

Implémentez un modèle de scoring

Soutenance du 16 octobre 2023

Yann Pham-Van

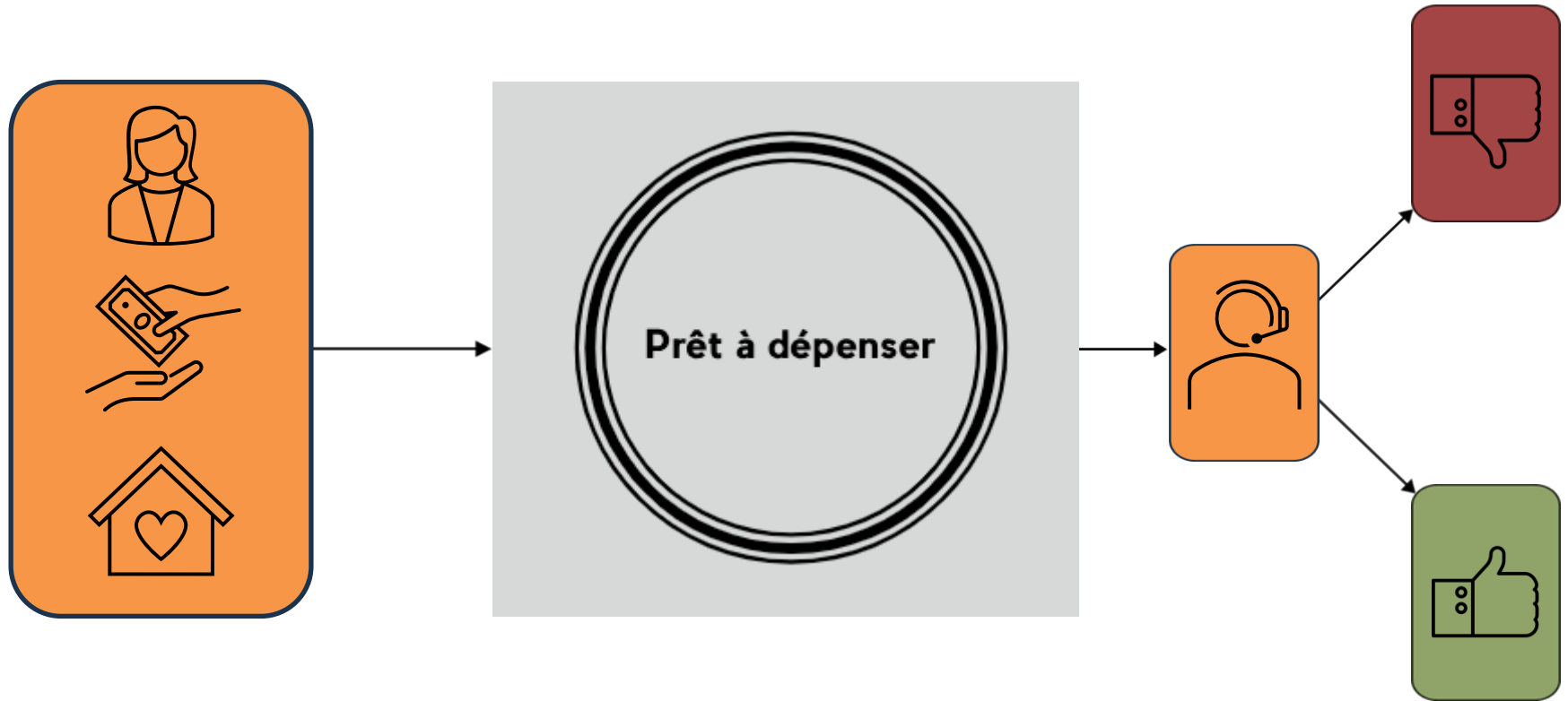
Implémentez un modèle de scoring

- Problématique
- Modélisation
- Déploiement
- Data drift
- Dashboard

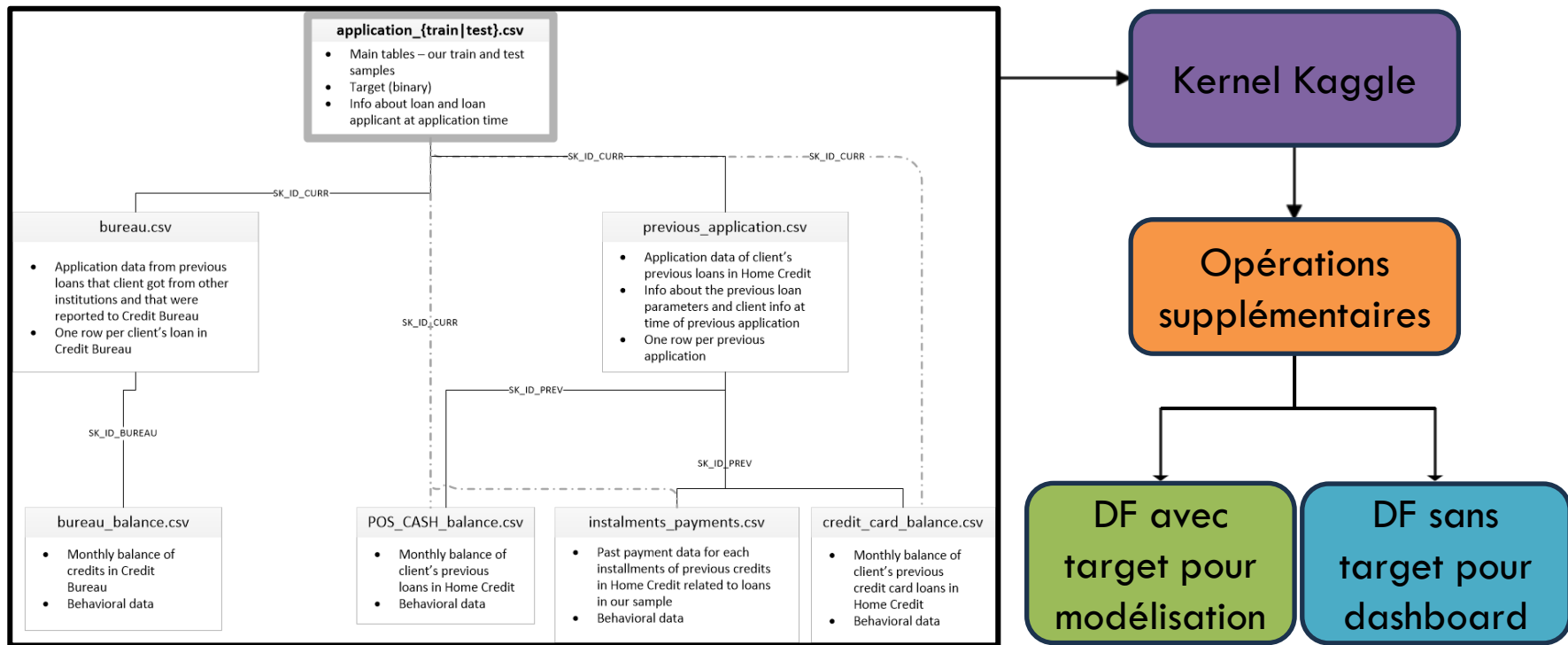
Implémentez un modèle de scoring

- **Problématique**
- Modélisation
- Déploiement
- Data drift
- Dashboard

Problématique



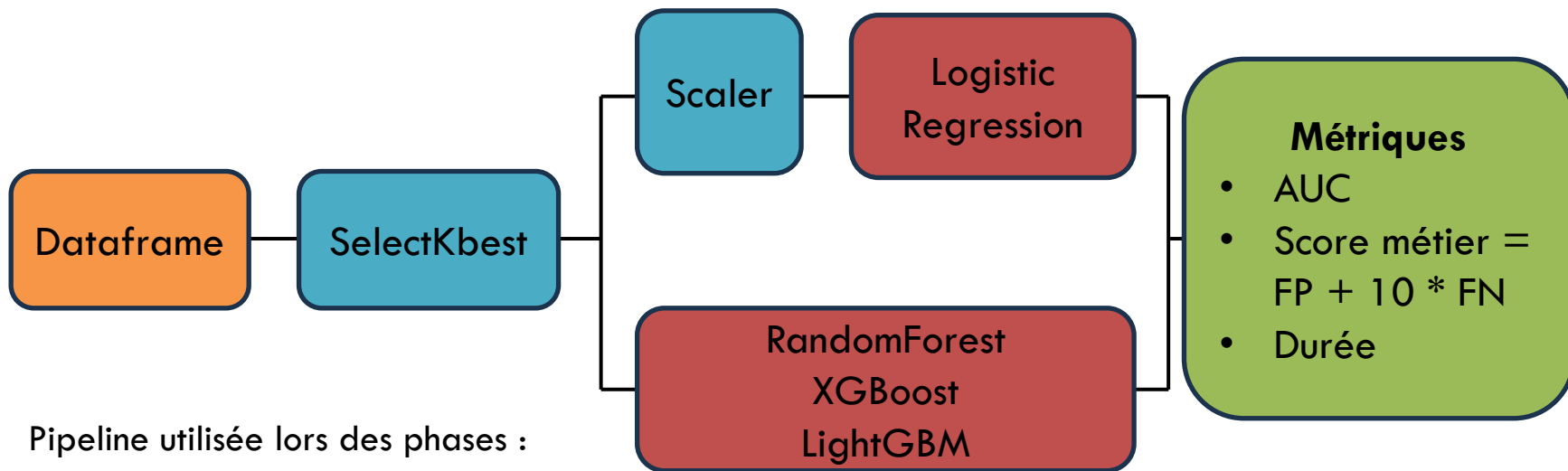
Problématique > dataset



Implémentez un modèle de scoring

- Problématique
- **Modélisation**
- Déploiement
- Data drift
- Dashboard

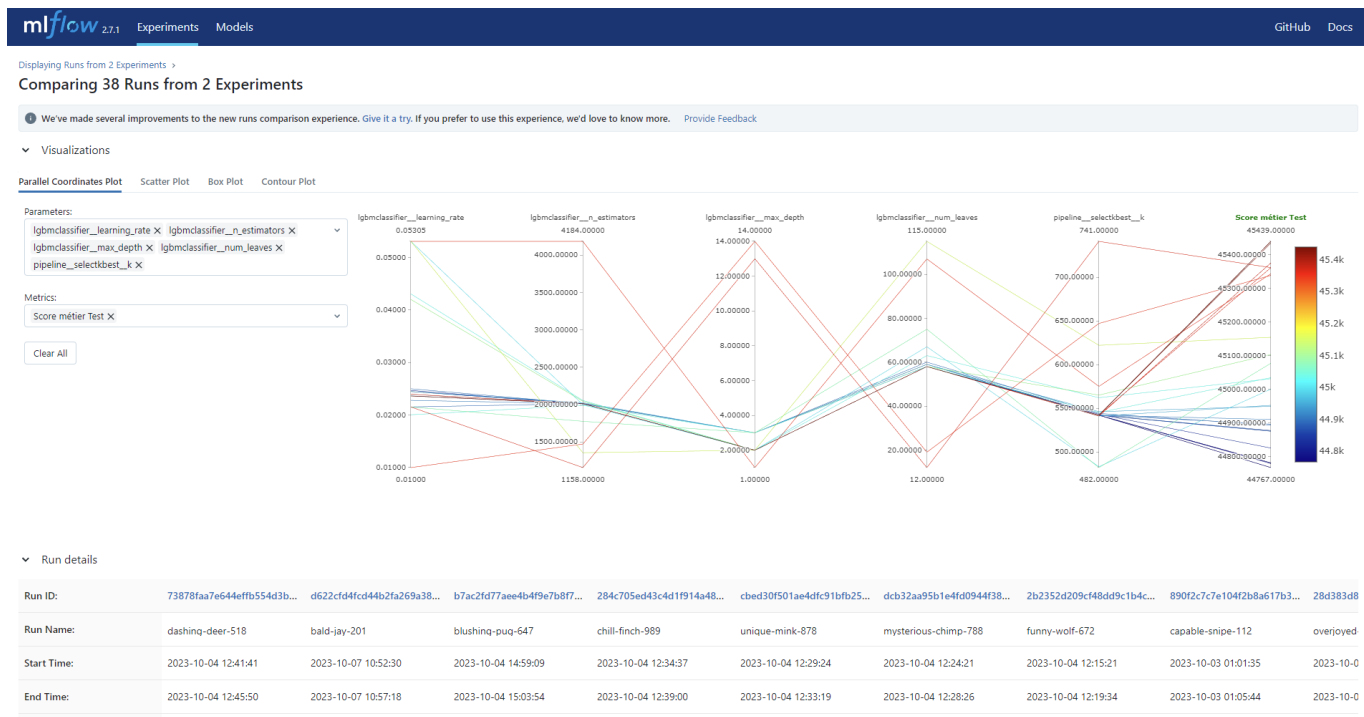
Modélisation > démarche, métriques



Pipeline utilisée lors des phases :

1. Modèles de base
2. Recherche hyperparamètres
3. Optimisation modèle retenu
4. Optimisation seuil de décision

Modélisation > MLFlow



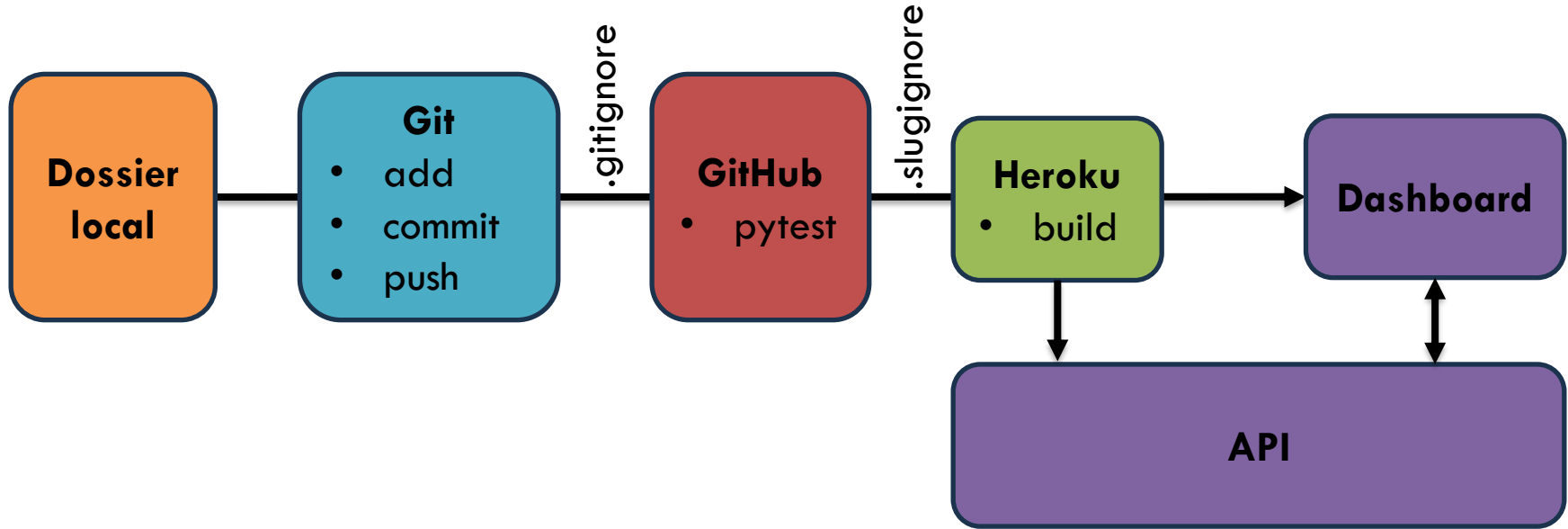
Modélisation > résultats

Phase 1 : modèles de base						Phase 2 : GridSearchCV				Ph. 3	Ph. 4
Modèle	DummyClassifier	RandomForest	LogisticRegression	XGBoost	LightGBM	RandomForest	LogisticRegression	XGBoost	LightGBM	LightGBM RandomSearchCV	LightGBM avec threshold optimisé
AUC	0,5	0,723	0,772	0,766	0,781	0,732	0,773	0,774	0,777	0,785	
Score métier	73290	73202	46421	48016	45502	51724	46175	46829	46083	44767	44705
Durée (s)		48	30	98	27						

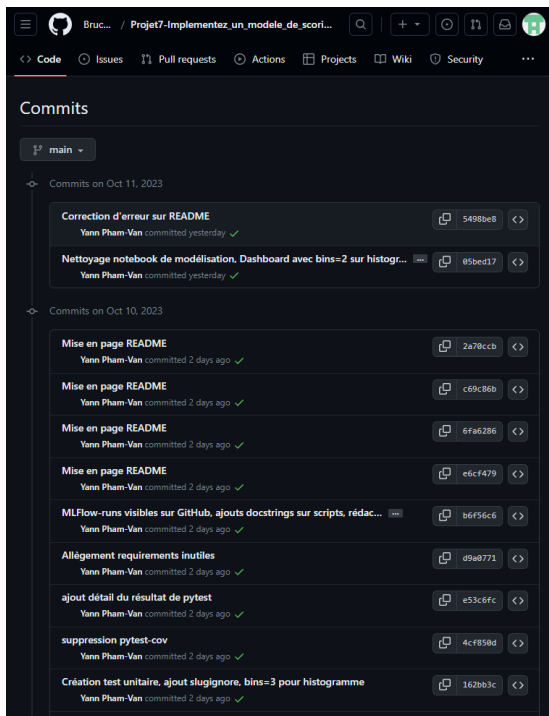
Implémentez un modèle de scoring

- Problématique
- Modélisation
- **Déploiement**
- Data drift
- Dashboard

Déploiement > pipeline

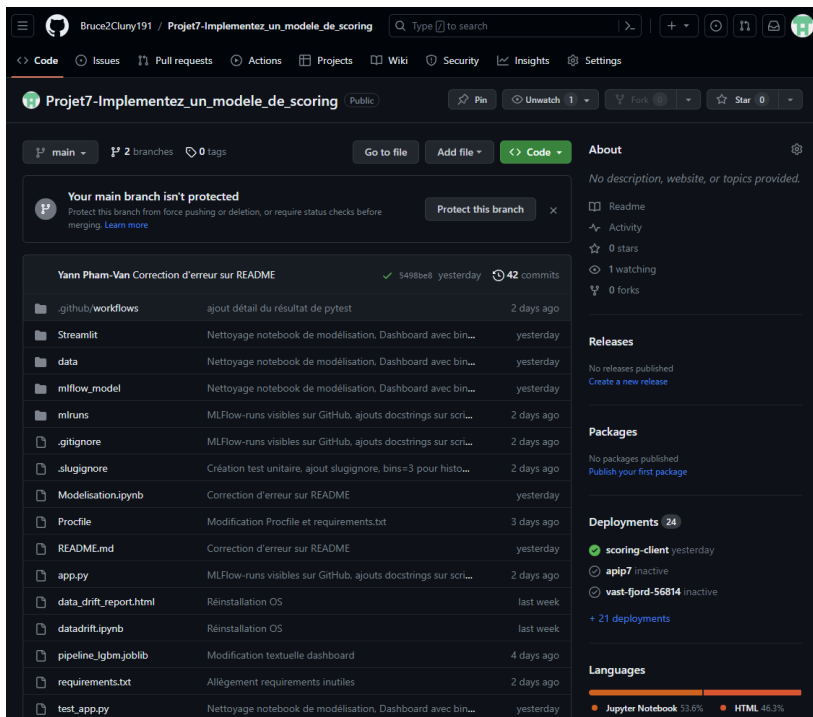


Déploiement > commits



- 42 commits
- Répartis sur 6 semaines

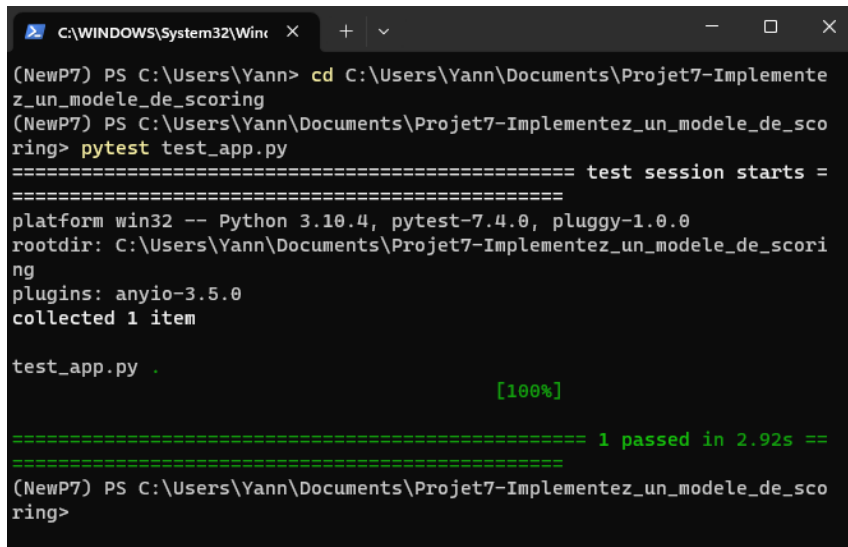
Déploiement > GitHub



- 1 repository principal
- 1 repository secondaire
→ déploiement du dashboard à part

Déploiement > test unitaire

- Pytest
- Vérification que la structure des datas d'un client correspond aux attentes du modèle



```
C:\WINDOWS\System32\Win... X + v
(NewP7) PS C:\Users\Yann> cd C:\Users\Yann\Documents\Projet7-Implementez_un_modele_de_scoring
(NewP7) PS C:\Users\Yann\Documents\Projet7-Implementez_un_modele_de_scoring> pytest test_app.py
===== test session starts =====
platform win32 -- Python 3.10.4, pytest-7.4.0, pluggy-1.0.0
rootdir: C:\Users\Yann\Documents\Projet7-Implementez_un_modele_de_scoring
plugins: anyio-3.5.0
collected 1 item

test_app.py .

[100%]

===== 1 passed in 2.92s =====
(NewP7) PS C:\Users\Yann\Documents\Projet7-Implementez_un_modele_de_scoring>
```

Déploiement > Heroku

The screenshot shows the Heroku dashboard for an application named 'scoring-client'. At the top, there's a navigation bar with the Heroku logo, a search bar, and user profile information. Below this, the app's name 'scoring-client' is displayed along with a 'main' branch selector. A navigation menu includes 'Overview', 'Resources', 'Deploy', 'Metrics', 'Activity', 'Access', and 'Settings'. The main content area is divided into several sections: 1. 'Add this app to a pipeline' with instructions and a dropdown menu to 'Choose a pipeline'. 2. 'Deployment method' showing 'Heroku Git' as the selected method, with 'GitHub Connected' and 'Container Registry' as alternatives. 3. 'App connected to GitHub' showing the app is connected to a specific GitHub repository, with links to 'Releases in the activity feed' and 'Automatic deploys from main'. 4. 'Automatic deploys' with a note about changing the main deploy branch from 'master' to 'main' and a confirmation that deploys from 'main' are enabled.

HEROKU

Jump to Favorites, Apps, Pipelines, Spaces...

Personal > scoring-client

GitHub Bruce2Cluny191/Projet7-Implementez_un_modele_de_scoring main

Overview Resources Deploy Metrics Activity Access Settings

Add this app to a pipeline

Create a new pipeline or choose an existing one and add this app to a stage in it.

Add this app to a stage in a pipeline to enable additional features

Pipelines let you connect multiple apps together and promote code between them. [Learn more](#)

Pipelines connected to GitHub can enable review apps, and create apps for new pull requests. [Learn more](#)

Choose a pipeline

Deployment method

Heroku Git Use Heroku CLI

GitHub Connected

Container Registry Use Heroku CLI

App connected to GitHub

Code diffs, manual and auto deploys are available for this app.

Connected to Bruce2Cluny191/Projet7-Implementez_un_modele_de_scoring by Bruce2Cluny191

Releases in the [activity feed](#) (link to GitHub to view commit diffs)

Automatic deploys from main

Automatic deploys

Enables a chosen branch to be automatically deployed to this app.

You can now change your main deploy branch from "master" to "main" for both manual and automatic deploys, please follow the instructions [here](#).

Automatic deploys from main are enabled

Every push to main will deploy a new version of this app. **Deploys happen automatically:** be sure that this branch in GitHub is always in a deployable state and any tests have passed before you push. [Learn more](#)

- 1 app API
- 1 app dashboard
- Déploiement automatique depuis GitHub

Déploiement > API

FastAPI 0.1.0 OAS 3.1
/openapi.json

default

POST /invocations Operate

Appel des datas du client choisi pour retourner la probabilité

Parameters Cancel

No parameters

Request body required application/json

```
{  
  "client_choice": 0  
}
```

Execute

Responses

Quelques clients à tester :

- 275922
- 178317
- 103516
- 251956
- 115290
- 122215

Implémentez un modèle de scoring

- Problématique
- Modélisation
- Déploiement
- **Data drift**
- Dashboard

Data drift

Dataset Drift

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

121
Columns

9
Drifted Columns

0.0744
Share of Drifted Columns

Data Drift Summary

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

Search

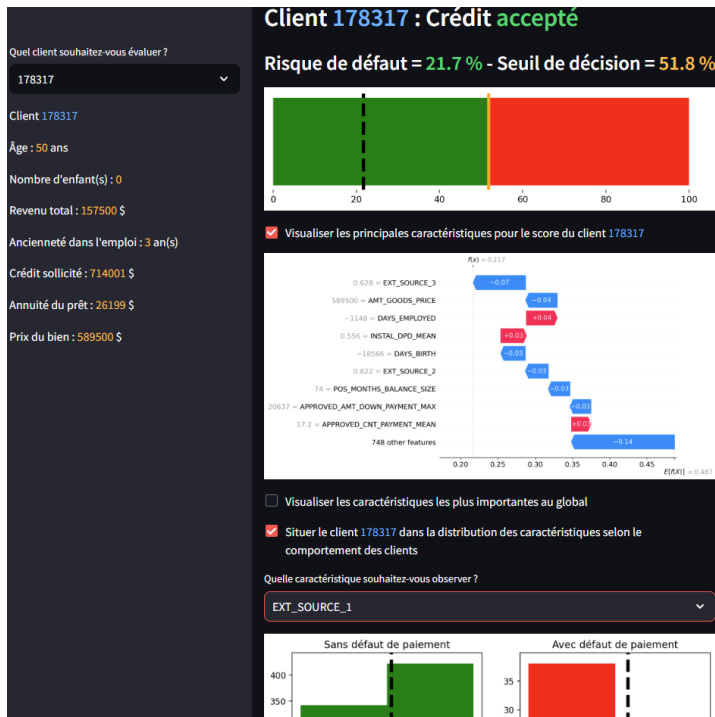
Column	Type	Reference Distribution	Current Distribution	Data Drift	Stat Test	Drift Score
> AMT_REQ_CREDIT_BUREAU_QRT	num			Detected	Wasserstein distance (normed)	0.359052
> AMT_REQ_CREDIT_BUREAU_MON	num			Detected	Wasserstein distance (normed)	0.281765
> AMT_GOODS_PRICE	num			Detected	Wasserstein distance (normed)	0.210785
> AMT_CREDIT	num			Detected	Wasserstein distance (normed)	0.207334

- 9 variables drift $> 10\%$
- Global 7,438% $< 50\%$
→ **Pas de drift**
- 3 variables parmi les plus importantes à surveiller

Implémentez un modèle de scoring

- Problématique
- Modélisation
- Déploiement
- Data drift
- **Dashboard**

Dashboard



- Critères d'accessibilité du WCAG
- Jauge d'acceptation avec situation relative du client
- Principales explications du score
- Position du client dans la distribution des variables selon la classification

Avez-vous des questions ?