift	Stat Test
AMT_REQ_CREDIT_BUREAU_QRT	num			Detected	Wasserstei distance (normed)
AMT_REQ_CREDIT_BUREAU_MON	num			Detected	Wasserstei distance (normed)
AMT_GOODS_PRICE	num			Detected	Wasserstei distance (normed)
AMT_CREDIT	num			Detected	Wasserstei distance (normed)
AMT_ANNUITY	num			Detected	Wasserstei distance (normed)
AMT_REQ_CREDIT_BUREAU_WEEK	num			Detected	Wasserstei distance (normed)
NAME_CONTRACT_TYPE	cat			Detected	Jensen-Shannon distance
DAYS_LAST_PHONE_CHANGE	num			Detected	Wasserstei distance (normed)
FLAG_EMAIL	num			Detected	Jensen-Shannon distance
FLAG_DOCUMENT_3	num			Not Detected	Jensen-Shannon distance

Rows per page: 10 rows 1-10 of 121

121	9	0.0744
Columns	Drifted Columns	Share of Drifted Columns

Data Drift Summary

Drift is detected for 7.438% of columns (9 out of 121).

Search

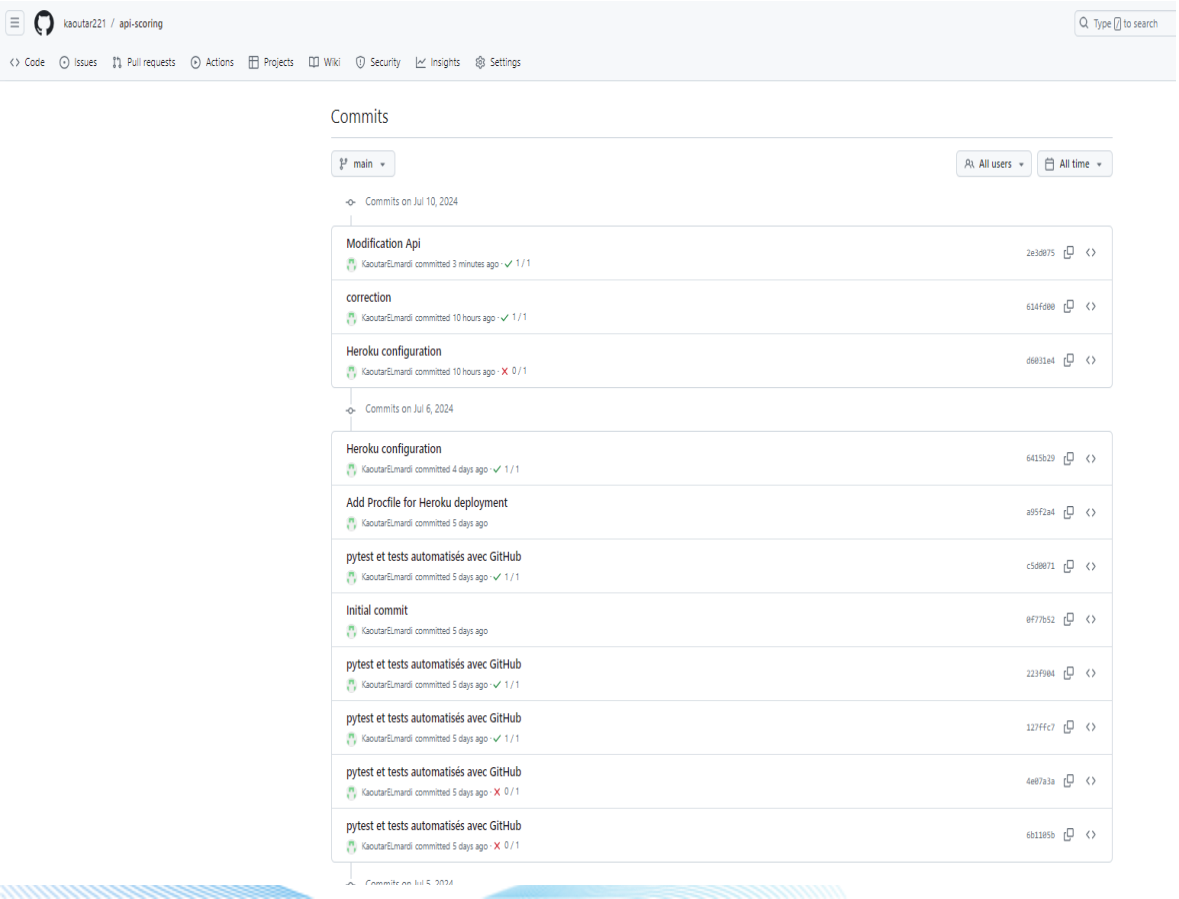
Column	Type	Reference Distribution	Current Distribution	Data Drift	Stat Test
EXT_SOURCE_3	num			Not Detected	Wasserstei distance (normed)
FLAG_DOCUMENT_16	num			Not Detected	Jensen-Shannon distance
AMT_REQ_CREDIT_BUREAU_YEAR	num			Not Detected	Wasserstei distance (normed)
LIVINGAPARTMENTS_AVG	num			Not Detected	Wasserstei distance (normed)
LIVINGAPARTMENTS_MEDI	num			Not Detected	Wasserstei distance (normed)
LIVINGAPARTMENTS_MODE	num			Not Detected	Wasserstei distance (normed)
FLOORSMAX_AVG	num			Not Detected	Wasserstei distance (normed)
FLOORSMAX_MEDI	num			Not Detected	Wasserstei distance (normed)
EXT_SOURCE_2	num			Not Detected	Wasserstei distance (normed)
FLOORSMAX_MODE	num			Not Detected	Wasserstei distance (normed)

Rows per page: 10 rows 11-20 of 121

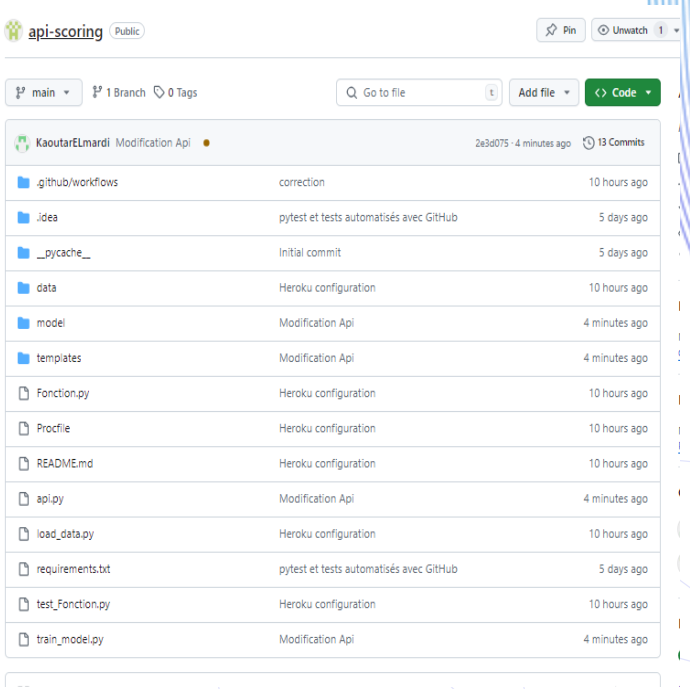
04

Pipeline de Déploiement
Scoring Client via l'API sur le Cloud

Processus de versionnement avec Git



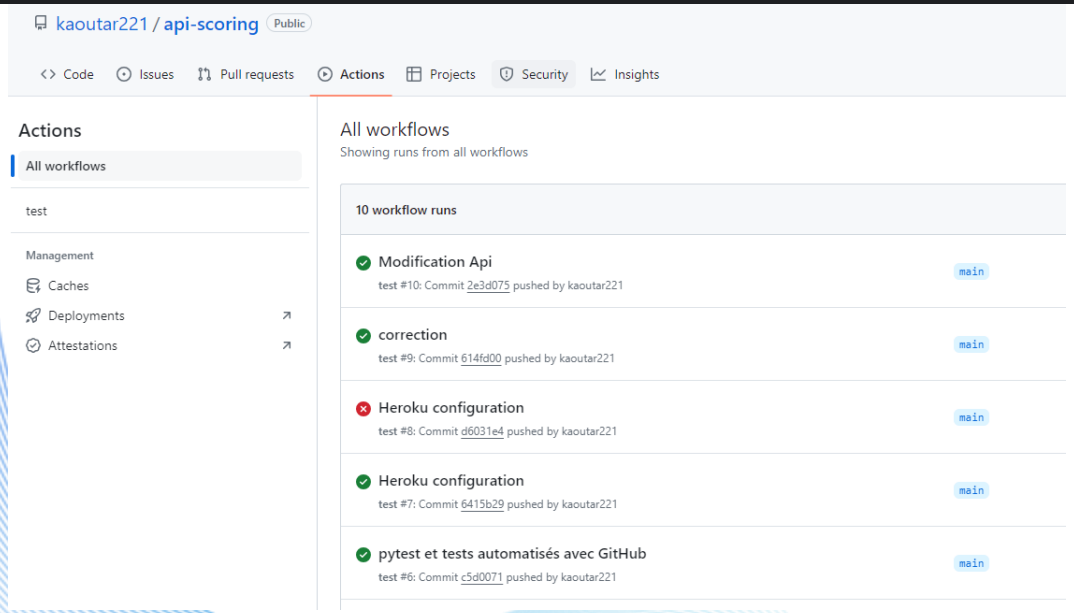
Lien:https://github.com/kaoutar221/api-scoring.git

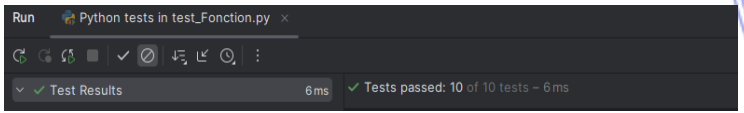


Présentation des Tests Unitaires

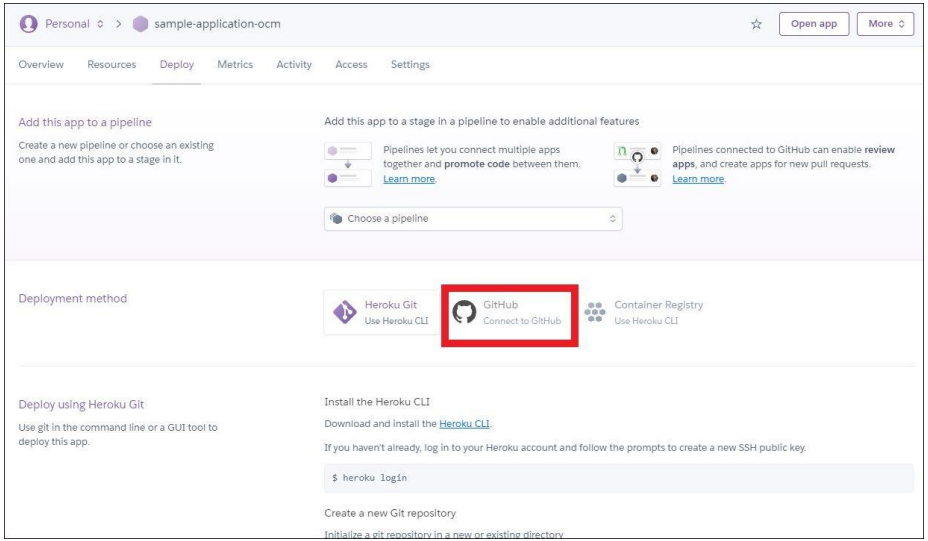
```python
import pytest
import json
from Fonction import app

KaoutarELmardi
def test_homepage():
    with app.test_client() as client:
        response = client.get('/')
        assert response.data.decode('utf-8') == "Application, model, and data have been successfully loaded."
```



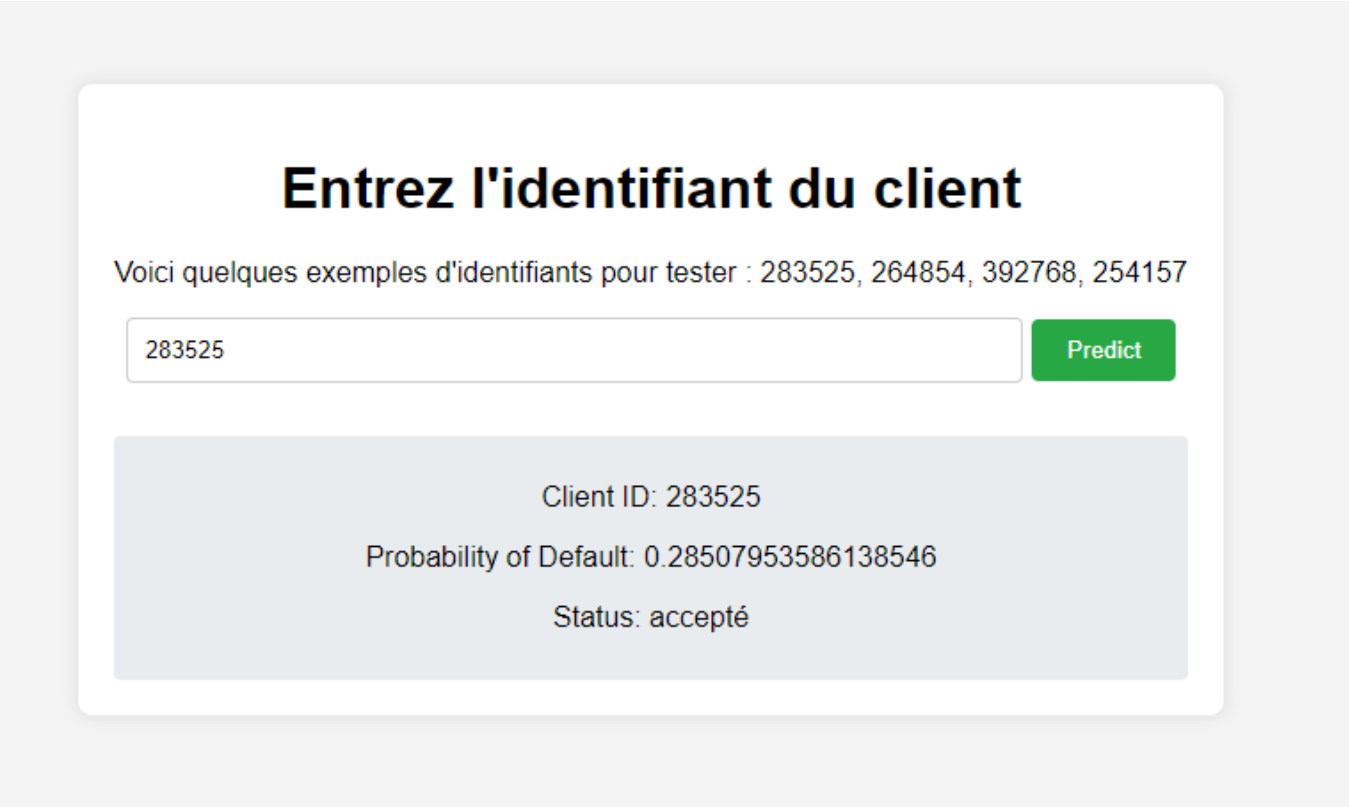


Exécution du Déploiement de l'API



Lien:https://api-scoring-app-04b1217cf097.herokuapp.com/

Exemple de résultats



Conclusion

Thanks!

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**