UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Prévision de prêts, solution et déploiement
Sorbonne Data Analytics

Ibrahim Guoual Belhamidi

Problematic

Nous faisons partie d'une équipe dans le secteur de la banque de détail, confrontée à un taux de défaut plus élevé que prévu sur les prêts personnels. Ces prêts jouent un rôle essentiel dans les revenus de la banque, mais présentent un risque important de non-remboursement de la part des emprunteurs.

Objectif

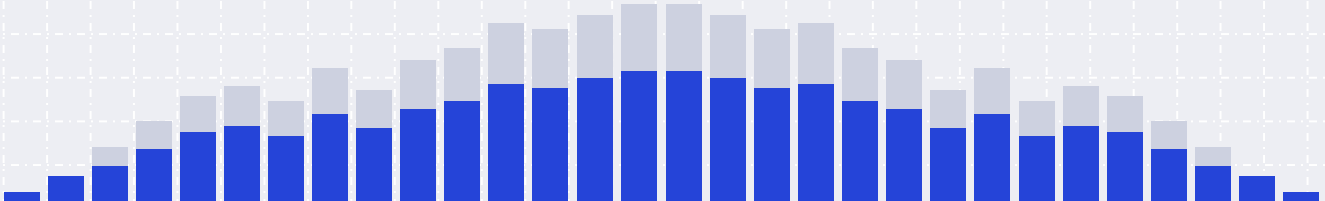
L'objectif est de développer un modèle prédictif permettant d'estimer la probabilité de défaut de chaque client, basé sur ses caractéristiques. Cela aidera la banque à anticiper les pertes potentielles et à allouer le capital nécessaire pour maintenir sa stabilité financière.

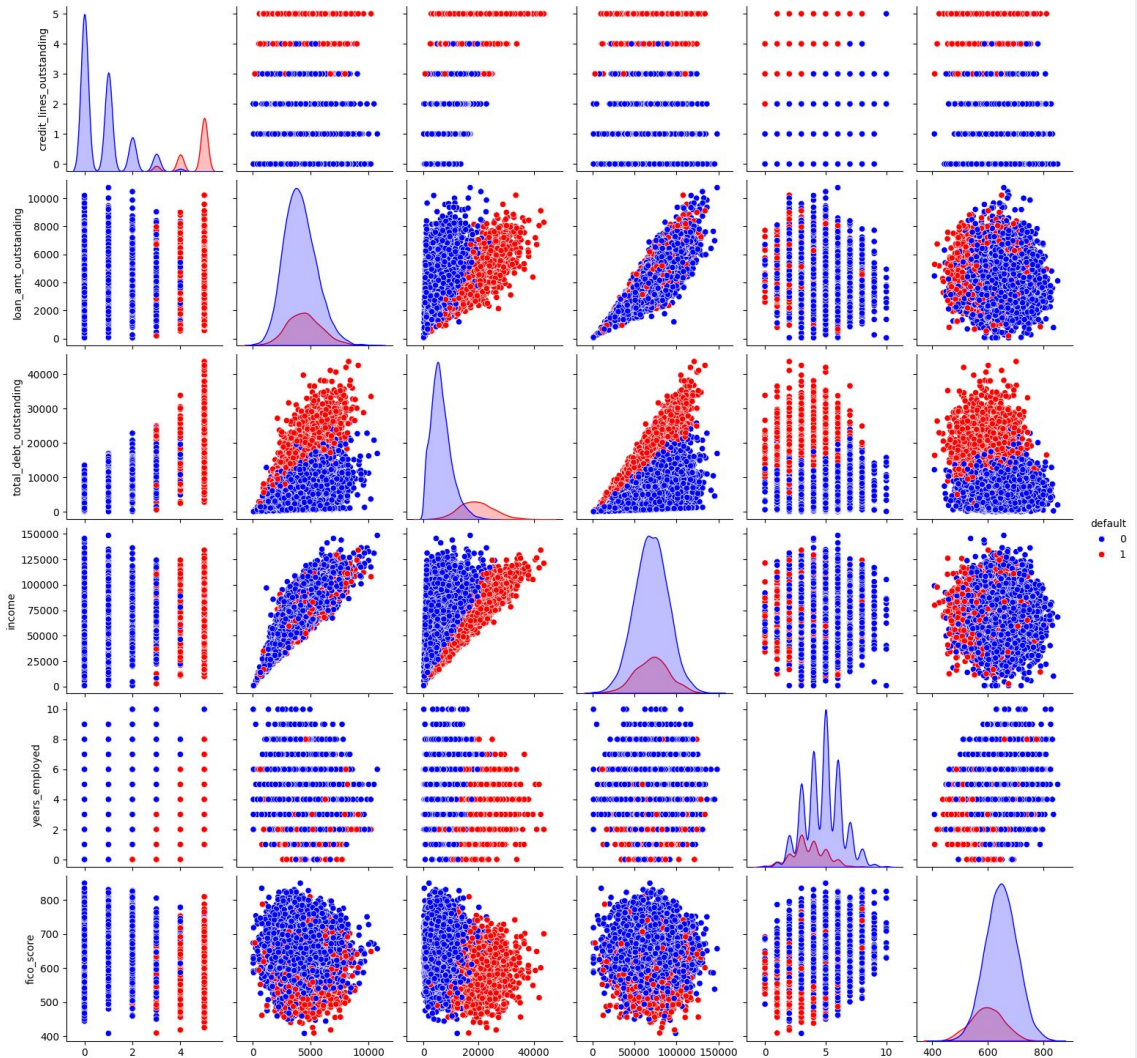
1851(18.51%)

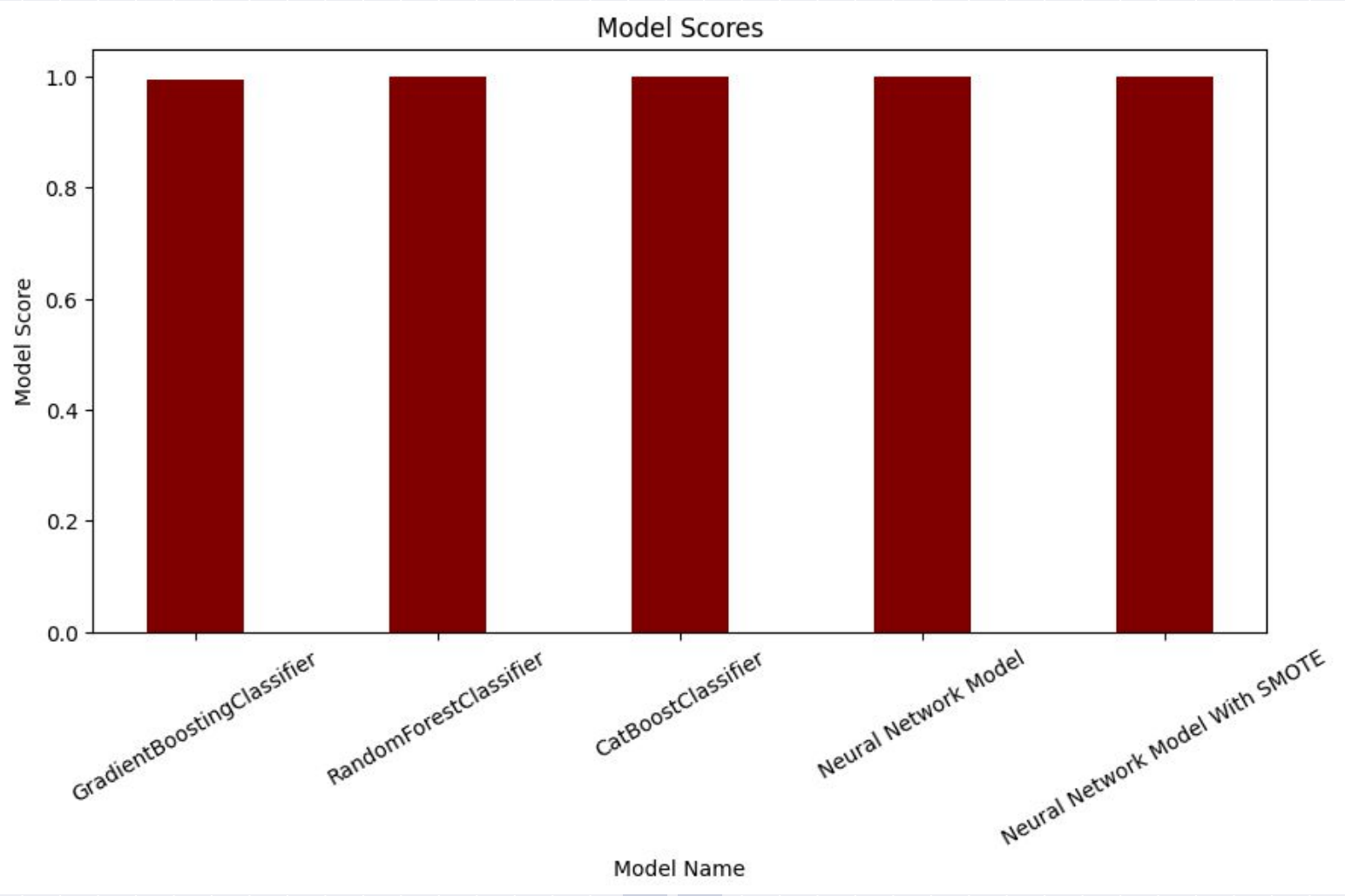
clients qui ont effectivement manqué leurs paiements

8149(81.49%)

clients qui n'ont pas manqué leurs paiements

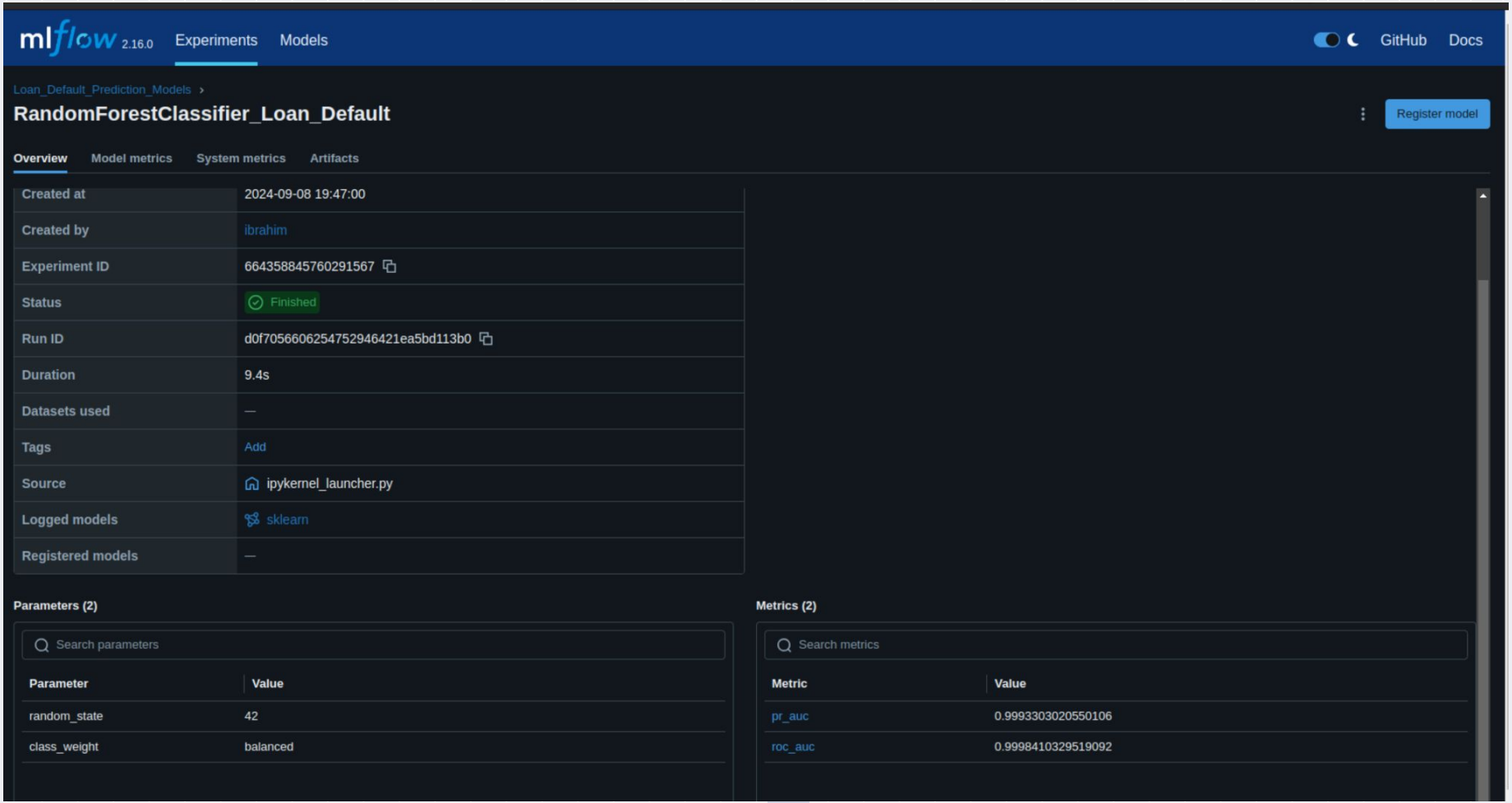



credit_lines_outstanding
loan_amt_outstanding
total_debt_outstanding
income
years_employed
fico_score
default
0
1


Model Scores
Model Score
1.0
0.8
0.6
0.4
0.2
0.0
GradientBoostingClassifier
RandomForestClassifier
CatBoostClassifier
Neural Network Model
Neural Network Model With SMOTE
Model Name

Columns Group by

Run Name	Created	Dataset	Duration	Source	Models
NN_with_SMOTE_and_R...	34 minutes ago	-	3.7min	ipykern...	keras
Neural_Network_Loan_P...	36 minutes ago	-	2.3min	ipykern...	keras
CatBoostClassifier_Loan...	37 minutes ago	-	5.6s	ipykern...	sklearn
RandomForestClassifier_...	37 minutes ago	-	7.2s	ipykern...	sklearn
NN_with_SMOTE_and_R...	53 minutes ago	-	4.3min	ipykern...	keras
Neural_Network_Loan_P...	58 minutes ago	-	2.6min	ipykern...	keras
Neural_Network_Loan_P...	1 hour ago	-	2.4min	ipykern...	-
CatBoostClassifier_Loan...	1 hour ago	-	7.0s	ipykern...	sklearn
RandomForestClassifier_...	1 hour ago	-	9.4s	ipykern...	sklearn


mlflow 2.16.0
Experiments
Models
GitHub
Docs
Loan_Default_Prediction_Models ›
RandomForestClassifier_Loan_Default
Register model
Overview
Model metrics
System metrics
Artifacts
Created at
2024-09-08 19:47:00
Created by
ibrahim
Experiment ID
664358845760291567
Status
Finished
Run ID
d0f7056606254752946421ea5bd113b0
Duration
9.4s
Datasets used
—
Tags
Add
Source
ipykernel_launcher.py
Logged models
sklearn
Registered models
—
Parameters (2)
Search parameters
Parameter
Value
random_state
42
class_weight
balanced
Metrics (2)
Search metrics
Metric
Value
pr_auc
0.9993303020550106
roc_auc
0.9998410329519092

Loan_Default_Prediction_Models ›

CatBoostClassifier_Loan_Default

Register model

Overview | Model metrics | System metrics | Artifacts

Description

No description

Details

Created at	2024-09-08 20:18:18
Created by	ibrahim
Experiment ID	664358845760291567
Status	Finished
Run ID	a3c3a58a29a64589b4c5ea60775e737f
Duration	7.0s
Datasets used	—
Tags	Add
Source	ipykernel_launcher.py
Logged models	sklearn
Registered models	—

Parameters (2)

Search parameters

Parameter	Value
random_state	42
class_weights	[1, 97]

Metrics (2)

Search metrics

Metric	Value
pr_auc	0.9993487073843894
roc_auc	0.9998461140904327

Neural_Network_Loan_Prediction

Created at	2024-09-08 20:21:12
Created by	ibrahim
Experiment ID	664358845760291567
Status	Failed
Run ID	37dc5e9395d945e1bd31871fa290dabe
Duration	2.4min
Datasets used	—
Tags	Add
Source	ipykernel_launcher.py
Logged models	—
Registered models	—

Parameters (8)

Search parameters

Parameter	Value
optimizer	adam
loss_function	binary_crossentropy
epochs	50
layer1_units	64
input_dim	6

Metrics (4)

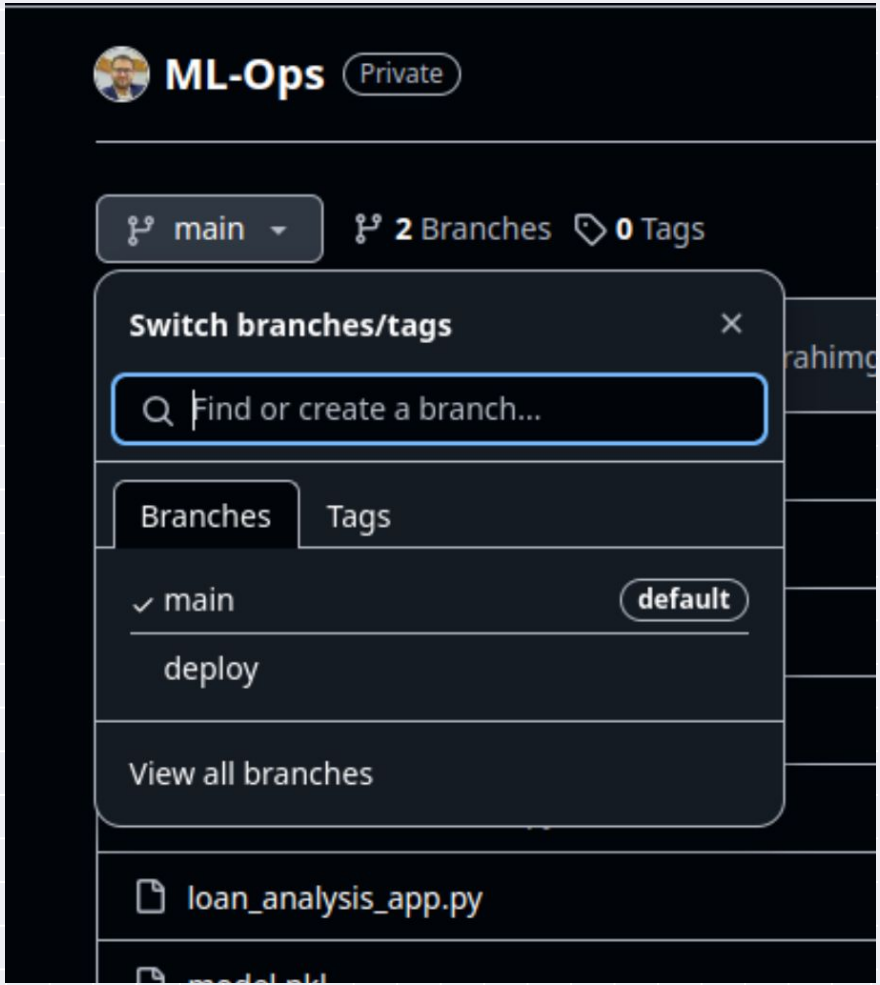
Search metrics

Metric	Value
test_accuracy	0.9959999918937683
test_loss	0.012040859088301659
roc_auc	0.9999234644179121
precision_recall_auc	0.9996432182264513

CI-CD:

Dans notre cas, nous avons créé deux branches : l'une est notre branche de développement où une image Docker sera créée, et l'autre, « deploy », poussera cette image vers le registre de conteneurs Azure. C'est considéré comme un bon modèle de conception pour le déploiement. En pratique, seul le responsable de l'équipe ou le chef de projet aura l'autorité pour déployer le projet (après fusion dans la branche principale).

Deploy URL:
https://loan-analysis-app.bluewave-fad57f1e.francecentral.azurecontainerapps.io

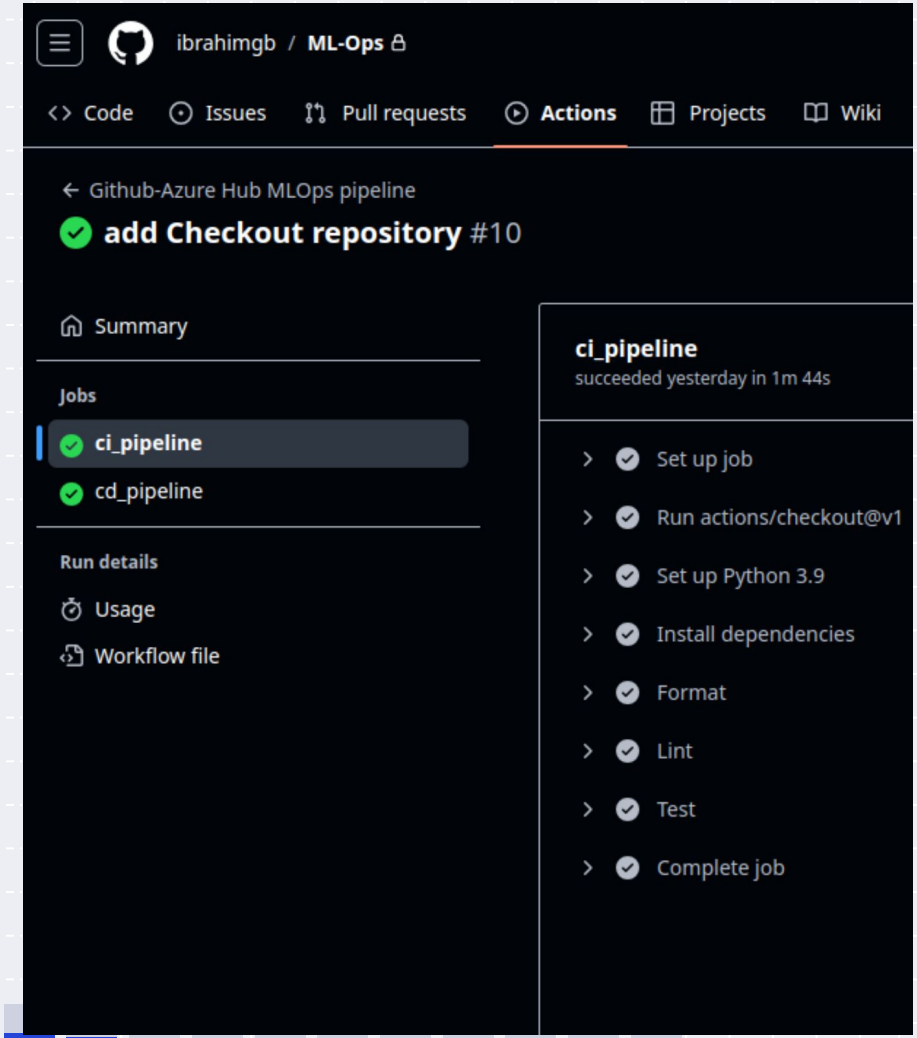


CI:

Ce pipeline est un workflow GitHub Actions utilisé pour le MLOps, spécifiquement pour un dépôt de projet. Il se compose de deux tâches principales :

Pipeline CI (Intégration Continue) : Cette tâche veille à ce que le code soit extrait, que les dépendances soient installées, et elle exécute des étapes de formatage, de linting et de tests pour vérifier la qualité et la fonctionnalité du code. La configuration implique :

- Configuration de Python 3.9
- Installation des dépendances nécessaires
- Exécution des vérifications de formatage du code
- Linting pour faire respecter les normes de code
- Exécution des tests pour s'assurer que tout fonctionne correctement



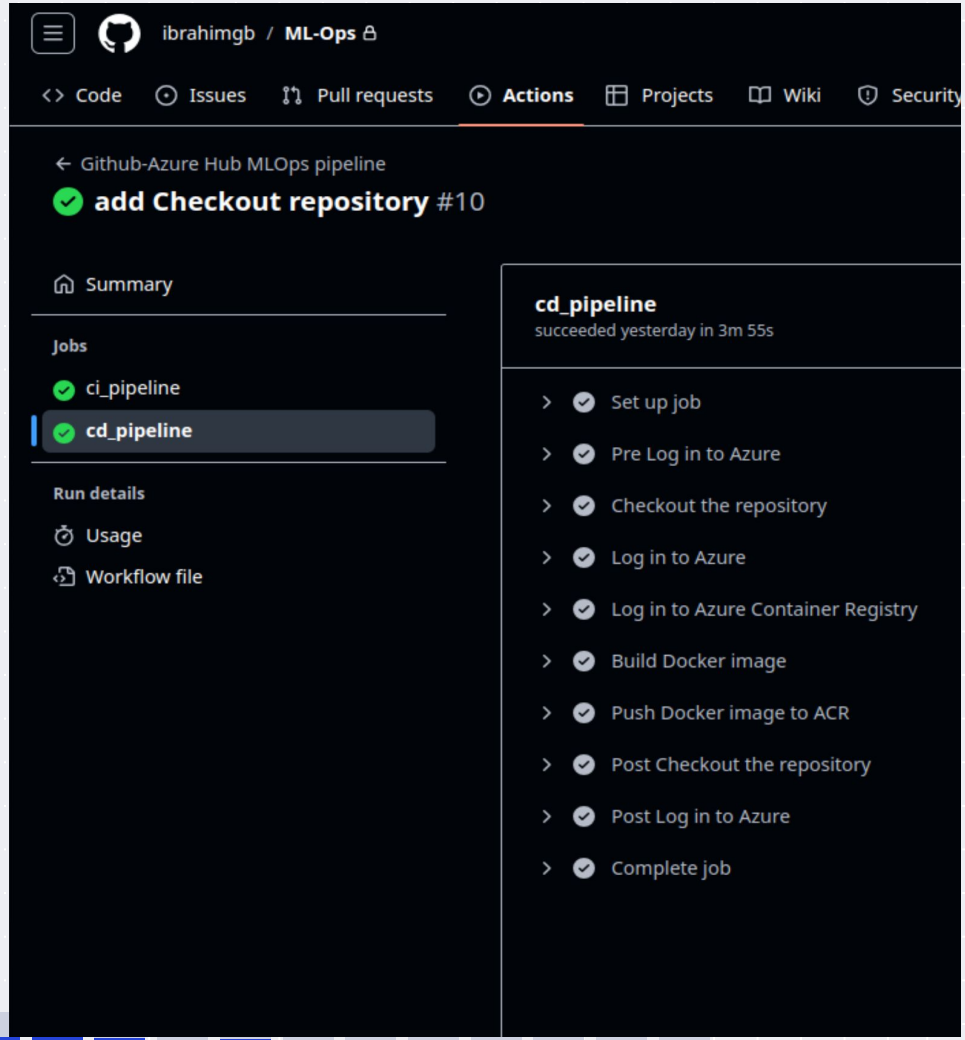
CD:

Ce pipeline est le Pipeline CD (Déploiement Continu) du workflow GitHub Actions. Il gère le processus de déploiement automatisé pour le projet MLOps, avec les étapes suivantes :

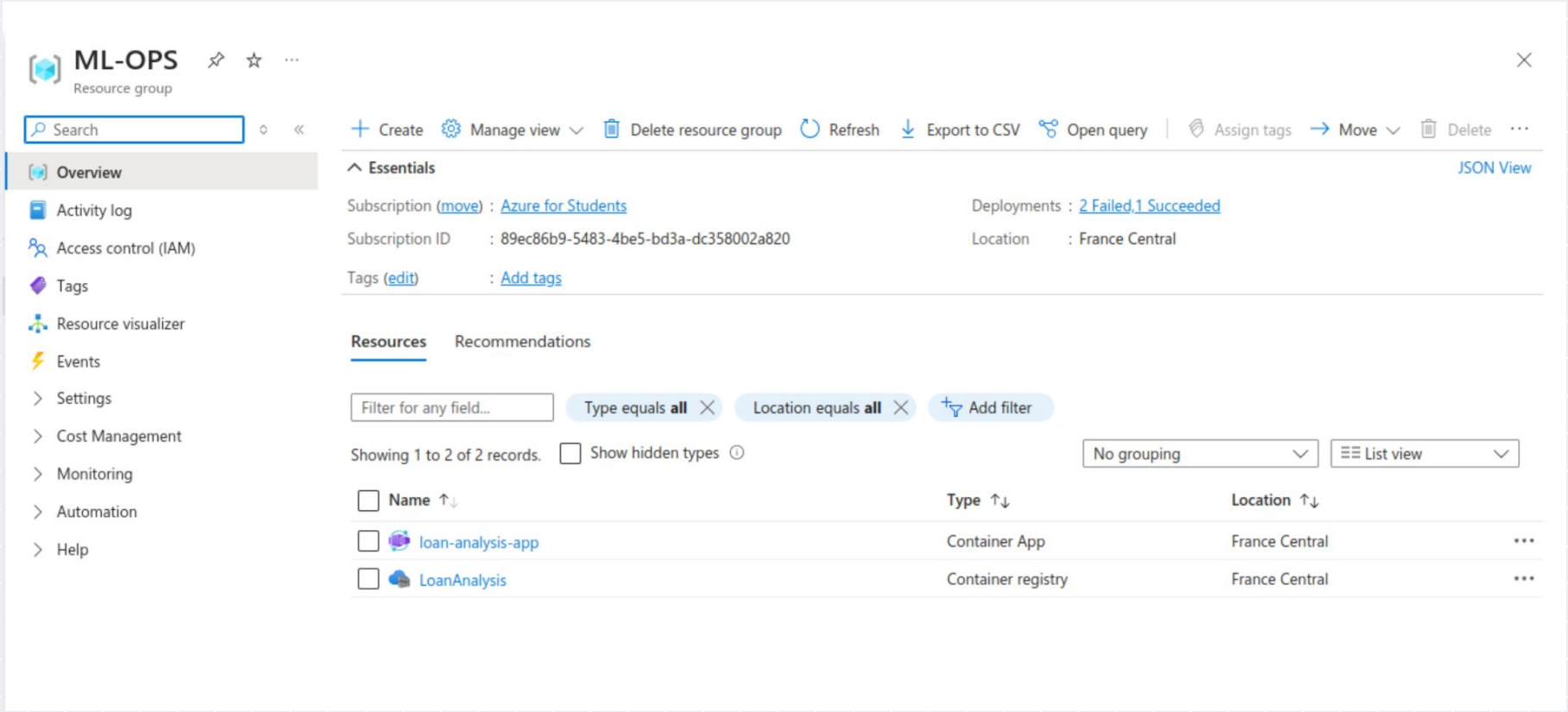
Connexion et Configuration :
- Pré-connexion à Azure
- Extraction du dépôt
- Connexion à Azure et au Registre de Conteneurs Azure (ACR)

Containerisation :
- Création d'une image Docker de l'application
- Pousser l'image Docker vers l'ACR

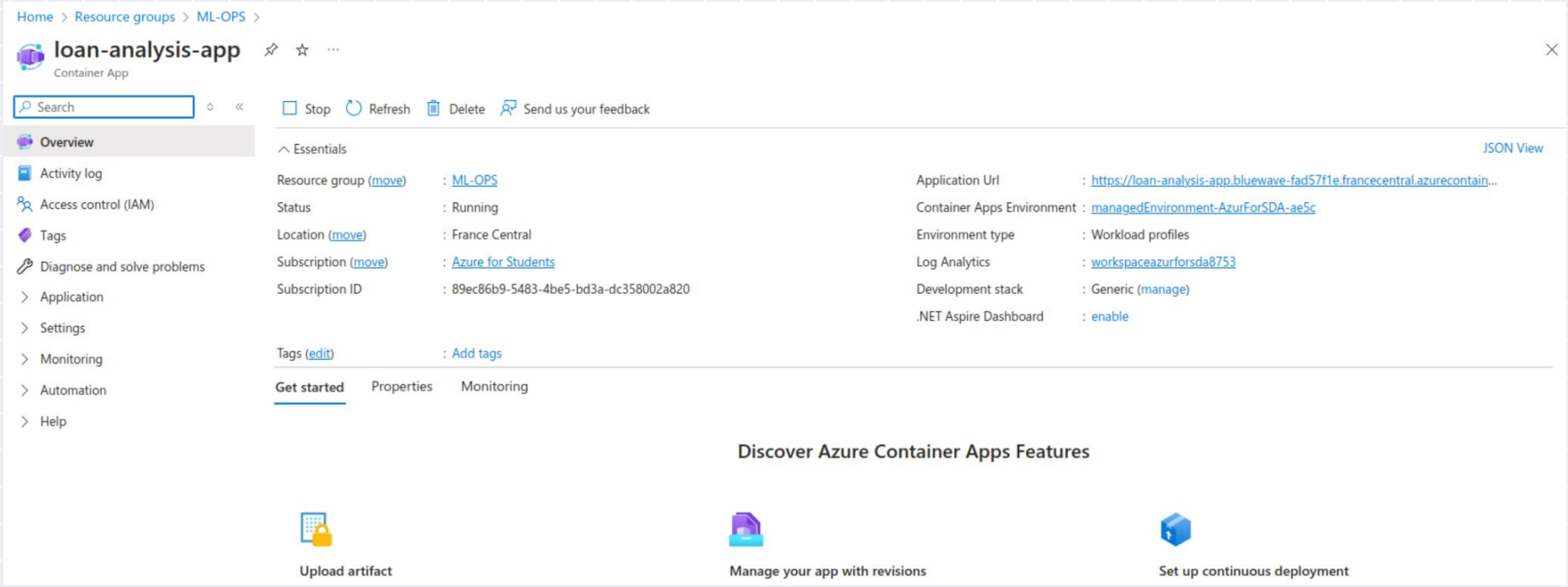
Post-Déploiement :
- Post-extraction du dépôt et connexion finale à Azure
- Finalisation de la tâche



Azure resource group:

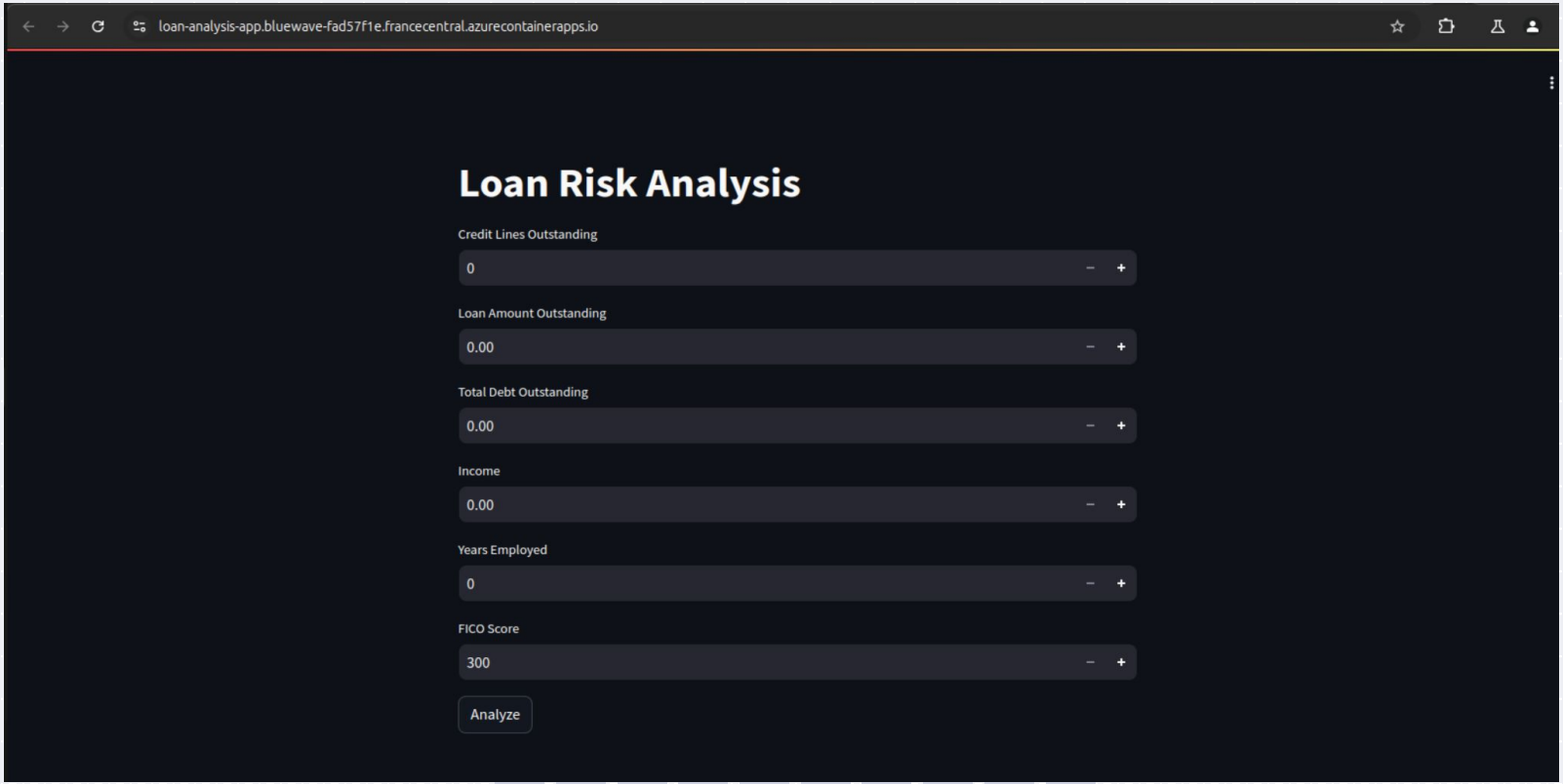


Azure resource group:



Solution déployée :

URL: https://loan-analysis-app.bluewave-fad57f1e.francecentral.azurecontainerapps.io



Merci !