



Projet 7: Implémentez un modèle de scoring

Kaoutar El Mardi

jury: Daouda Thioye

Table of contents

01

Introduction

02

Problématique
Et jeu de Données

03

Modélisation et
MLFlow UI

04

Feature Importance
Globale et Locale
Data Drift

05

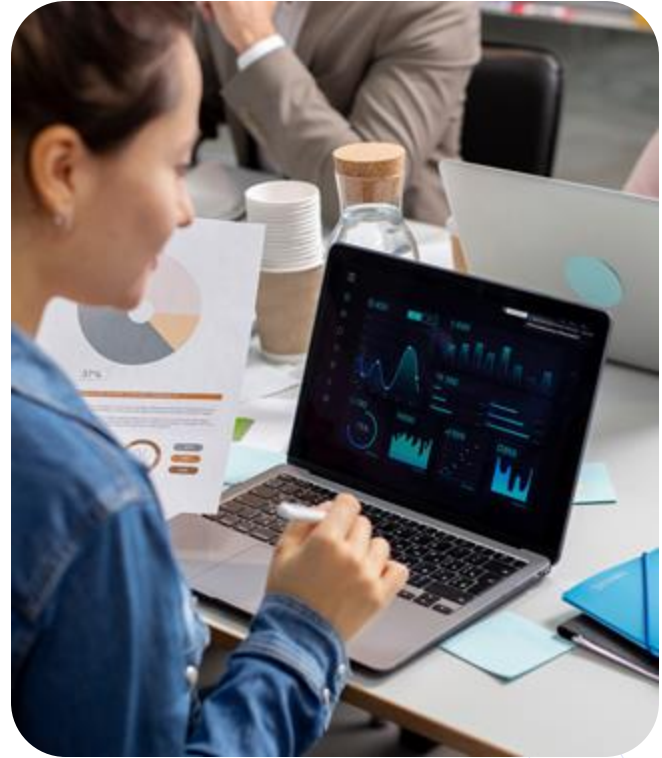
Pipeline de
Déploiement
Scoring Client via
l'API sur le Cloud

06

Conclusion

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.

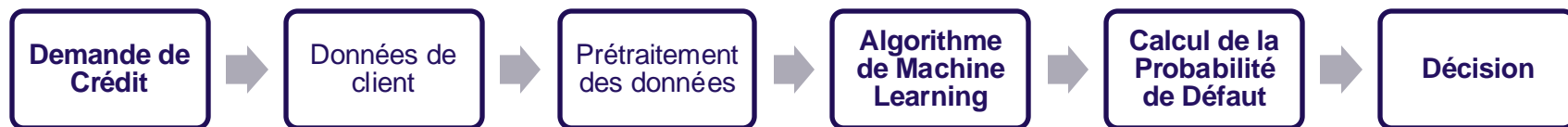




01

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
```

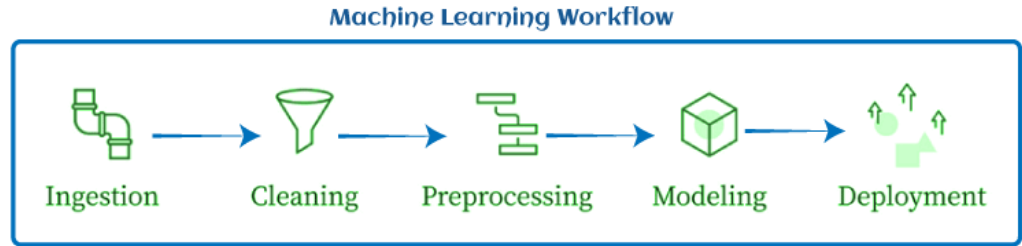
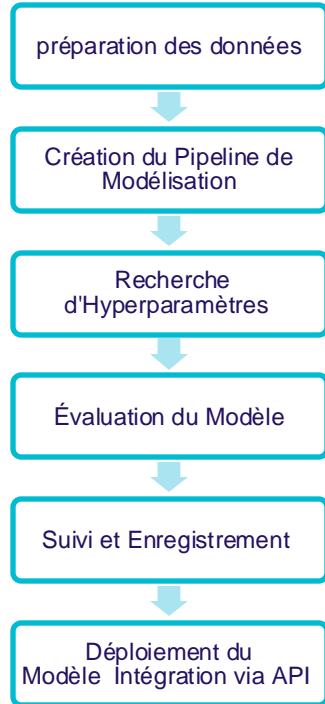
SK_ID_CURR	Identifiant unique pour chaque client dans le jeu de données.
TARGET	Variable cible indiquant si le client a fait défaut sur le prêt (1 = défaut, 0 = non défaut).
CODE_GENDER	Code représentant le genre du client (0 = féminin, 1 = masculin, par exemple).
FLAG_OWN_CAR	Indicateur binaire indiquant 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.
HOURLY_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

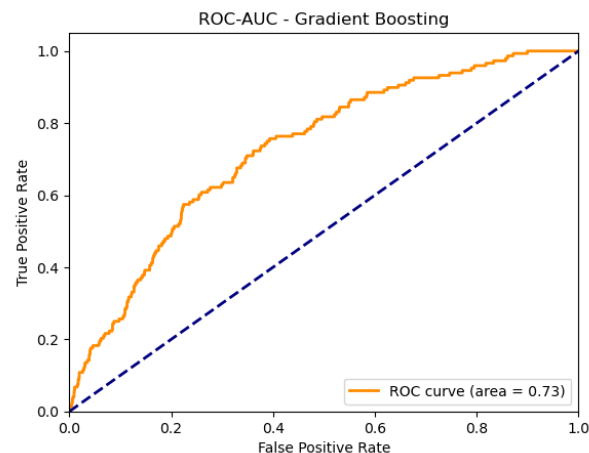
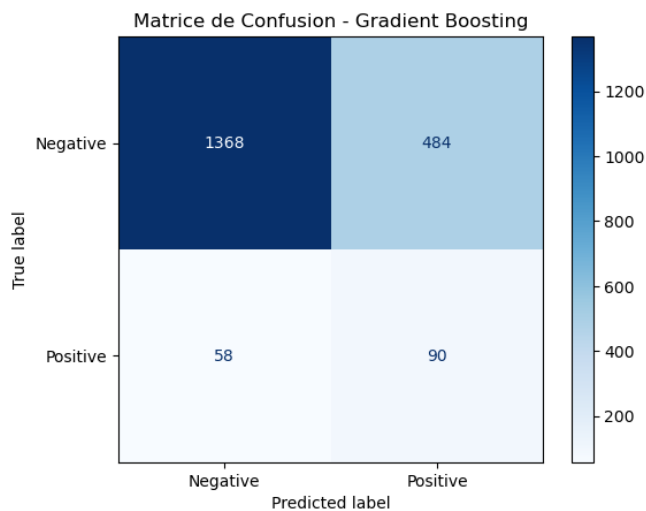
Modélisation et MLFlow UI

Description du pipeline de modélisation



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

Experiments

Search Experiments

☒ models



models 🕒 Provide Feedback 📝 Add Description

Q metrics.rmse < 1 and params.model = "tree" 🕒 Time created ▾ State: Active ▾ Datasets ▾ Sort: Created ▾ Columns ▾ Group by

Table Chart Evaluation Experimental

<input type="checkbox"/>		Run Name	Created	Dataset	Duration	Source	Models
<input type="checkbox"/>		Gradient Boosting_ROC_A...	🟢 17 minutes ago	-	4.1s	C:\Users\...	sklearn
<input type="checkbox"/>		Gradient Boosting_Fbeta	🟢 17 minutes ago	-	3.7s	C:\Users\...	sklearn
<input type="checkbox"/>		SVC_ROC_AUC	🟢 32 minutes ago	-	5.9s	C:\Users\...	sklearn
<input type="checkbox"/>		SVC_Fbeta	🟢 32 minutes ago	-	7.0s	C:\Users\...	sklearn
					3.1s	C:\Users\...	sklearn
					5.3s	C:\Users\...	sklearn

models >

Gradient Boosting_Fbeta

Overview Model metrics System metrics Artifacts

Details

Created at	2024-07-10 23:15:52
Created by	kaout
Status	🟢 Finished
Run ID	991b4122f01144ffb7739ca1c978c4b
Duration	3.7s
Datasets used	—
Tags	Add
Source	C:\Users\kaout\anaconda3\Lib\site-packages\ipykernel_launcher.py
Logged models	sklearn
Registered models	—

Parameters (6)

Q Search parameters	
Parameter	Value
classifier__learning_rate	0.016262871409461376
classifier__max_depth	2
classifier__min_samples_leaf	40
classifier__min_samples_split	85
classifier__n_estimators	276
classifier__subsample	0.6861413832808716

Metrics (4)

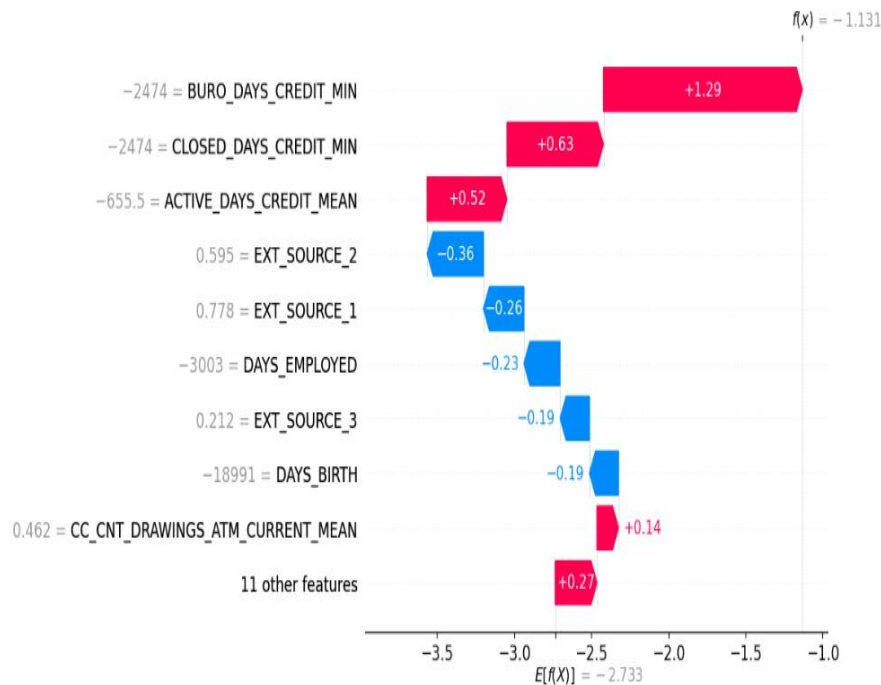
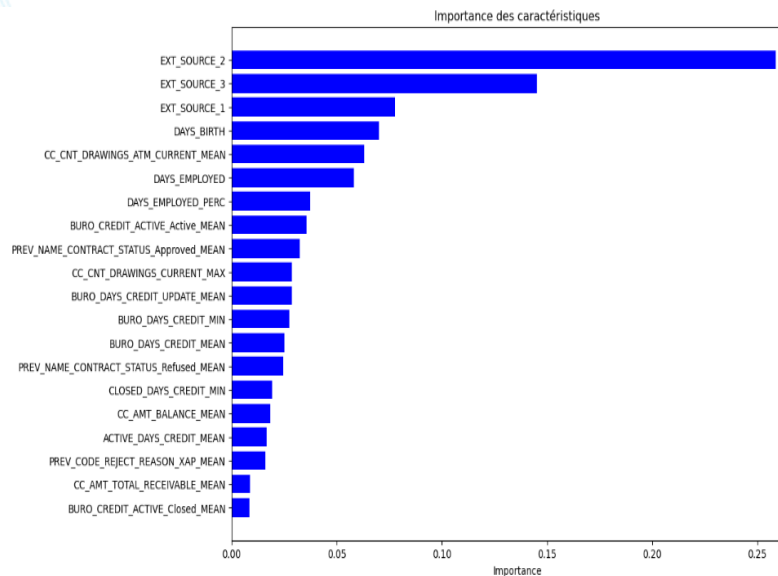
Q Search metrics	
Metric	Value
fbeta_test	0.6658024860833419
fbeta_train	0.6834193184082586
roc_auc_test	0.6733845076177689
roc_auc_train	0.6904436945783198



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).

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

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).

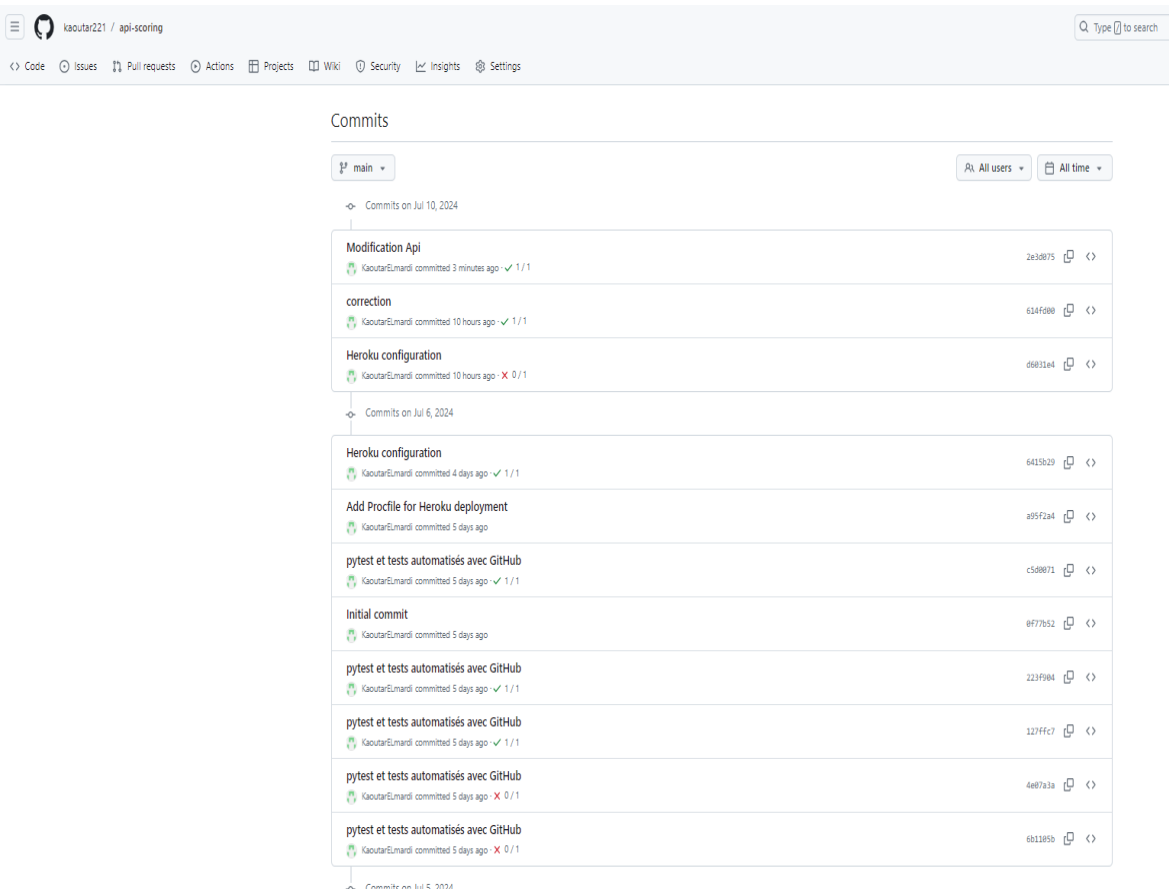
Q Search X					
Column	Type	Reference Distribution	Current Distribution	Data Drift	Stat Test
> EXT_SOURCE_3	num			Not Detected	Wasserstein distance (normed)
> FLAG_DOCUMENT_16	num			Not Detected	Jensen-Shannon distance
> AMT_REQ_CREDIT_BUREAU_YEAR	num			Not Detected	Wasserstein distance (normed)
> LIVINGAPARTMENTS_AVG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



04

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

Processus de versionnement avec Git



The screenshot shows the GitHub interface for the repository 'api-scoring' by user 'kaoutar221'. The top navigation bar includes links for Code, Issues, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. The main content area is titled 'Commits' and shows a list of commits on the 'main' branch. The commits are grouped by date, with the most recent group being 'Commits on Jul 10, 2024'. The commits in this group are: 'Modification Api' (2e3d075, 3 minutes ago), 'correction' (634f608, 10 hours ago), and 'Heroku configuration' (66831e4, 10 hours ago). Below this, there is a group for 'Commits on Jul 6, 2024' with 13 commits, including 'Heroku configuration', 'Add Profile for Heroku deployment', 'pytest et tests automatisés avec GitHub', 'Initial commit', and several other updates to 'pytest et tests automatisés avec GitHub'.

Commits

main

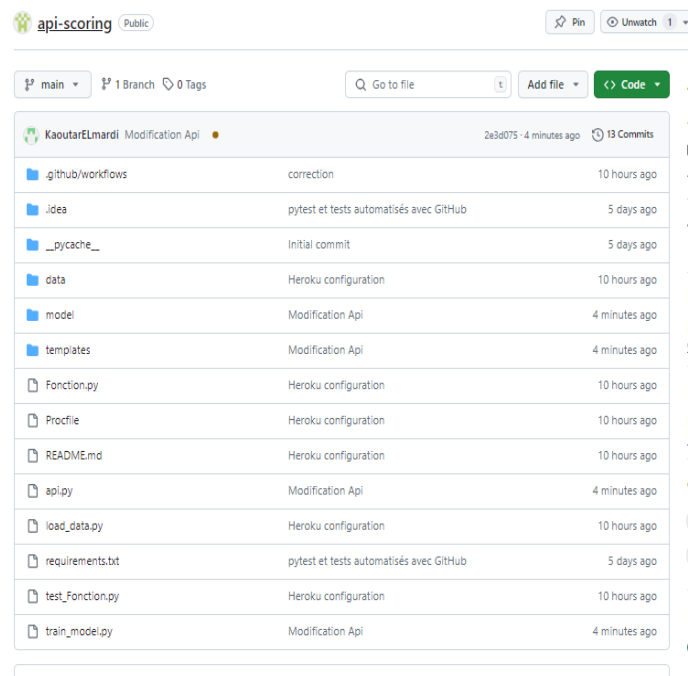
Commits on Jul 10, 2024

Commit Message	Commit Hash	Time Ago	Status
Modification Api	2e3d075	3 minutes ago	✓ 1 / 1
correction	634f608	10 hours ago	✓ 1 / 1
Heroku configuration	66831e4	10 hours ago	✗ 0 / 1

Commits on Jul 6, 2024

Commit Message	Commit Hash	Time Ago	Status
Heroku configuration	6415b29	4 days ago	✓ 1 / 1
Add Profile for Heroku deployment	a95f2a4	5 days ago	
pytest et tests automatisés avec GitHub	c5d8672	5 days ago	✓ 1 / 1
Initial commit	8f77b52	5 days ago	
pytest et tests automatisés avec GitHub	223f984	5 days ago	✓ 1 / 1
pytest et tests automatisés avec GitHub	127ffc7	5 days ago	✓ 1 / 1
pytest et tests automatisés avec GitHub	4e87a3a	5 days ago	✗ 0 / 1
pytest et tests automatisés avec GitHub	6b13f5b	5 days ago	✗ 0 / 1

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



The screenshot shows the file structure of the 'api-scoring' repository. The repository is public and has 1 branch and 0 tags. The file structure is as follows:

- .github/workflows: correction (10 hours ago)
- idea: pytest et tests automatisés avec GitHub (5 days ago)
- __pycache__: Initial commit (5 days ago)
- data: Heroku configuration (10 hours ago)
- model: Modification Api (4 minutes ago)
- templates: Modification Api (4 minutes ago)
- Fonction.py: Heroku configuration (10 hours ago)
- Profile: Heroku configuration (10 hours ago)
- README.md: Heroku configuration (10 hours ago)
- api.py: Modification Api (4 minutes ago)
- load_data.py: Heroku configuration (10 hours ago)
- requirements.txt: pytest et tests automatisés avec GitHub (5 days ago)
- test_Fonction.py: Heroku configuration (10 hours ago)
- train_model.py: Modification Api (4 minutes ago)

Présentation des Tests Unitaires

```
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."
```

kaoutar221 / api-scoring (Public)

<> Code Issues Pull requests Actions Projects Security Insights

Actions

All workflows

test

Management

Caches

Deployments

Attestations

All workflows

Showing runs from all workflows

10 workflow runs

Modification Api

test #10: Commit 2e3d075 pushed by kaoutar221

main

correction

test #9: Commit 614fd00 pushed by kaoutar221

main

Heroku configuration

test #8: Commit d6031e4 pushed by kaoutar221

main

Heroku configuration

test #7: Commit 6415b29 pushed by kaoutar221

main

pytest et tests automatisés avec GitHub

test #6: Commit c5d0071 pushed by kaoutar221

main

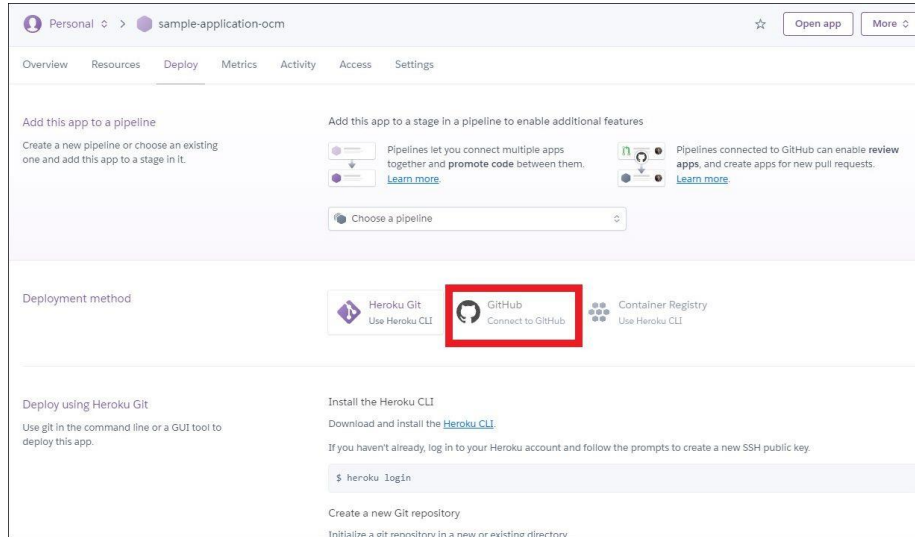
Run Python tests in test_Fonction.py

Test Results

6ms

Tests passed: 10 of 10 tests - 6ms

Exécution du Déploiement de l'API



The screenshot shows the Heroku dashboard for an application named 'sample-application-ocm'. The 'Deploy' tab is selected. Under 'Add this app to a pipeline', there are instructions and a 'Choose a pipeline' dropdown. The 'Deployment method' section shows three options: 'Heroku Git' (selected), 'GitHub' (highlighted with a red box), and 'Container Registry'. Below this, the 'Deploy using Heroku Git' section provides instructions on installing the Heroku CLI and logging in, with a terminal snippet: `$ heroku login`.



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

Exemple de résultats

Entrez l'identifiant du client

Voici quelques exemples d'identifiants pour tester : 283525, 264854, 392768, 254157

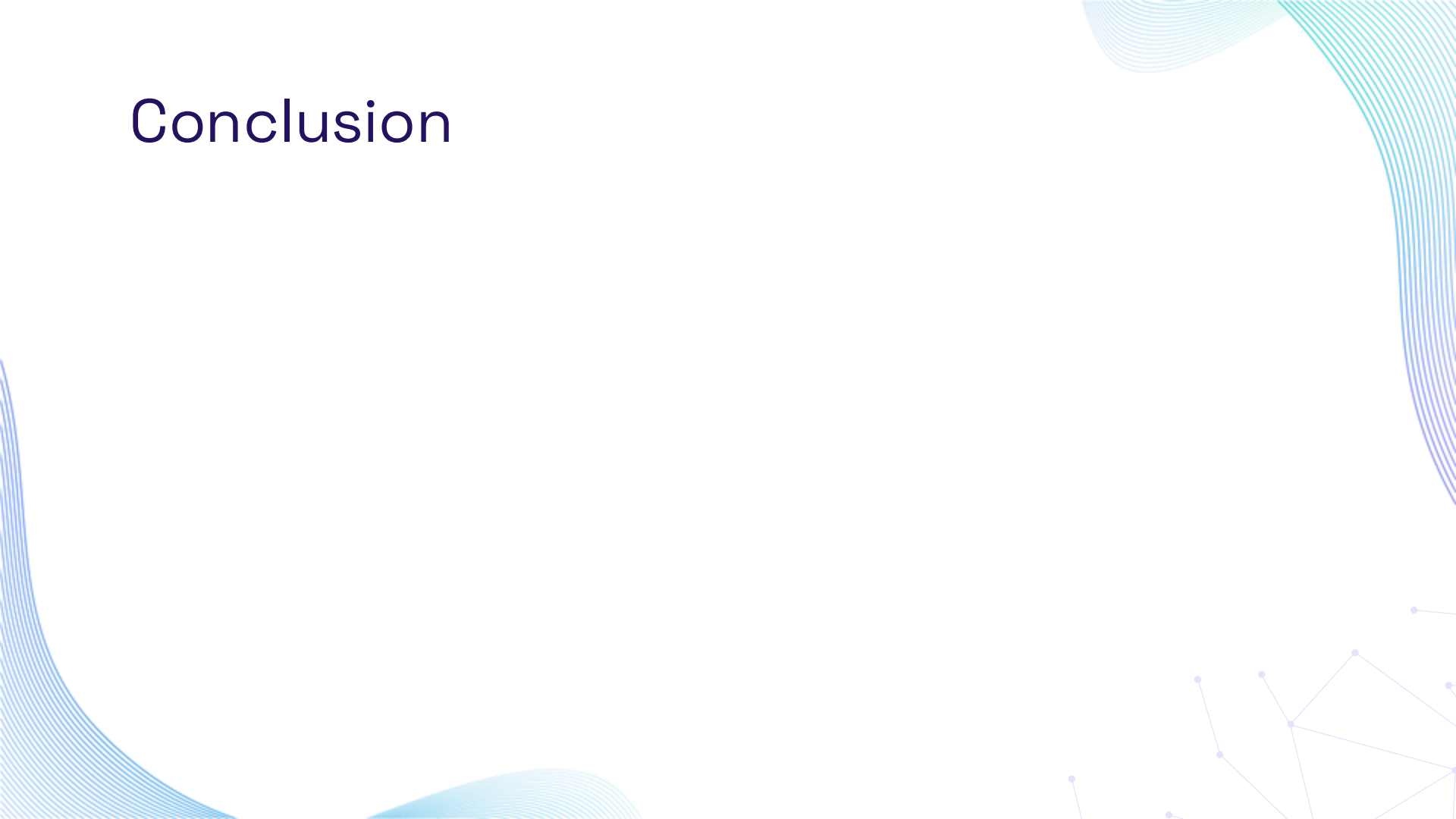
Predict

Client ID: 283525

Probability of Default: 0.28507953586138546

Status: accepté

Conclusion



The background features abstract, flowing wavy lines in shades of blue and teal. In the bottom right corner, there is a network diagram consisting of several nodes connected by thin lines, forming a web-like structure.

Thanks!

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