

# Projet 7

Implémentez un modèle de scoring

Alexandre ROGUES - Parcours DataScientist - OpenClassRooms

# Objectif du projet

- Développement d'un modèle de scoring de crédit
- Déploiement en production

# Contexte

## L'entreprise

- augmenter la part d'individus bénéficiaires de crédits
- minimiser le risque pour l'entreprise

## Le besoin

- **automatiser la décision :**
  - réduire le délai de réponse,
  - diminuer les ressources nécessaires pour produire la réponse
- **augmenter la précision :**
  - prendre en compte des données comportementales, et issues d'autres sources



## Objectif

- **réduire les pertes**
  - produire des prévisions fiables
- **accroître la transparence**
  - fournir des explications claires sur la décision

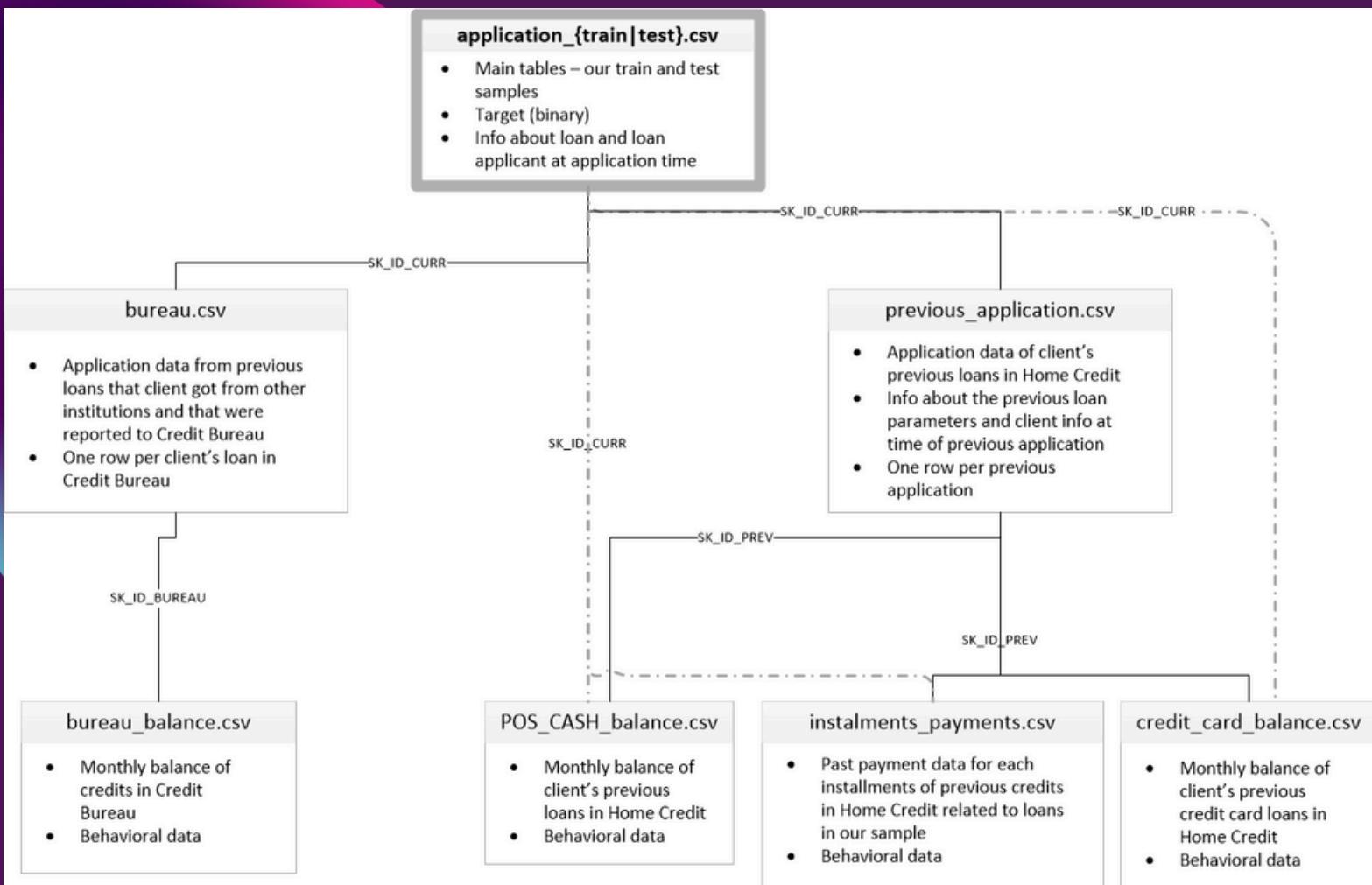
**train set**  
307511 individus  
122 colonnes

**test set**  
48744 individus  
121 colonnes

Dataset issu d'une  
compétition Kaggle

# Le jeu de données

7 fichiers csv, environ 4Go



Données individuelles

Age, genre, famille, biens

...

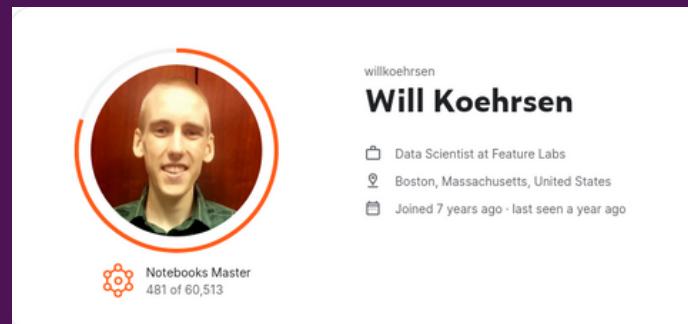
Données financières

revenus, endettements,  
autres établissements,

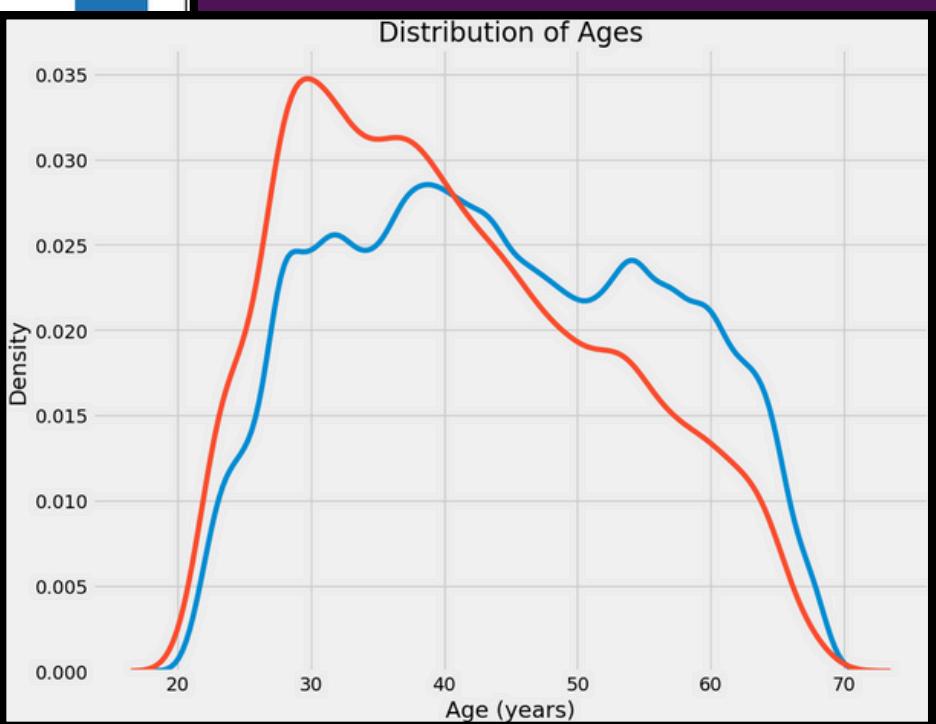
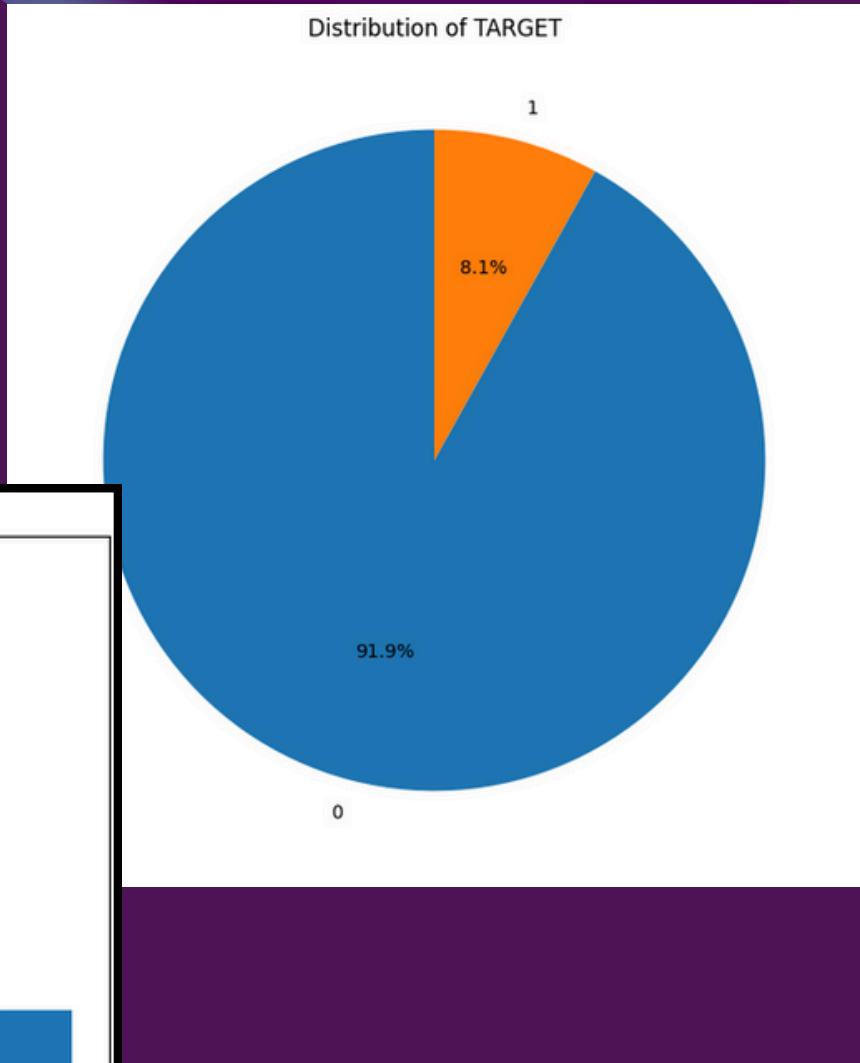
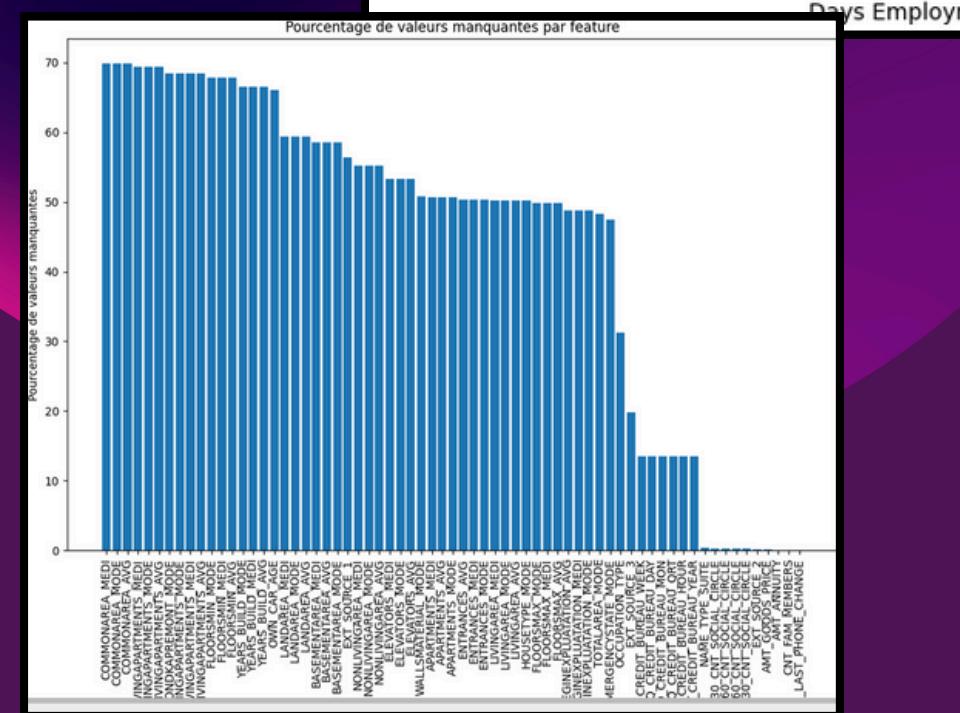
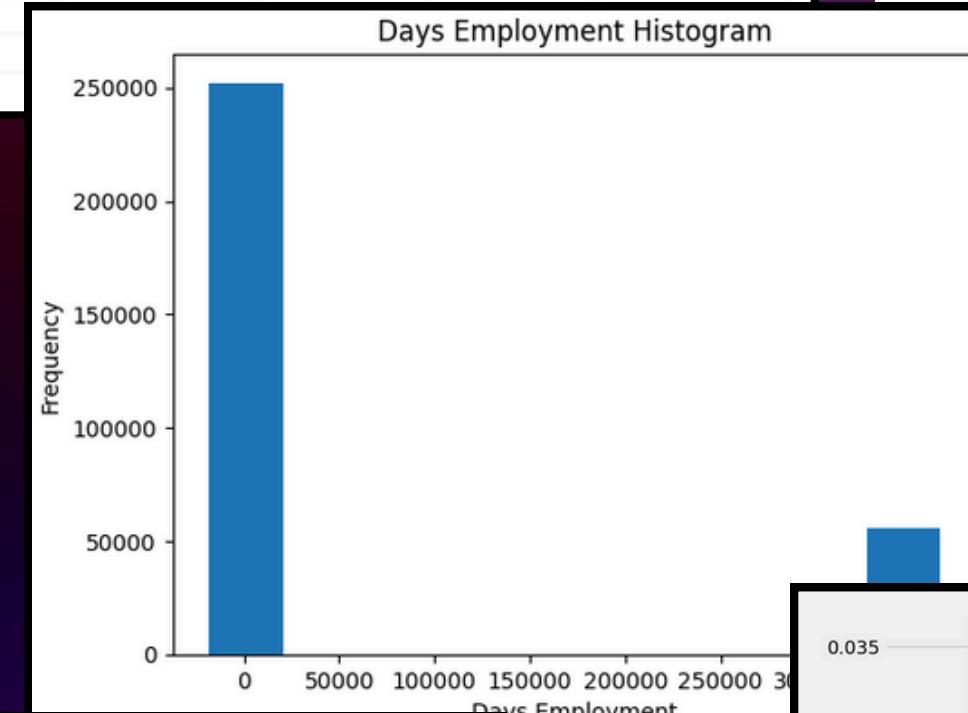
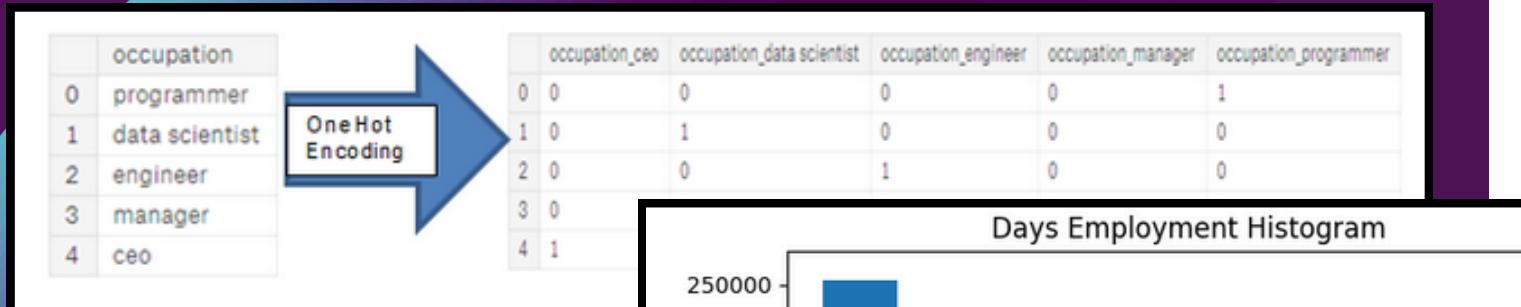
# Préparation des données

## Analyse exploratoire

*appui sur Kaggle existant*



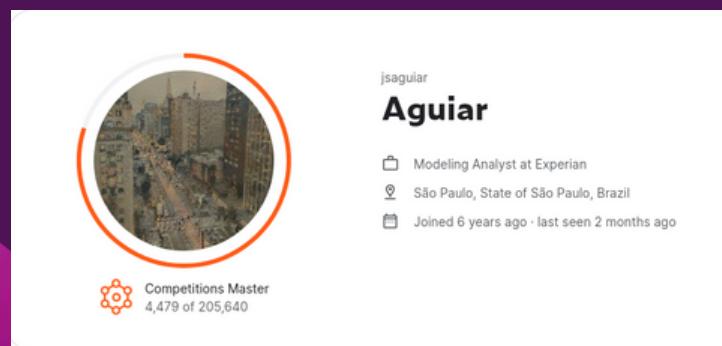
- Unbalanced dataset
- Valeurs manquantes
- One-hot encoding
- Valeurs absurdes
- Recherche corrélations



# Préparation des données

## Feature engineering

*appui sur Kaggle existant*



- Variables polynomiales
- Variables métier
- Intégration des autres dataframes

CREDIT\_INCOME\_PERCENT : le pourcentage du montant du crédit par rapport aux revenus d'un client  
ANNUITY\_INCOME\_PERCENT : le pourcentage de la rente du prêt par rapport au revenu d'un client  
CREDIT\_TERM : la durée du paiement en mois (puisque la rente est le montant mensuel dû)  
DAYS\_EMPLOYED\_PERCENT : le pourcentage de jours employés par rapport à l'âge du client

```
# Pourcentage passé à travailler de la vie d'une personne  
df['DAYS_EMPLOYED_PERC'] = df['DAYS_EMPLOYED'] / df['DAYS_BIRTH']  
  
# Rapport revenu total d'un individu et montant total du crédit  
df['INCOME_CREDIT_PERC'] = df['AMT_INCOME_TOTAL'] / df['AMT_CREDIT']  
  
# Revenu moyen par membre de la famille  
df['INCOME_PER_PERSON'] = df['AMT_INCOME_TOTAL'] / df['CNT_FAM_MEMBERS']  
  
# Rapport entre montant annuité et le revenu total d'un individu  
df['ANNUITY_INCOME_PERC'] = df['AMT_ANNUITY'] / df['AMT_INCOME_TOTAL']  
  
# Proportion du crédit remboursée annuellement  
df['PAYMENT_RATE'] = df['AMT_ANNUITY'] / df['AMT_CREDIT']
```

# features

120 => 240 => 765

# Stratégie de modélisation

## Train/Test Split

- 80/20
- stratify sur Target

## Imputation

- KNN
- 3 neighbors

## Normalisation

- Standard Scaler

## Gestion déséquilibre

- SMOTE 0.3

**pipeline**

# Modélisation

## Dummy Classifier

- **Description:**

Un modèle de base utilisé pour comparer les performances de modèles plus complexes.

- **Fonctionnement:**

Fait des prédictions simples basées sur des règles triviales comme la distribution des classes dans les données d'entraînement.

- **Utilisation:**

Sert de référence pour évaluer la pertinence des autres modèles.

- **Avantage:**

Facile à implémenter, offre une ligne de base pour les performances.

## Régression Linéaire

- **Description:**

Modèle statistique qui prédit une valeur continue en trouvant la relation linéaire entre les variables indépendantes et la variable dépendante.

- **Fonctionnement:**

Minimise la somme des carrés des différences entre les prédictions et les valeurs réelles.

- **Utilisation:**

Convient pour les problèmes de régression où la relation entre les variables est linéaire.

- **Avantage:**

Simple, interprétable, et efficace pour des relations linéaires.

## Random Forrest

- **Description:**

Modèle d'ensemble basé sur la construction de multiples arbres de décision indépendants et leur agrégation.

- **Fonctionnement:**

Utilise un échantillonnage bootstrap et une sélection aléatoire de features pour chaque arbre, puis agrège les prédictions par vote majoritaire (classification) ou moyenne (régression).

- **Utilisation:**

Efficace pour les problèmes de classification et de régression avec des données complexes et non linéaires.

- **Avantage:**

Réduit le risque de surapprentissage et améliore la généralisation grâce à la diversité des arbres.

## XG Boost

- **Description:**

Implémentation avancée du gradient boosting qui optimise les performances des modèles d'ensemble en utilisant des arbres de décision.

- **Fonctionnement:**

Ajoute des arbres séquentiellement, chaque arbre corrigeant les erreurs des arbres précédents, avec une régularisation pour éviter le surapprentissage.

- **Utilisation:**

Très performant pour les compétitions de machine learning et les problèmes avec de grandes quantités de données et des relations complexes.

- **Avantage:**

Efficace, flexible, et capable de gérer les interactions complexes entre les features, avec des options de régularisation avancées.

# Modélisation

## Dummy Classifier

AUC: 0.5  
Accuracy: 0.92  
Cost: 490

## Régression Linéaire

AUC: 0.60  
Accuracy: 0.71  
Cost: 417

## Random Forrest

AUC: 0.63  
Accuracy: 0.91  
Cost: 491

## XG Boost

AUC: 0.61  
Accuracy: 0.91  
Cost: 476

Résultat sur sample 0.15

# Modélisation

## Dummy Classifier

AUC: 0.5  
Accuracy: 0.92  
Cost: 7480

## Régression Linéaire

AUC: 0.73  
Accuracy: 0.70  
Cost: 5158

## Random Forrest

AUC: 0.68  
Accuracy: 0.91  
Cost: 7473

## XG Boost

AUC: 0.71  
Accuracy: 0.91  
Cost: 7309

Résultat sur sample 0.15

# make-scorer

## Objectif :

Evaluer les modèles en tenant compte des coûts spécifiques des erreurs

## Coût des erreurs:

- Faux Négatif (FN): Mauvais client prédit bon client (crédit accordé à un client risqué).
  - Coût élevé: Perte en capital (10x le coût d'un FP)
- Faux Positif (FP): Bon client prédit mauvais (crédit refusé à un bon client).
  - Coût modéré: Manque à gagner en marge

```
from sklearn.metrics import make_scorer, confusion_matrix

# Définir une fonction pour calculer le coût
def custom_cost(y_true, y_pred):
    conf_matrix = confusion_matrix(y_true, y_pred)
    fp = conf_matrix[0][1]
    fn = conf_matrix[1][0]
    cost = fp + 10 * fn
    return cost

# Créer un make_scorer à partir de la fonction custom_cost
cost_scorer = make_scorer(custom_cost, greater_is_better=False)
```

# GridSearchCV

```
Fitting 3 folds for each of 27 candidates, totalling 81 fits
Best Hyperparameters: {'xgb_learning_rate': 0.01, 'xgb_max_depth': 9, 'xgb_n_estimators': 300}
Best model saved to ../models/best_pipeline_model.pkl
Best Model - AUC: 0.637412562198024
Best Model - Accuracy: 0.9235772357723577
Best Model - Cost: 470
```

## Fonctionnement du Pipeline

### 1. Standardisation des Données

- **StandardScaler:** Met les caractéristiques à une moyenne de 0 et un écart-type de 1.
- **Objectif:** Améliorer la convergence et les performances des algorithmes.

### 2. Modèle de Classification

- **XGBoost Classifier:** Algorithme de gradient boosting pour la classification binaire.
- **Paramètres Testés:**
  - Taux d'apprentissage (learning\_rate)
  - Profondeur maximale des arbres (max\_depth)
  - Nombre d'estimateurs (n\_estimators)

pourquoi utiliser un pipeline ?

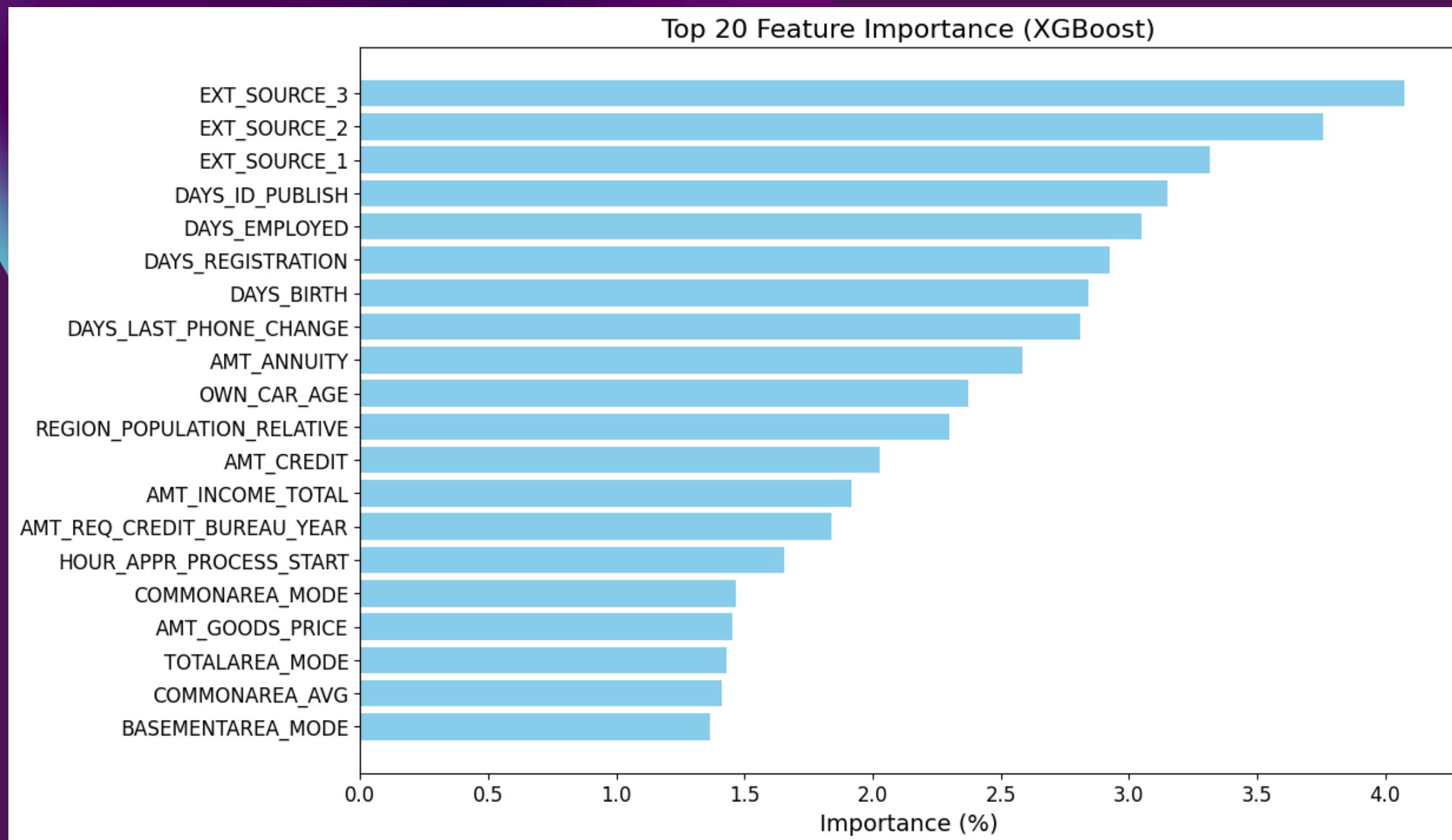
Lors de la validation croisée, il est crucial que les étapes de prétraitement soient ajustées uniquement sur les données d'entraînement et non sur les données de validation. Les pipelines assurent cette séparation.

## Avantages de l'Approche

- **Optimisation Automatisée:** GridSearchCV teste plusieurs combinaisons de paramètres.
- **Log Tracking avec MLflow:** Enregistre les expérimentations et résultats pour la traçabilité.
- **Sauvegarde et Déploiement:** Enregistre le meilleur modèle pour une utilisation future.

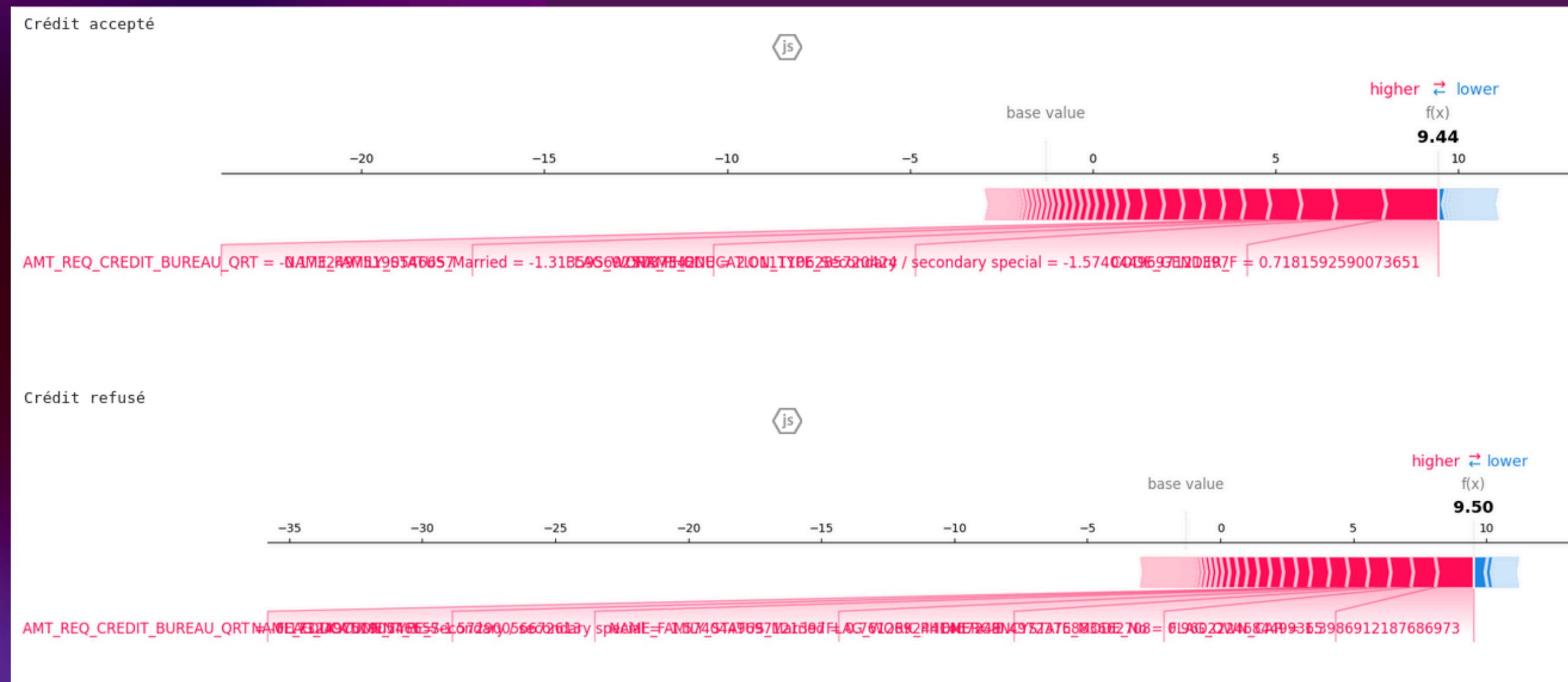
# shap

## importances globales



# shap

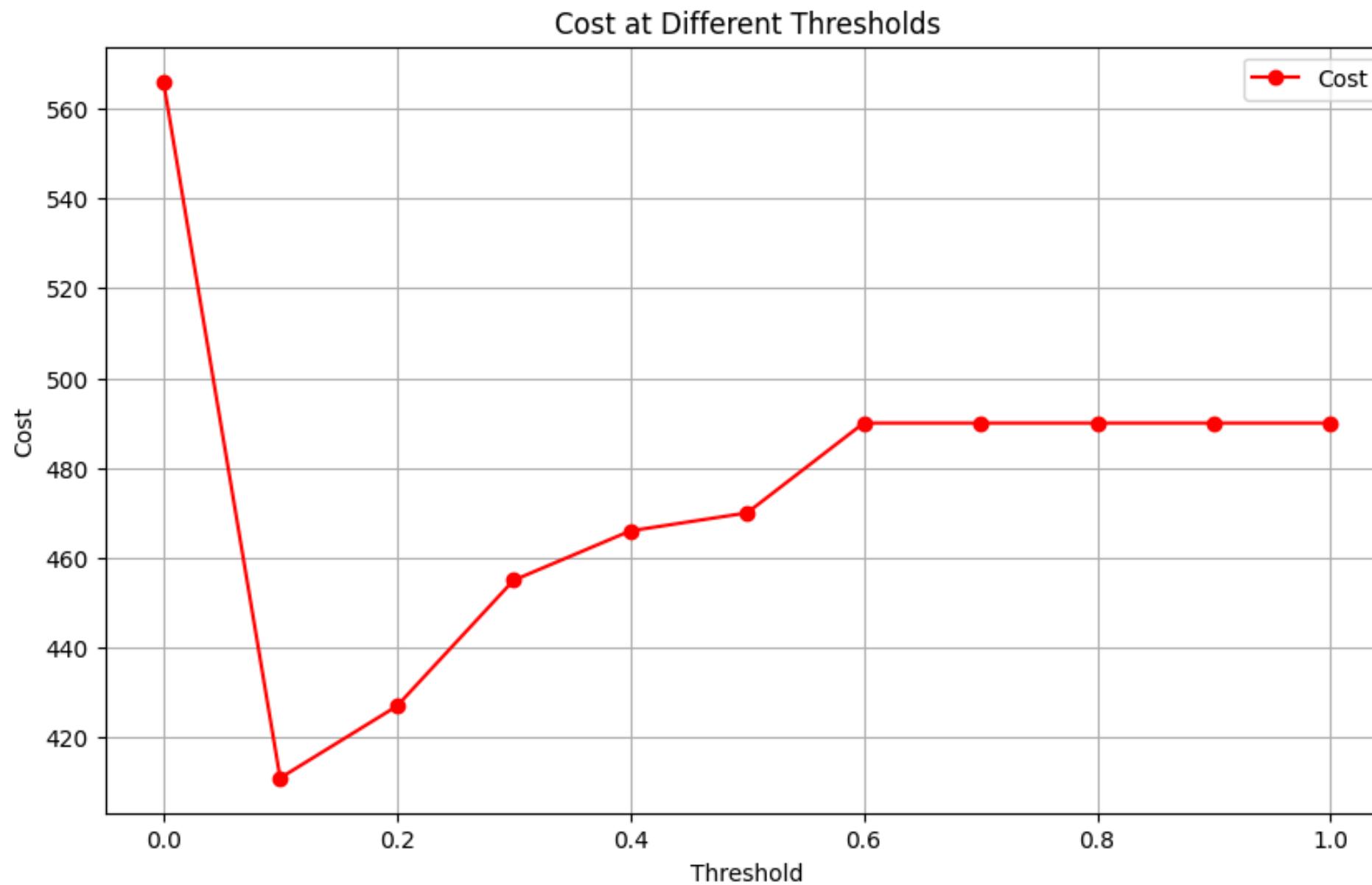
## importances locales



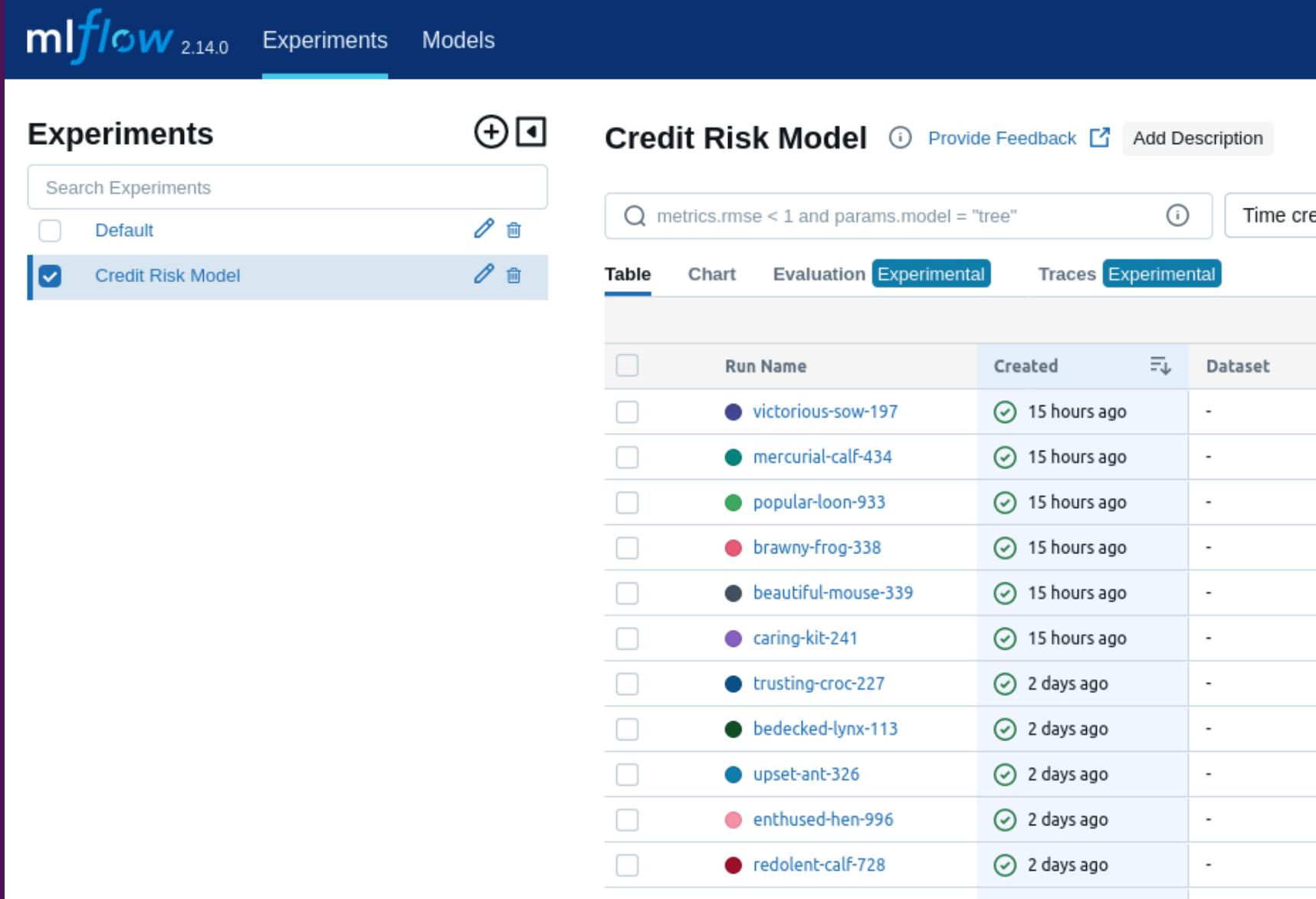
# Optimisation du seuil en fonction du coût

recherche du meilleur seuil

0.1 semble minimiser le coût



# Tracking MLFlow



The screenshot shows the MLflow UI interface. At the top, there's a navigation bar with the MLflow logo, version 2.14.0, and links for 'Experiments' and 'Models'. Below this is the 'Experiments' section, which includes a search bar and two filter buttons: 'Default' and 'Credit Risk Model' (which is checked). A table lists ten runs under the 'Experimental' tab, each with a run name, creation time (ranging from 15 hours ago to 2 days ago), and a dataset column showing a dash. The table has columns for 'Run Name', 'Created', and 'Dataset'.

	Run Name	Created	Dataset
1	victorious-sow-197	15 hours ago	-
2	mercurial-calf-434	15 hours ago	-
3	popular-loon-933	15 hours ago	-
4	brawny-frog-338	15 hours ago	-
5	beautiful-mouse-339	15 hours ago	-
6	caring-kit-241	15 hours ago	-
7	trusting-croc-227	2 days ago	-
8	bedecked-lynx-113	2 days ago	-
9	upset-ant-326	2 days ago	-
10	enthused-hen-996	2 days ago	-
11	redolent-calf-728	2 days ago	-

```
(stabadevP7) alexandre@alexandre-desktop:~/Documents/DataScientist/Projet7/notebooks$ mlflow ui  
[2024-07-05 09:06:43 +0200] [5296] [INFO] Starting gunicorn 22.0.0  
[2024-07-05 09:06:43 +0200] [5296] [INFO] Listening at: http://127.0.0.1:5000 (5296)  
[2024-07-05 09:06:43 +0200] [5296] [INFO] Using worker: sync  
[2024-07-05 09:06:43 +0200] [5297] [INFO] Booting worker with pid: 5297  
[2024-07-05 09:06:43 +0200] [5298] [INFO] Booting worker with pid: 5298  
[2024-07-05 09:06:43 +0200] [5299] [INFO] Booting worker with pid: 5299  
[2024-07-05 09:06:43 +0200] [5300] [INFO] Booting worker with pid: 5300  
[2024-07-05 09:08:17 +0200] [5296] [INFO] Handling signal: winch
```

Enregistrement des hyperparamètres pour assurer la reproductibilité

VPS chez OVH  
install Debian + Yunohost  
subdomain projet7.rogues.fr  
ssl + dns  
install my\_webapp module  
config nginx  
gunicorn

# API de prédition

Bienvenue dans l'interface de prédition de remboursement

Client ID:

Predict

Prediction: 0

Global Feature Importances

Feature	Description	Importance
CODE_GENDER_F	nan	0.0489999871850014
FLAG_DOCUMENT_3	Did client provide document 3	0.0480000041723251
EMERGENCYSTATE_MODE_No	nan	0.04500000178813934
NAME_EDUCATION_TYPE_Higher education	nan	0.03599998450279236
AMT_REQ_CREDIT_BUREAU_DAY	Number of enquiries to Credit Bureau about the client one day before application (excluding one hour before application)	0.03500000014901161

Local Feature Importances

Feature	Description	Importance
OCCUPATION_TYPE_Cleaning staff	nan	0.32499998807907104
WEEKDAY_APPR_PROCESS_START_TUESDAY	nan	0.2879999876022339
BASEMENTAREA_MODE	Normalized information about building where the client lives, What is average ( AVG suffix) modus ( MODE suffix) median ( MEDIAN suffix)	0.240999966621399

VPS chez OVH

install Debian + Yunohost

subdomain projet7.rogues.fr

ssl + dns

install my\_webapp module

config nginx

gunicorn

# API de prédition

The screenshot shows the OVH Cloud Control Panel interface. On the left, there's a sidebar with various service links like Bare Metal Cloud, Hosted Private Cloud, Public Cloud, etc. The main area displays the configuration of a VPS instance named "VPS Debian Yunohost pour rogues.fr". It includes sections for "Votre VPS" (Status: Actif), "Votre configuration" (Model: VPS vps2020-value-1-2-40, vCores: 1, Memory: 2 Go, Storage: 40 Go), and "IP" (IPv4: 137.74.198.208). There are also tabs for "Accueil", "DNS Secondaire", "Backup automatisé", and "Disque additionnel". A message at the top indicates the instance is associated with a free IPv6.

This screenshot shows the "Zone DNS actuelle" (Current DNS Zone) interface. It lists all the DNS records for the domain "projet7.rogues.fr". The records include:

Type	Name	Value
A	*	137.74.198.208
A	dataScientist	137.74.198.208
A	unicorn	137.74.198.208
A	projet7	137.74.198.208
A	@	137.74.198.208
A	streamlit	137.74.198.208
A	testraw	137.74.198.208
AAAA	*	2001:41d0:401:3000::5535
AAAA	dataScientist	2001:41d0:401:3000::5535
AAAA	unicorn	2001:41d0:401:3000::5535
AAAA	projet7	2001:41d0:401:3000::5535
AAAA	@	2001:41d0:401:3000::5535
AAAA	streamlit	2001:41d0:401:3000::5535
AAAA	testraw	2001:41d0:401:3000::5535
CAA	@	0 issue "letsencrypt.org"
CNAME	muc	rogues.fr.
CNAME	pubsub	rogues.fr.
CNAME	vjud	rogues.fr.
CNAME	xmpp-upload	rogues.fr.
CNAME	dataScientist	rogues.fr.
CNAME	unicorn	rogues.fr.
CNAME	projet7	rogues.fr.
MX	@	10 dataScientist.rogues.fr.
MX	streamlit	10 unicorn.rogues.fr.
MX	testraw	10 projet7.rogues.fr.
MX	streamlit	10 rogues.fr.
MX	testraw	10 streamlit.rogues.fr.
MX	streamlit	10 testraw.rogues.fr.
SRV	_xmpp-client._tcp	0 5 5222 rogues.fr.
SRV	_xmpp-server._tcp	0 5 5269 rogues.fr.
TXT	_dmarc.dataScientist	"v=DMARC1; p=none"
TXT	_dmarc.unicorn	"v=DMARC1; p=none"
TXT	_dmarc.projet7	"v=DMARC1; p=none"
TXT	_dmarc	"v=DMARC1; p=none"
TXT	_dmarc.streamlit	"v=DMARC1; p=none"
TXT	_dmarc.testraw	"v=DMARC1; p=none"
TXT	dataScientist	"v=spf1 a mx -all"
TXT	unicorn	"v=spf1 a mx -all"
TXT	mail._domainkey.dataScientist	"v=DKIM1; h=sha256; k=rsa; p=MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBi
TXT	mail._domainkey.unicorn	"v=DKIM1; h=sha256; k=rsa; p=MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBi
TXT	mail._domainkey.projet7	"v=DKIM1; h=sha256; k=rsa; p=MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBi
TXT	mail._domainkey	"v=DKIM1; h=sha256; k=rsa; p=MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBi
TXT	mail._domainkey.streamlit	"v=DKIM1; h=sha256; k=rsa; p=MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBi
TXT	mail._domainkey.testraw	"v=DKIM1; h=sha256; k=rsa; p=MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBi
TXT	projet7	"v=spf1 a mx -all"
TXT	@	"v=spf1 a mx -all"
TXT	streamlit	"v=spf1 a mx -all"
TXT	testraw	"v=spf1 a mx -all"

This screenshot shows the domain management interface for "projet7.rogues.fr". It includes sections for "Lien" (Link: https://projet7.rogues.fr), "Autorité de certification SSL" (SSL Certificate Authority: Let's Encrypt, valid until 84 days), "Registre de nom de domaine" (Domain Name Registry: Voir le domaine parent rogues.fr), and "Applications installées sur le domaine" (Installed Applications: My Webapp). Below these, there are sections for "Features", "DNS", and "Certificate". A green box says "Bravo ! Vous utilisez un certificat Let's Encrypt valide !" (Bravo ! You are using a valid Let's Encrypt certificate!). A yellow box says "Renouvellement du certificat Let's Encrypt" (Renewal of the Let's Encrypt certificate).

```
ssh -i id_rsa arogues@137.74.198.208
```

VPS chez OVH

install Debian + Yunohost

subdomain projet7.rogues.fr

ssl + dns

install my\_webapp module

config nginx

gunicorn

# API de prédition

## clone repo

```
arogues@rogues:/var/www/my_webapp__2/www/Projet7$ ls -ll
total 700
drwxr-xr-x 2 arogues arogues 4096 Jul  4 10:41 csv_files
drwxr-xr-x 4 arogues arogues 4096 Jul  4 10:41 misc
drwxr-xr-x 4 arogues arogues 4096 Jul  4 10:41 mlruns
drwxr-xr-x 2 arogues arogues 4096 Jul  4 10:41 models
drwxr-xr-x 4 arogues arogues 4096 Jul  4 10:41 notebooks
-rw-r--r-- 1 arogues arogues    26 Jul  4 10:41 Procfile
-rw-r--r-- 1 arogues arogues 12891 Jul  4 10:41 README.md
-rw-r--r-- 1 arogues arogues     48 Jul  4 10:41 requirements.txt
drwxr-xr-x 4 arogues arogues 4096 Jul  4 10:42 src
drwxr-xr-x 2 arogues arogues 4096 Jul  4 10:41 templates
-rw-r--r-- 1 arogues arogues 659726 Jul  4 10:41 tree.txt
arogues@rogues:/var/www/my_webapp__2/www/Projet7$ 
```

VPS chez OVH

install Debian + Yunohost

subdomain projet7.rogues.fr

ssl + dns

install my\_webapp module

config nginx

gunicorn

# API de prédition

```
rogues@rogues:/etc/nginx/conf.d/projet7.rogues.fr.d/my_webapp_2.d$ cat projet7.conf
location /site/projet7/ {
    proxy_pass http://127.0.0.1:8000; # Port où Gunicorn écoute
    proxy_set_header Host $host;
    proxy_set_header X-Real-IP $remote_addr;
    proxy_set_header X-Forwarded-For $proxy_add_x_forwarded_for;
    proxy_set_header X-Forwarded-Proto $scheme;

    # WebSocket support
    proxy_http_version 1.1;
    proxy_set_header Upgrade $http_upgrade;
    proxy_set_header Connection "upgrade";
}
```

# nginx

VPS chez OVH

install Debian + Yunohost

subdomain projet7.rogues.fr

ssl + dns

install my\_webapp module

config nginx

gunicorn

# API de prédition

```
arogues@rogues:/etc/systemd/system$ cat my_webapp.service
[Unit]
Description=Gunicorn instance to serve my_webapp
After=network.target

[Service]
User=arogues
Group=arogues
WorkingDirectory=/var/www/my_webapp_2/www/Projet7/src
ExecStart=/home/arogues/.local/bin/gunicorn --workers 3 --bind 0.0.0.0:8000 appV2:app

[Install]
WantedBy=multi-user.target
```

## my\_webapp.service

puis exécution des instructions suivantes

**sudo systemctl daemon-reload**

**sudo systemctl restart my\_webapp.service**

**sudo systemctl status my\_webapp.service**

```
arogues@rogues:/var/www/my_webapp_2/www$ sudo systemctl status my_webapp.service
[sudo] password for arogues:
● my_webapp.service - Gunicorn instance to serve my_webapp
   Loaded: loaded (/etc/systemd/system/my_webapp.service; enabled; vendor preset: enabled)
   Active: active (running) since Mon 2024-07-08 12:10:29 UTC; 9h ago
     Main PID: 828653 (gunicorn)
        Tasks: 7 (limit: 2301)
       Memory: 514.2M
          CPU: 16.715s
        CGroup: /system.slice/my_webapp.service
                └─828653 /usr/bin/python3 /home/arogues/.local/bin/gunicorn --workers 3 --bind 0.0.0.0:8000 appV2:app

Jul 08 12:10:29 rogues.fr systemd[1]: Started Gunicorn instance to serve my_webapp.
Jul 08 12:10:29 rogues.fr gunicorn[828653]: [2024-07-08 12:10:29 +0000] [828653] [INFO] Starting gunicorn 22.0.0
Jul 08 12:10:29 rogues.fr gunicorn[828653]: [2024-07-08 12:10:29 +0000] [828653] [INFO] Listening at: http://0.0
Jul 08 12:10:29 rogues.fr gunicorn[828653]: [2024-07-08 12:10:29 +0000] [828653] [INFO] Using worker: sync
Jul 08 12:10:29 rogues.fr gunicorn[828655]: [2024-07-08 12:10:29 +0000] [828655] [INFO] Booting worker with pid:
Jul 08 12:10:29 rogues.fr gunicorn[828656]: [2024-07-08 12:10:29 +0000] [828656] [INFO] Booting worker with pid:
Jul 08 12:10:29 rogues.fr gunicorn[828657]: [2024-07-08 12:10:29 +0000] [828657] [INFO] Booting worker with pid:
```

VPS chez OVH

install Debian + Yunohost

subdomain projet7.rogues.fr

ssl + dns

install my\_webapp module

config nginx

gunicorn

# API de prédition

```
arogues@rogues:/etc/systemd/system$ cat my_webapp.service
[Unit]
Description=Gunicorn instance to serve my_webapp
After=network.target

[Service]
User=arogues
Group=arogues
WorkingDirectory=/var/www/my_webapp_2/www/Projet7/src
ExecStart=/home/arogues/.local/bin/gunicorn --workers 3 --bind 0.0.0.0:8000 appV2:app

[Install]
WantedBy=multi-user.target
```

## my\_webapp.service

puis exécution des instructions suivantes

**sudo systemctl daemon-reload**

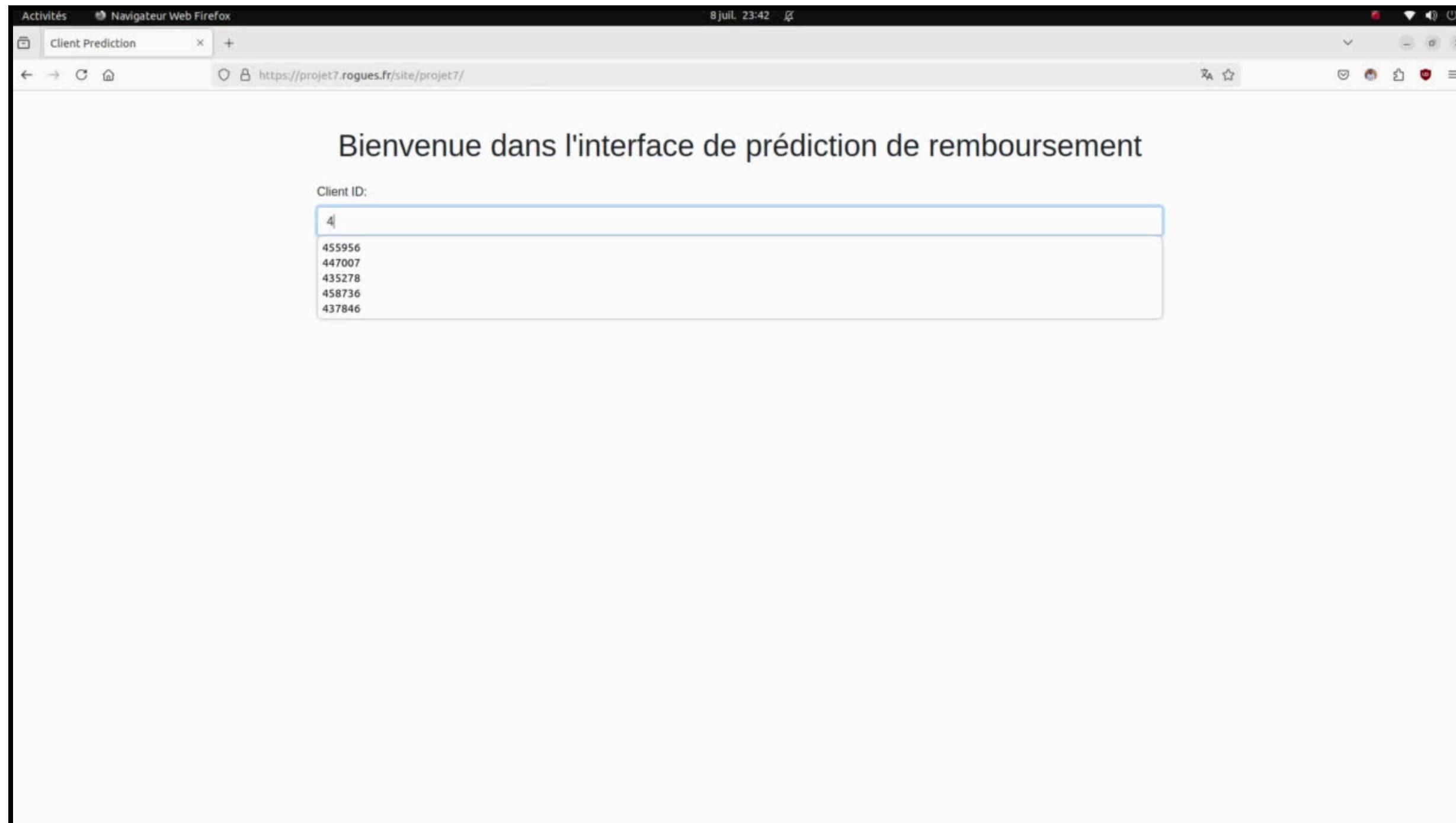
**sudo systemctl restart my\_webapp.service**

**sudo systemctl status my\_webapp.service**

```
arogues@rogues:/var/www/my_webapp_2/www$ sudo systemctl status my_webapp.service
[sudo] password for arogues:
● my_webapp.service - Gunicorn instance to serve my_webapp
   Loaded: loaded (/etc/systemd/system/my_webapp.service; enabled; vendor preset: enabled)
   Active: active (running) since Mon 2024-07-08 12:10:29 UTC; 9h ago
     Main PID: 828653 (gunicorn)
        Tasks: 7 (limit: 2301)
       Memory: 514.2M
          CPU: 16.715s
        CGroup: /system.slice/my_webapp.service
                └─828653 /usr/bin/python3 /home/arogues/.local/bin/gunicorn --workers 3 --bind 0.0.0.0:8000 appV2:app

Jul 08 12:10:29 rogues.fr systemd[1]: Started Gunicorn instance to serve my_webapp.
Jul 08 12:10:29 rogues.fr gunicorn[828653]: [2024-07-08 12:10:29 +0000] [828653] [INFO] Starting gunicorn 22.0.0
Jul 08 12:10:29 rogues.fr gunicorn[828653]: [2024-07-08 12:10:29 +0000] [828653] [INFO] Listening at: http://0.0
Jul 08 12:10:29 rogues.fr gunicorn[828653]: [2024-07-08 12:10:29 +0000] [828653] [INFO] Using worker: sync
Jul 08 12:10:29 rogues.fr gunicorn[828655]: [2024-07-08 12:10:29 +0000] [828655] [INFO] Booting worker with pid:
Jul 08 12:10:29 rogues.fr gunicorn[828656]: [2024-07-08 12:10:29 +0000] [828656] [INFO] Booting worker with pid:
Jul 08 12:10:29 rogues.fr gunicorn[828657]: [2024-07-08 12:10:29 +0000] [828657] [INFO] Booting worker with pid:
```

# API de prédition



# github actions

## Unit Test

Le script de test vérifie deux aspects principaux de l'application Flask :

1. La page d'accueil est accessible et affiche le message prévu.
2. La route de prédiction fonctionne correctement, renvoie une réponse HTTP 200, et inclut les prédictions et les importances des caractéristiques dans la réponse.

Ces tests permettent de s'assurer que les fonctionnalités clés de l'application sont en bon état de marche après des modifications ou des mises à jour.

## Script de déploiement

1. Se déclenche à chaque push sur la branche master.
2. Vérifie le code du dépôt.
3. Configure une connexion SSH sécurisée.
4. Se connecte au serveur VPS et se rend dans le répertoire de déploiement.
5. Crée un nouveau dossier avec un nom basé sur la date et l'heure actuelles.
6. Clone la branche master du dépôt GitHub dans ce nouveau dossier.
7. Affiche un message indiquant que la nouvelle version du projet a été copiée sur le serveur et fournit des instructions pour finaliser manuellement le déploiement en rechargant et redémarrant le service.

Cela permet de s'assurer que chaque nouvelle version de l'application est automatiquement copiée sur le serveur, tout en laissant le contrôle de la phase finale du déploiement.



# github actions

Activités Navigateur Web Firefox

Workflow runs - Stabadev

https://github.com/Stabadev/Projet7/actions

8 juil. 23:29

Importer les marque... Débuter avec Firefox projet7 Easy Install TP-Link A... Administratif Soutien Scolaire Recettes Code formations Saaaaalsa OpenClassRooms Gouvernance et jeux ... formation

Stabadev / Projet7

Type to search

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

All workflows Showing runs from all workflows

Filter workflow runs

7 workflow runs

	Event	Status	Branch	Actor
● modification script deploy.yml, spécification clonage branche master	now	Queued	master	
● modification script deploy.yml, spécification clonage branche master	now	Queued	master	
✓ Mise à jour du script deploy.yml	6 minutes ago	58s	master	
✓ Mise à jour du script deploy.yml	6 minutes ago	20s	master	
✓ rajout fichier csv	9 hours ago	58s	master	
✗ rajout fichier csv	9 hours ago	24s	master	
✗ Ajout github actions workflow, et MAJ requirements et client ID et csv	10 hours ago	56s	master	
✗ Ajout github actions workflow, et MAJ requirements et client ID et csv	10 hours ago	16s	master	
✗ ajout workflow déploiement	4 days ago	17s	master	

New workflow

All workflows Deploy to VPS Python application Management Caches Attestations Runners

Python application #4: Commit 011d4e2 pushed by Stabadev

Deploy to VPS #5: Commit 011d4e2 pushed by Stabadev

Python application #3: Commit 411dd0c7 pushed by Stabadev

Deploy to VPS #4: Commit 411dd0c7 pushed by Stabadev

Python application #2: Commit 83e05e1 pushed by Stabadev

Deploy to VPS #3: Commit 83e05e1 pushed by Stabadev

Python application #1: Commit c3ae37b pushed by Stabadev

Deploy to VPS #2: Commit c3ae37b pushed by Stabadev

https://github.com/Stabadev/Projet7/actions/runs/9847206017 y to VPS #1: Commit 636e6e8 pushed by Stabadev

Ouvrir le lien dans un nouvel onglet  
Ouvrir le lien dans une nouvelle fenêtre  
Ouvrir le lien dans une fenêtre de navigation privée  
Marquer le lien...  
Enregistrer la cible du lien sous...  
Enregistrer le lien dans Pocket  
Copier le lien  
Copier le lien sans le pistage du site  
Envoyer le lien à un appareil  
Recherche DuckDuckGo pour « modification sc... »  
Inspecter les propriétés d'accessibilité  
Inspecter  
Bloquer l'élément



# Data Drift

## Qu'est-ce que le Data Drift ?

Changement des distributions des données entre l'entraînement et la production.

Peut dégrader les performances du modèle.  
Nécessite une surveillance régulière.

## Pourquoi surveiller le Data Drift ?

Assurer la performance continue du modèle.  
Identifier les dérives dans les sources de données.  
Prendre des actions correctives rapidement.

## Utilisation d'Evidently pour détecter le Data Drift :

### Charger les Données :

application\_train pour l'entraînement.  
application\_test pour la production.

### Prétraitement des Données :

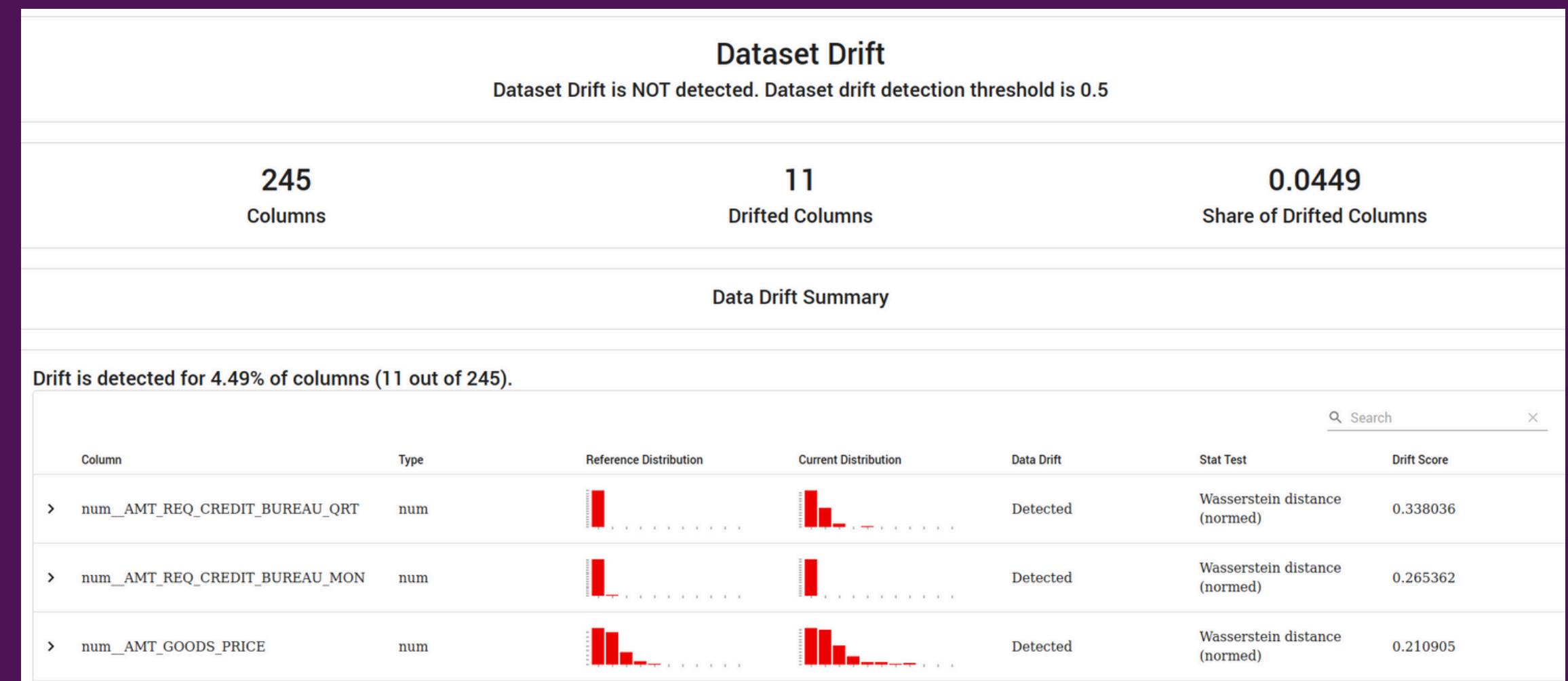
Imputation des valeurs manquantes.  
Standardisation des données.  
Encodage des variables catégorielles.

### Configurer Evidently :

Utiliser DataDriftPreset pour les métriques.  
Comparer les features entre les datasets.

### Générer le Rapport :

Créer un rapport HTML  
Visualiser les dérives entre les datasets.



### Interprétation des Résultats :

#### Pas de Dataset Drift Global :

Le rapport indique que le dataset global ne présente pas de dérive significative. Le seuil de détection utilisé est de 0.5, ce qui signifie que la dérive globale des données est en dessous de ce seuil.

#### Nombre de Colonnes Analyées :

245 colonnes ont été analysées pour détecter la dérive des données. Cela inclut toutes les features du dataset d'entraînement et de test.

#### Colonnes avec Dérive :

11 colonnes montrent une dérive significative. Ces colonnes ont des distributions de données qui ont changé entre l'entraînement et la production.

#### Proportion de Colonnes avec Dérive :

4.49% des colonnes montrent une dérive.

fin