



PRÊT À DÉPENSER

Implémentez un modèle de scoring



PRÊT À DÉPENSER

Summary

Objectives and context 

The data set 

Modelling 

MLOps 

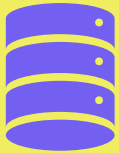
Data Drift 

Demonstration 

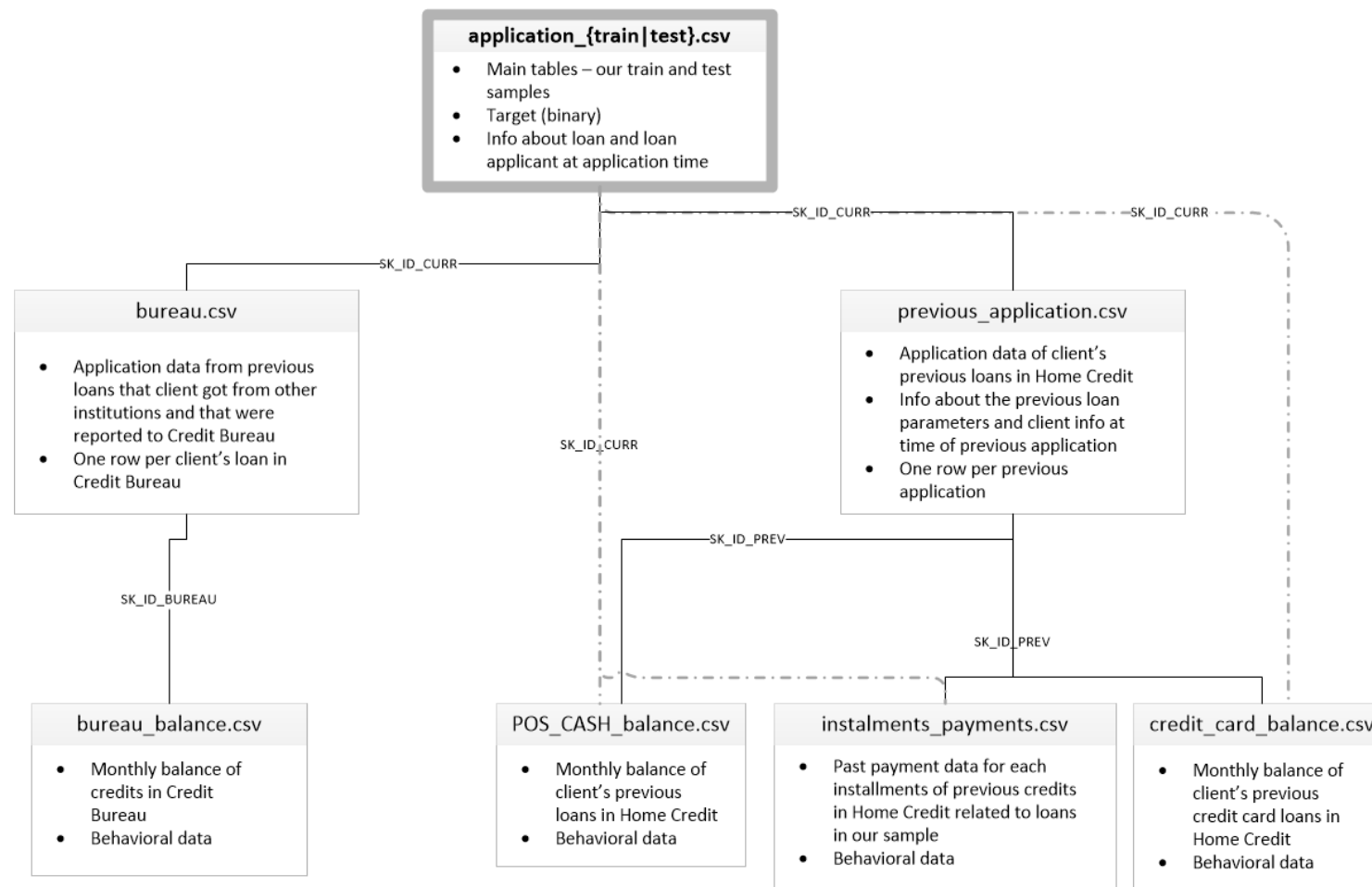


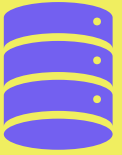
Objectives and context

- Une clientèle avec peu d'historique
- Demande de transparence dans les décisions d'octroi de crédit
- Prédire la probabilité de défaut de paiement
- Mise en place d'un tableau de bord interactif

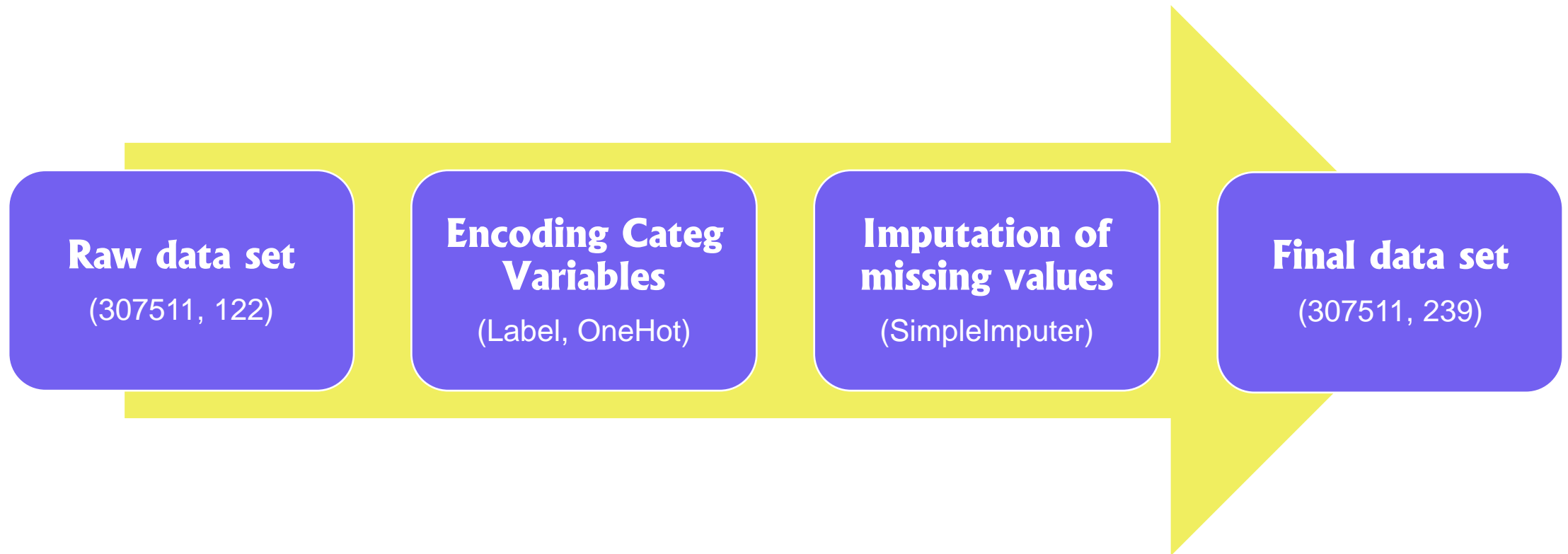


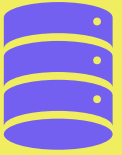
The data set





The data set – Features Engineering

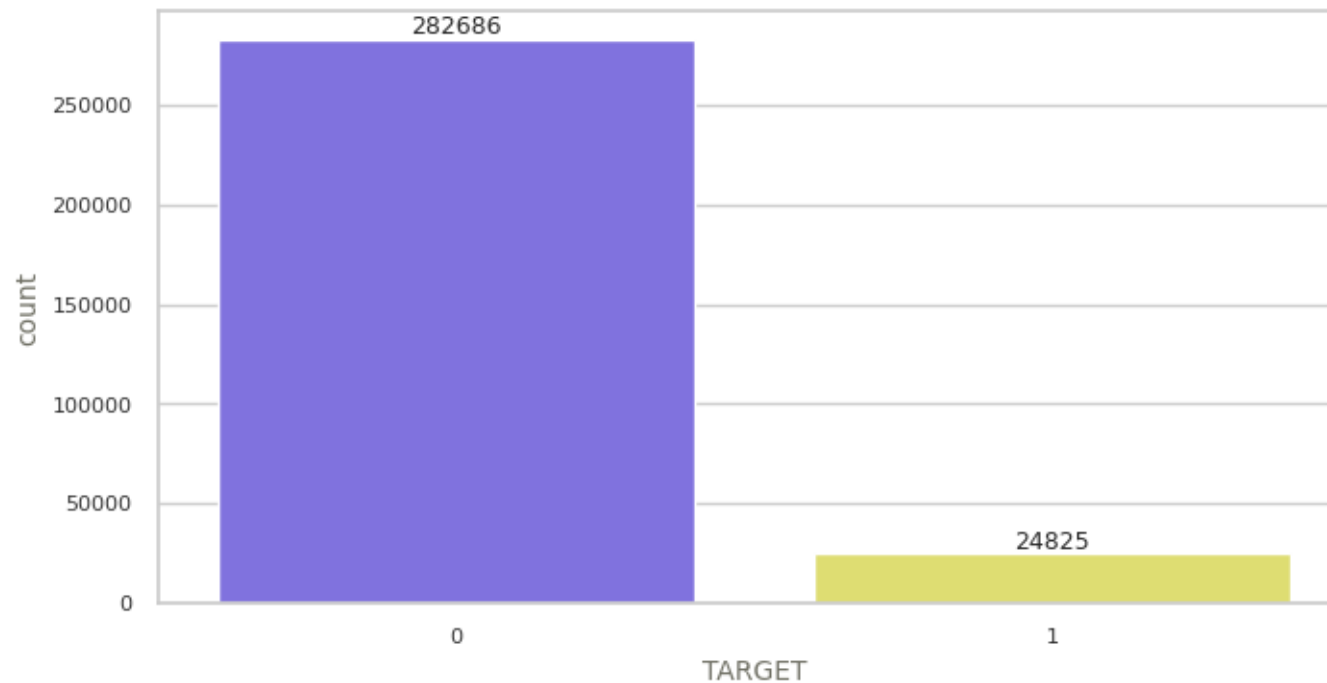




The data set - Target

0 le prêt a été remboursé **1** le prêt n'a pas été remboursé

Répartition des demandeurs de prêt en mesure de le rembourser





Modelling – Pipeline d'entraînement

Normalisation

(MinMaxScaler)

**Réduction de
dimension**

(RFE)

**Optimisation du
seuil de décision**

(TunedThresholdClassifierCV)

**Optimisation
des
hyperparamètres**

(GridSearchCV)



Modelling – Mesures

Fonction personnalisée pour refléter l'impact financier

$$custom_score = 1 * FP + 10 * FN$$

Métriques de performance

- **AUC** (Area Under the Curve) : Évalue la capacité globale du modèle à distinguer les classes, permettant d'optimiser le seuil de décision.
- **F1-Score** : Balance précision et rappel pour assurer une performance équilibrée, complémentaire à la minimisation des FN.



Modelling – Tracking Neptune.io

Experiments						
List All runs Search runs or build a query						
Id	Creation Time	Owner	Name	metrics/score_test_auc	...score_test_f1	...de_cout
OC-146	2024/10/30 10:11:27	remi.rogulski	LGBM_rawdata_rfe_dill_v1	75.509	26.942	-32909.3
OC-144	2024/10/29 22:41:58	remi.rogulski	LGBM_rawdata_rfe_dill_v1	75.509	26.942	-32909.3
OC-143	2024/10/29 21:01:43	remi.rogulski	LGBM_rawdata_rfe_dill_v1	75.509	26.942	-32909.3
OC-141	2024/10/29 17:13:17	remi.rogulski	LogisticRegression_rawdata_rfe_dill_v1	74.426	25.825	-33788.3
OC-136	2024/10/25 21:17:30	remi.rogulski	LGBM_rawdata_rfe_dill_v1	73.121	0	-4969
OC-135	2024/10/25 20:27:24	remi.rogulski	LogisticRegression_rawdata_rfe_dill_v1	73.573	0.06	-4975



Modelling – Synthèse des résultats

R

OC-144 X

All metadata

OC-144

OC-144 > metrics

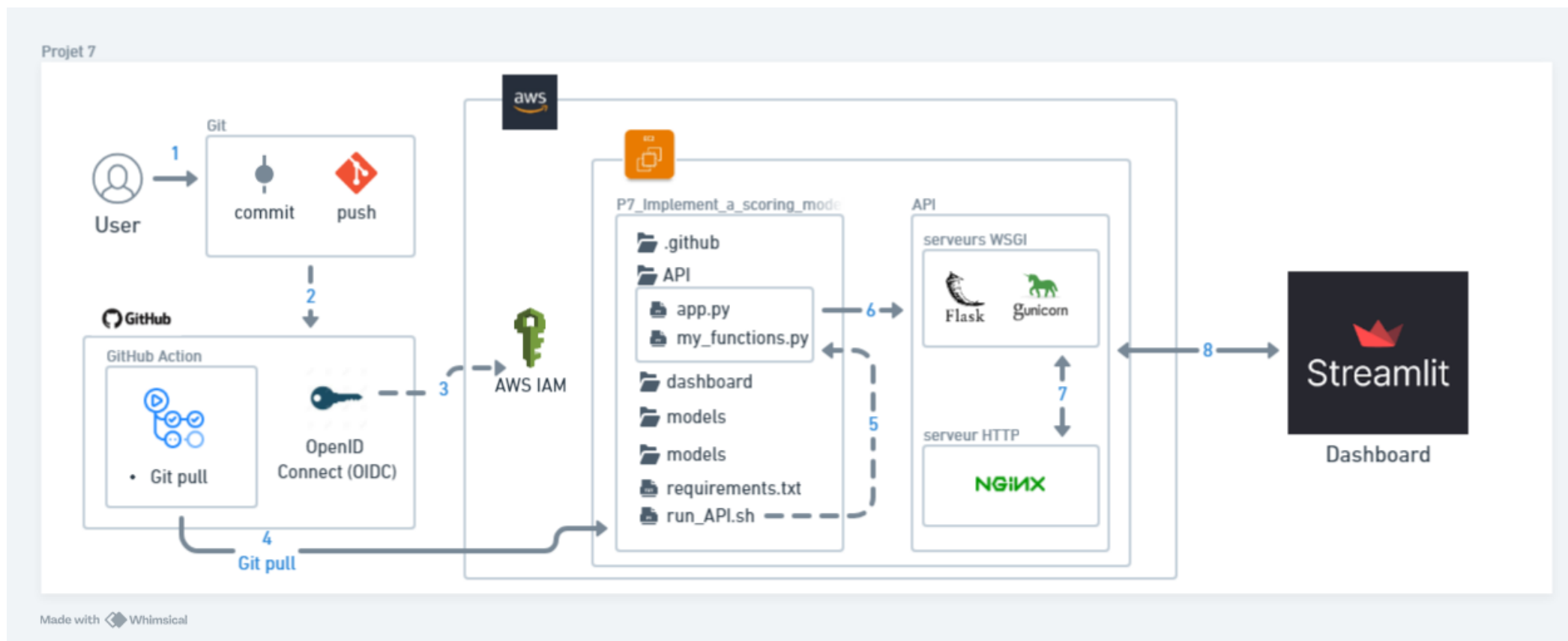
Search fields

Start typing to select fields...

NAME	PREVIEW
## best_Threshold	0.2
## best_threshold_Unite_cout	-31188.6
## best_Unite_de_cout	-32909.25
## score_test_auc	75.509
## score_test_f1	26.942
## score_train_auc	80.407
## score_train_f1	29.593



MLOps – Architecture





MLOps – Repository

P7_Implement_a_scoring_modelPublic

PinUnwatch1Fork0Star0

master1 BranchTags

Go to fileAdd fileCode

RR-DataSciencesMAJ du projet - Correction erreur dashboard.py8a6f9ea · 4 minutes ago185 Commits

.github/workflows	MAJ du projet - Ajout du notebook de modélisation et du ra...	18 minutes ago
API	MAJ du projet - Ajout du notebook de modélisation et du ra...	18 minutes ago
dashboard	MAJ du projet - Correction erreur dashboard.py	4 minutes ago
models	V3 - Add new models and explainer	2 months ago
tests	MAJ du projet - Ajout du notebook de modélisation et du ra...	18 minutes ago
.gitignore	Initial commit	4 months ago
README.md	MAJ du projet - Ajout du notebook de modélisation et du ra...	18 minutes ago
modelling.ipynb	MAJ du projet - Ajout du notebook de modélisation et du ra...	18 minutes ago
rapport_data_drift.html	MAJ du projet - Ajout du notebook de modélisation et du ra...	18 minutes ago
requirements.txt	MAJ du projet - Ajout du notebook de modélisation et du ra...	18 minutes ago
run_API.sh	MAJ du projet - Ajout du notebook de modélisation et du ra...	18 minutes ago

About

No description, website, or topics provided.

ReadmeActivity0 stars1 watching0 forks

Releases

No releases published
[Create a new release](#)


Packages

No packages published
[Publish your first package](#)

Languages

HTML 59.2%Jupyter Notebook 40.2%Other 0.6%

README



PRET A DEPENSER

[Projet 7] - Implémentez un modèle de scoring

Objectif du projet

Le projet Prêt à dépenser vise à développer un outil de scoring de crédit pour l'entreprise de prêt bancaire Prêt à dépenser. Cet outil permet de prédire la probabilité de remboursement d'un prêt par un client potentiel, ainsi que de visualiser et d'expliquer les décisions de crédit à l'aide de valeurs SHAP (SHapley Additive exPlanations). Le projet comprend un Dashboard interactif pour simuler des scénarios d'attribution de prêt, ainsi qu'une API pour effectuer des prédictions en temps réel.

Structure des dossiers

- API
 - app.py - Script principal pour le déploiement de l'API Flask.
 - my_functions.py - Regroupe les fonctions utilitaires nécessaires au prétraitement des données et à la gestion des modèles.
 - static/ - Stockez les images de fond et les éléments graphiques utilisés pour la page d'accueil de l'API.
- dashboard
 - dashboard.py - Fichier principal pour l'exécution de l'application Streamlit.
 - style.css - Fichier CSS pour la personnalisation de l'interface utilisateur.
 - images/ - Stockez les logos et autres images utilisés dans le Dashboard.
- models
- test
 - test_data.csv - Contient les données nécessaires pour tester l'application.
 - unite_test.py - Regroupe les fonctions de tests unitaires appliquées lors du déploiement.
- modelling.ipynb - Notebook de modélisation.
- rapport_data_drift.html - Rapport d'analyse du drift sur les données app_train et app test.
- requirements.txt - Liste des dépendances nécessaires pour exécuter le projet.
- run_API.sh - Fichier permettant le lancement de l'API.



MLOps – Commits

Commits		
master		
All users		
All time		
Commits on Oct 27, 2024		
MAJ du projet - Correction erreur dashboard.py		
RR-DataSciences committed 6 minutes ago	0 / 1	8a6f9ea
MAJ du projet - Correction erreur dashboard.py		
RR-DataSciences committed 10 minutes ago	1 / 1	1cce332
MAJ du projet - Ajout du notebook de modélisation et du rapport data drift		
RR-DataSciences committed 20 minutes ago	1 / 1	c4f662a
Commits on Oct 24, 2024		
MAJ du README		
RR-DataSciences committed 3 days ago	0 / 1	5e27ee7
MAJ du README		
RR-DataSciences committed 3 days ago	0 / 1	169257d
MAJ du README		
RR-DataSciences committed 3 days ago	0 / 1	de47495
MAJ du README		
RR-DataSciences committed 3 days ago	0 / 1	c45dc6f
MAJ du README		
RR-DataSciences committed 3 days ago	0 / 1	818d5b4
MAJ du README		
RR-DataSciences committed 3 days ago	0 / 1	00w29ba
MAJ du README		
RR-DataSciences committed 3 days ago	0 / 1	fc4dbcb
MAJ du README		
RR-DataSciences committed 3 days ago	0 / 1	e9933c4
MAJ du README		
RR-DataSciences committed 3 days ago	0 / 1	c80fbad
Ajout du README		
RR-DataSciences committed 3 days ago	0 / 1	23f8993
Ajout du README		
RR-DataSciences committed 3 days ago		d094c02
Commits on Oct 23, 2024		
Dashboard - Placement des éléments dans le container		
RR-DataSciences committed 4 days ago	0 / 1	ad14d13
Dashboard - Orga page prediction		
RR-DataSciences committed 4 days ago	0 / 1	ab23d13
Dashboard - couleur message d'indication		
RR-DataSciences committed 4 days ago	0 / 1	487e627
Commits on Oct 22, 2024		

Commits on Oct 22, 2024		
API - Logo		
RR-DataSciences committed 5 days ago	0 / 1	325f138
API - Logo		
RR-DataSciences committed 5 days ago	1 / 1	2ba28ad
API - Logo		
RR-DataSciences committed 5 days ago	1 / 1	5b275a4
API - Logo		
RR-DataSciences committed 5 days ago	1 / 1	9da782d
API - Logo		
RR-DataSciences committed 5 days ago	1 / 1	b230b91
API - Logo		
RR-DataSciences committed 5 days ago	1 / 1	30bd7c2
API - Logo		
RR-DataSciences committed 5 days ago	1 / 1	e42cce1
API - Logo		
RR-DataSciences committed 5 days ago	1 / 1	89b9727
API - Logo		
RR-DataSciences committed 5 days ago	0 / 1	a7097cf
API - Logo		
RR-DataSciences committed 5 days ago	1 / 1	3aa7f05
API - Logo		
RR-DataSciences committed 5 days ago	0 / 1	ad11452
API - Logo		
RR-DataSciences committed 5 days ago	0 / 1	e40d8fe
API - Logo		
RR-DataSciences committed 5 days ago	0 / 1	aa41f76
API - Logo		
RR-DataSciences committed 5 days ago	1 / 1	1eeccf0
Dashboard - Logo		
RR-DataSciences committed 5 days ago	0 / 1	a26f245
Dashboard - Logo		
RR-DataSciences committed 5 days ago	0 / 1	f611fd4
Dashboard - Logo		
RR-DataSciences committed 5 days ago	1 / 1	617b603
Dashboard - Logo		
RR-DataSciences committed 5 days ago	1 / 1	5442423
Previous Next		



MLOps – GitHub Actions

RR-DataSciences / P7_Implement_a_scoring_model

<> Code Issues Pull requests **Actions** Projects Wiki Security Insights Settings

← Déploiement sur EC2 AWS

✓ MAJ du projet - Correction erreur dashboard.py #167 Re-run all jobs

Summary

Jobs

✓ MajPullOnEC2

Run details

Usage

Workflow file

Annotations
1 warning

MajPullOnEC2
succeeded 2 hours ago in 4m 28s

Search logs

- > ✓ Set up job 3s
- > ✓ Git clone the repository 1s
- > ✓ configure aws credentials 0s
- > ✓ Sts GetCallerIdentity 4s
- > ✓ Setup SSH 1s
- > ✓ Connect & Execute commands on EC2 1s
- > ✓ Run unit tests 4s
- > ✓ Deploy API 4m 7s
- > ✓ Post configure aws credentials 0s
- > ✓ Post Git clone the repository 0s
- > ✓ Complete job 0s



MLOps – Unit testing

```
Run unit tests 4s

1 ▶ Run ssh -o StrictHostKeyChecking=no -i ~/.ssh/id_rsa ec2-user@ec2-54-171-37-232.eu-west-1.compute.amazonaws.com << EOF
14 Pseudo-terminal will not be allocated because stdin is not a terminal.
15 A newer release of "Amazon Linux" is available.
16   Version 2023.5.20240730:
17   Version 2023.5.20240805:
18   Version 2023.5.20240819:
19   Version 2023.5.20240903:
20   Version 2023.5.20240916:
21   Version 2023.5.20241001:
22   Version 2023.6.20241010:
23 Run "/usr/bin/dnf check-release-update" for full release and version update info
24   ,      #_
25   ~\     #####_      Amazon Linux 2023
26   ~ ~   \_#####\
27   ~ ~   \###|
28   ~ ~   \#/  ____  https://aws.amazon.com/linux/amazon-linux-2023
29   ~ ~   V~' '->
30   ~ ~      /
31   ~ ~. _  _/
32   ~ ~  _/ _/
33   ~ ~ _/m/'
34   ....
35   -----
36
37 Ran 4 tests in 0.052s
38 OK
39 Your selected dataframe has 2 columns.
40 There are 2 columns that have missing values.
41 Description avant RFE : (1, 239)
42 Description après RFE : (1, 71)
43 Résultats de prédiction : [0], Score : [[0.30213113 0.69786887]]
44 Description des valeurs SHAP : (1, 71)
```



Data Drift

Dataset Drift

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

240
Columns

10
Drifted Columns

0.0417
Share of Drifted Columns

Data Drift Summary

Drift is detected for 4.167% of columns (10 out of 240).

Search ×						
Column	Type	Reference Distribution	Current Distribution	Data Drift	Stat Test	Drift Score
> DAYS_BIRTH	num			Detected	Wasserstein distance (normed)	7.356832
> 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
> AMT_ANNUITY	num			Detected	Wasserstein distance (normed)	0.161102



PRÊT À DÉPENSER

Demo

Objectives and context 

The data set 

Modelling 

MLOps 

Data Drift 

Demonstration 