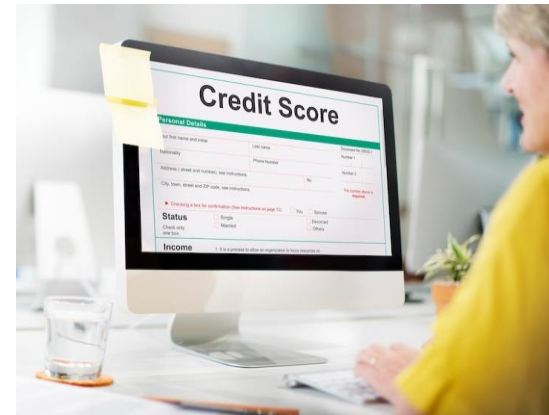




Implémenter un outil de « scoring crédit »



Projet : DataScientist #7

Oumou Faye

Mentor : Medina Hadjem

PLAN DE PRESENTATION



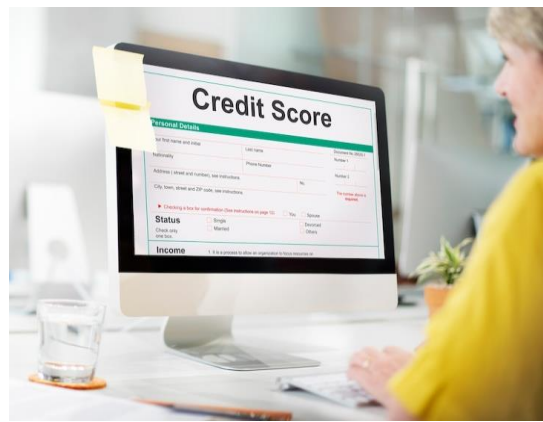
INTRODUCTION
-
MISSION



MODÉLISATION



DÉPLOIEMENT
-
CONCLUSION



INTRODUCTION - MISSION

INTRODUCTION - MISSION

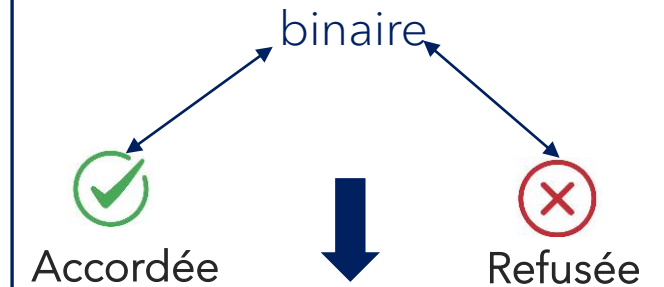


Objectif

Évaluer la probabilité de remboursement d'un client



Classifier la demande en deux catégories :



Automatiser la décision de d'octroi de crédit

Entreprise



Activité



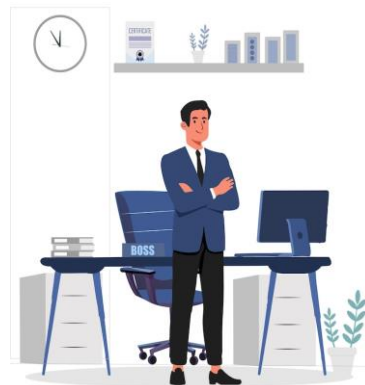
Proposition de crédit à la consommation
(client à historique limité)

Fonction

Data Scientist



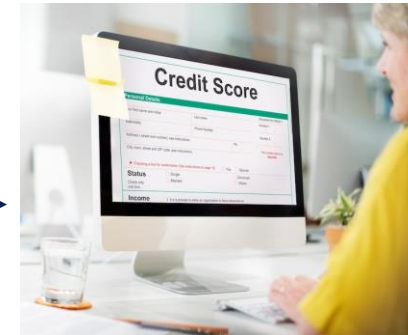
Manager



Mickaël

Mission

Mise en place d'un outil de « *scoring crédit* »



Prédire la probabilité de remboursement du client



Cibles : les nouveaux clients
et les clients avec un
historique bancaire limité



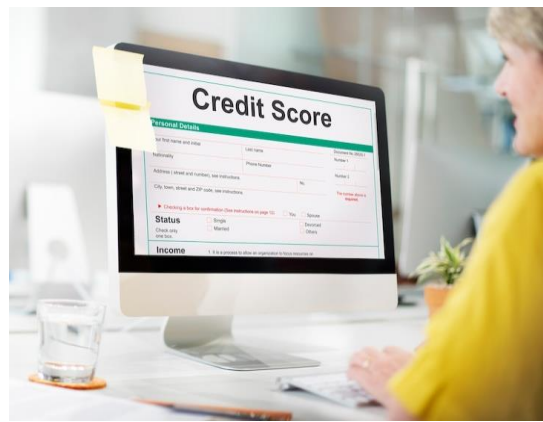
- Données issues du concours Kaggle Home Credit (#10 fichiers au format « .csv »)
- Données d'applications de crédit et de leur suivi comportemental.

1. HomeCredit_columns_description

Dictionnaire de données décrivant toutes les variables contenues dans les fichiers ci-dessous :

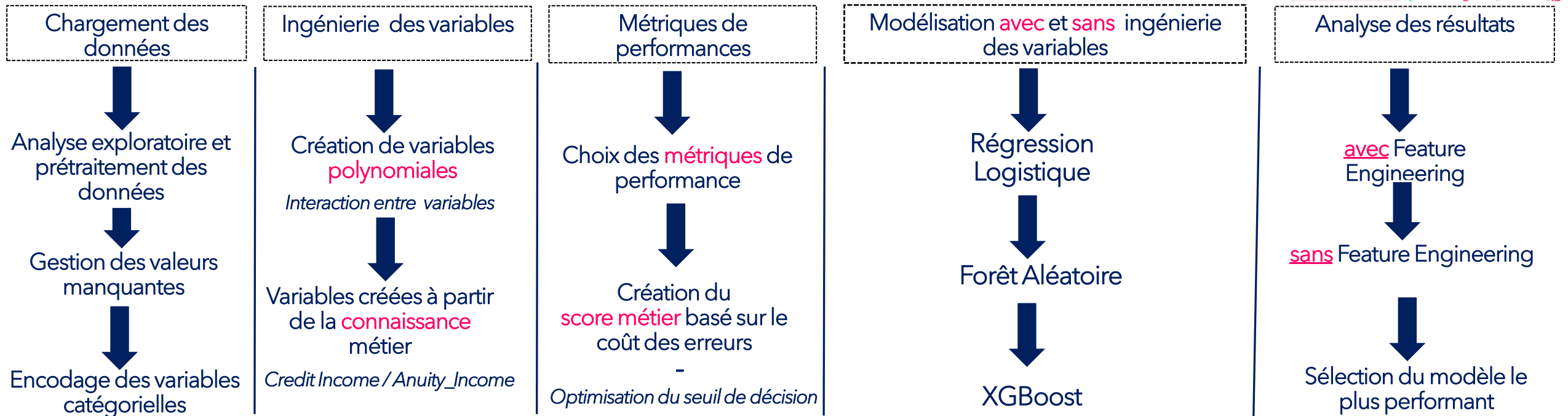
2. application_train 307.511 <u>lignes</u> / 122 <u>colonnes</u>	4. bureau	6. previous_application	8. installments_payments
3. application_test 48.744 <u>lignes</u> / 121 <u>colonnes</u>	5. bureau_balance	7. POS_CASH_balance	9. credit_card_balance
			10. sample_submission
Information de chaque demande de crédit. application_train intègre la variable cible ↓ ▪ Target : 1 = difficulté de paiement ▪ Target: 0 = Autre cas	↓ - Crédit contracté auprès d'autres institutions financières - Historique mensuel des crédits signalés <i>Données de comportements sur la durée de chaque crédit</i>	↓ - Historique des demandes de crédit. - Historiques mensuels des demandes de crédit enregistrées	↓ - Historique des paiements (effectués et manqués) - Nombre de carte de crédit détenu par les clients de la société - Identifiant unique du client

application_train = données utilisées pour l'entraînement lors de la modélisation



MODÉLISATION

PIPELINE DE MODÉLISATION



MODÉLISATION DU MODÈLE LE PLUS PERFORMANT : **XGBoost**



XGBoost - Modèle final retenu pour le scoring en production

METRIQUES DE PERFORMANCES



DETAIL DES MÉTRIQUES DE PERFORMANCES

- **AUC_global** : Score **AUC** global sur toutes les prédictions.
- **AUC_mean** : Moyenne des scores **AUC** Area Under the Curve sur les validations croisées.

- **Business_Score_Global** : Score métier normalisé sur toutes les prédictions.
- **Business_Score_Mean** : Moyenne du score métier normalisé (coût évité) sur les validations croisées (#5).

- **Accuracy_Global** : Précision globale du modèle (sur tout l'ensemble).
- **Accuracy_Mean** : Moyenne de la précision globale (toutes classes confondues).

- **Precision_Global** : Précision positive globale (proportion de bons positifs parmi les positifs prédits).
- **Precision_Mean** : Moyenne de la précision positive (sur les folds).

- **Recall_Global** : Rappel global : (proportion de vrais positifs détectés).
- **Recall_mean** : Moyenne du rappel (sur les folds).

- **F1_mean** : Moyenne du F1-score (équilibre précision/rappel).
- **F1_global** : F1-score global sur l'ensemble des prédictions.

DETAIL DU SCORE METIER

Étape 1 :

Transformation des probabilités du modèle en classes (0 ou 1) avec un **seuil de décision** (threshold = 0.5).

Étape 2 :

Identification des erreurs de prédiction :

- Faux Positifs (FP) = Clients **fiables refusés**
- Faux Négatifs (FN) = Clients **acceptés à tort**

SCORE METIER
> 0 :



Minimisation des
erreurs coûteuses

Étape 3 :

Calcul du **coût total des erreurs** :
 $\text{coût_total} = \text{FP} \times \text{coût_fp} + \text{FN} \times \text{coût_fn}$
(*coût_fn* = **10** & *coût_fp* = 1)

SCORE METIER
SUPERIEUR > 0 :



Modèle *économiquement intéressant*

Étape 4 :

Calcul du **pire coût possible** si le modèle se trompe à 100%.
 $\text{worst} = \text{np.sum}(y_true == 0) / \text{worst_fn} = \text{np.sum}(y_true == 1) / \text{worst_cost} = \text{worst_fp} * \text{cost_fp} + \text{worst_fn} * \text{cost_fn}$

Étape 5 :

Calcul du **score métier normalisé** entre 0 et 1 :
 $\text{score} = 1 - (\text{coût_total} / \text{coût_max})$

VISUALISATION

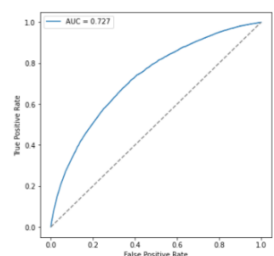
mlflow



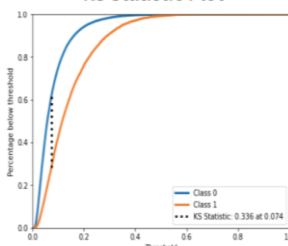
Model Registry :

« Credit Scoring Tool Baseline »

ROC CURVE



KS Statistic Plot



Experiment Name :

« Credit_Scoring_Tool_baseline »

=== RÉSULTATS VALIDATION CROISÉE ===

AUC moyen : 0.634 ± 0.003

AUC global : 0.634

Score métier global : 0.532

Score métier moyen : 0.532

Accuracy moyen : 0.919

Accuracy global : 0.919

Précision moyenne : 0.000

Précision globale : 0.000

Rappel moyen : 0.000

Rappel global : 0.000

F1-score moyen : 0.000

F1-score global : 0.000

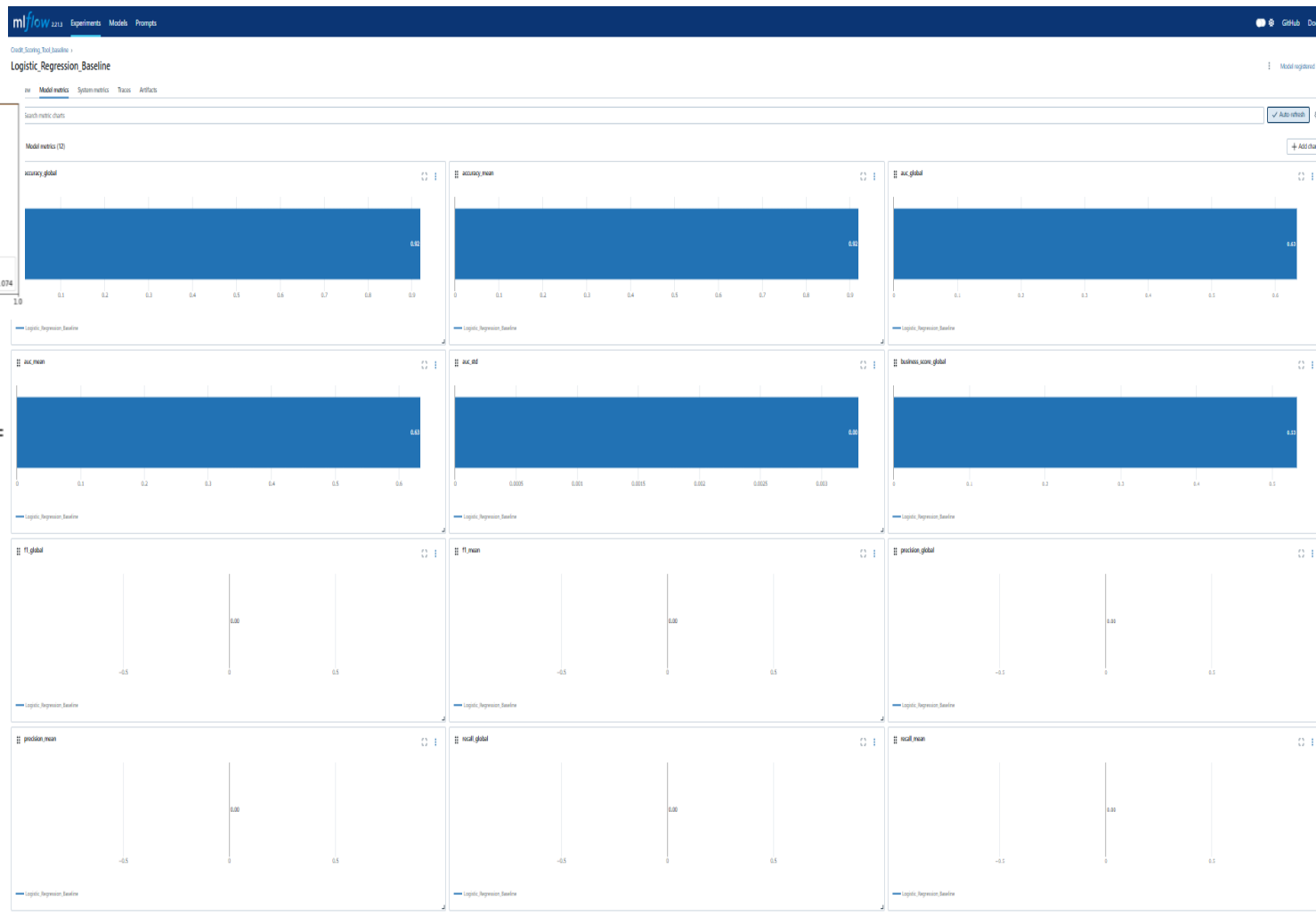
RESULTATS

Sans feature Engineering

Résultat de la modélisation avec l'algorithme de classification de la **Régression Linéaire**.

Modèle statistique:

- Objectif : Prédire une classe binaire (Crédit **Accordé**: =0 / Crédit **Refusé**: = 1)
- Visualisation : graphique présentant une droite qui minimise l'erreur entre les prédictions et les vraies valeurs
- Mise en place : calcul de la probabilité de remboursement et classification de la demande de crédit (accord / refus)



SCORE METIER

> 0 :



Modèle économiquement intéressant

- Modèle simple avec bonne précision globale 0.92%.
- Le modèle n'identifie pas les cas positifs (scores de précision, rappel et F1 nuls)
- La cause : le déséquilibre des classes.

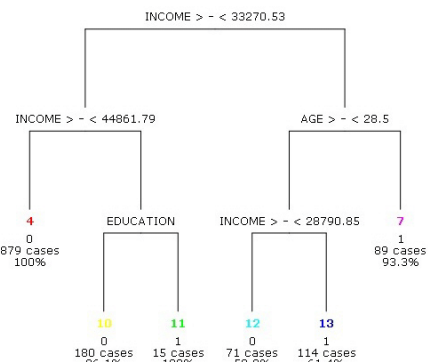


0 = (pas de difficulté de paiement) 282686 (91.93%) / 1=(difficulté de paiement) = nombre de 24825 (8.07%) 04/07/2025



Model Registry :

« Credit Scoring Tool Baseline »



Experiment Name :

« Credit_Scoring_Tool_baseline »

=== RÉSULTATS VALIDATION CROISÉE ===

AUC moyen : 0.712 ± 0.001

AUC global : 0.712

Score métier moyen : 0.533

Score métier global : 0.533

Accuracy moyen : 0.919

Accuracy global : 0.919

Précision moyenne : 0.657

Précision globale : 0.667

Rappel moyen : 0.001

Rappel global : 0.001

F1-score moyen : 0.002

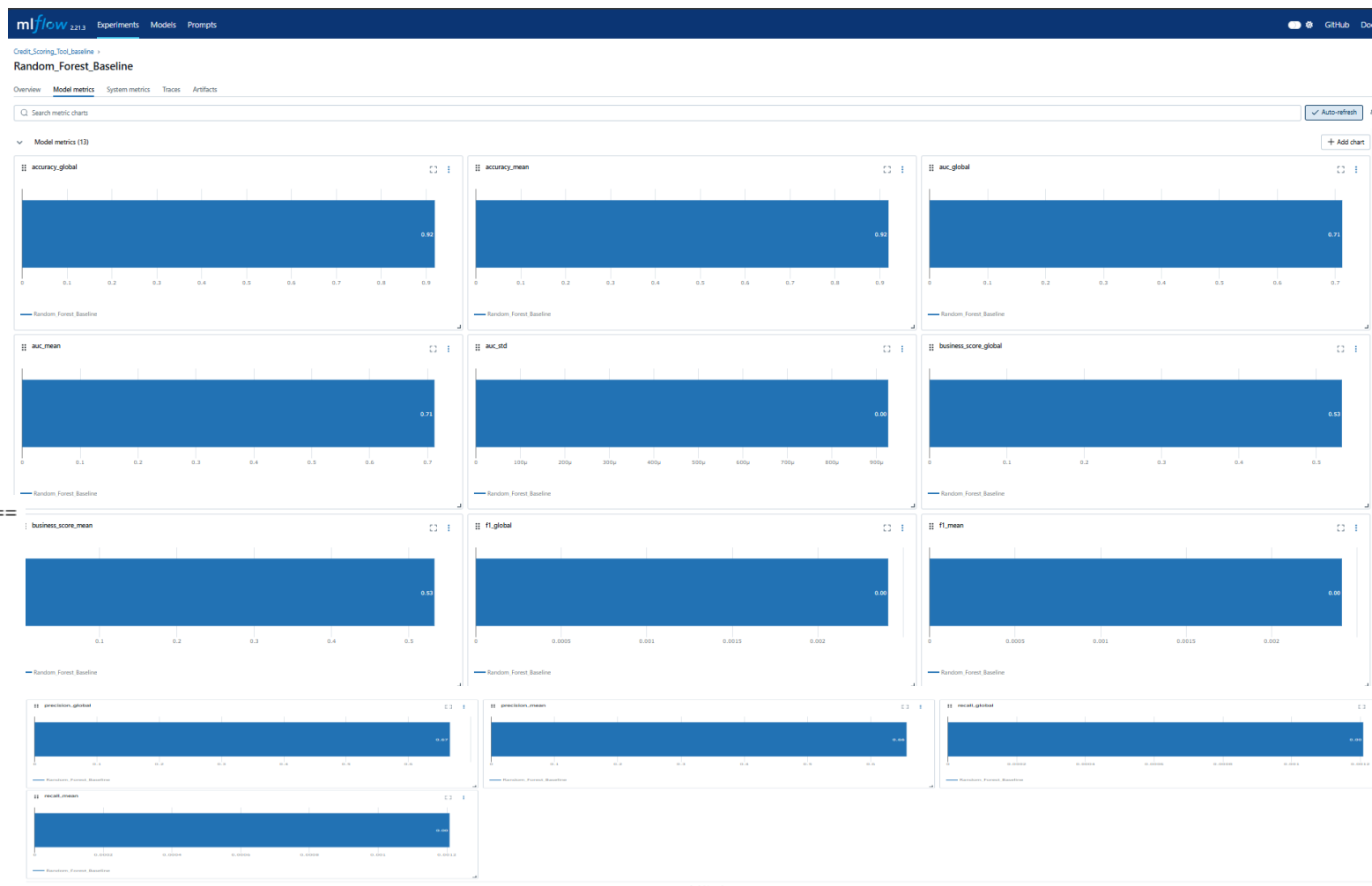
F1-score global : 0.002

RESULTATS

Sans feature Engineering

Résultat de la modélisation avec l'algorithme de classification de la **Random Forest**.

- **Objectif** : Classifier 0 = Crédit accordé / 1 = Crédit refusé
- **Principe** : Combine plusieurs arbres aléatoires pour améliorer la précision
- **Usage** : Robuste, réduit le surapprentissage, très stable.



- Score AUC en légère augmentation mais, le modèle est limité concernant la détection des clients à risque avec un F1-score quasi nul.
- Le déséquilibre des classes est aussi la raison pour laquelle le modèle ne parvient pas à détecter des clients à risques.



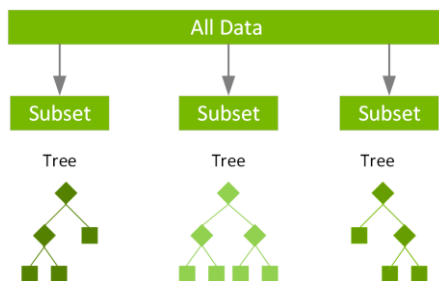
VISUALISATION

mlflow



Model Registry :

« Credit Scoring Tool Baseline »



Experiment Name :

« Credit_Scoring_Tool_baseline »

=== RÉSULTATS VALIDATION CROISÉE ===

AUC moyen : 0.751 ± 0.001

AUC global : 0.751

Score métier moyen : 0.536

Score métier global : 0.536

Accuracy moyen : 0.920

Accuracy global : 0.920

Précision moyenne : 0.631

Précision globale : 0.627

Rappel moyen : 0.008

Rappel global : 0.008

F1-score moyen : 0.016

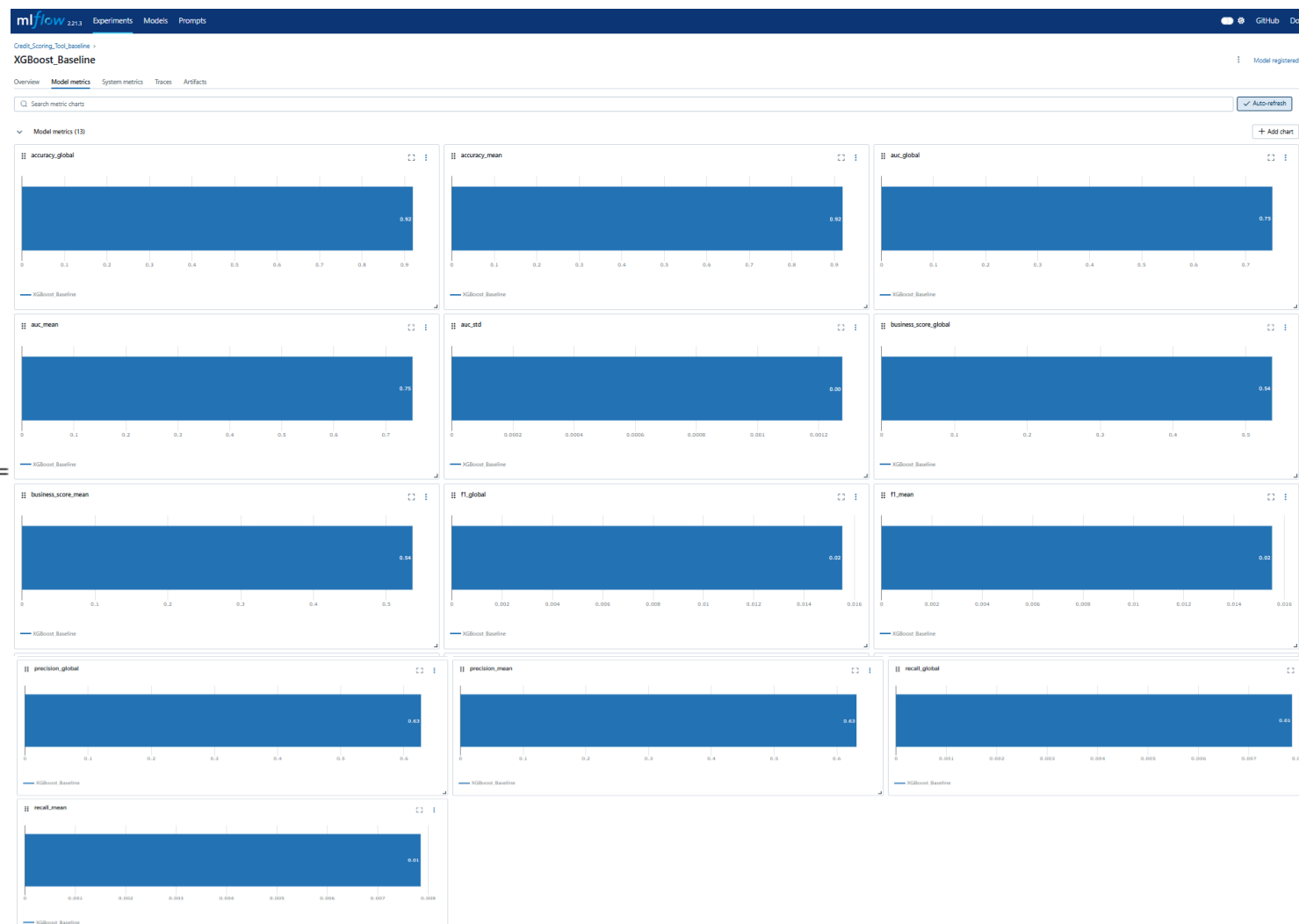
F1-score global : 0.016

RESULTATS

Sans feature Engineering

Résultat de la modélisation avec l'algorithme de classification de la **XGBoost**:

- **Objectif** : **Booster** les performances de prédiction
- **Principe** : Entraîne les arbres en série, chaque arbre corrige les erreurs du précédent
- **Avantage** : Très Performant, gère les données manquantes



- AUC également en légère augmentation 0.751.
- Le déséquilibre des classes est encore le résultat de la sous performance sur les classes minoritaires.

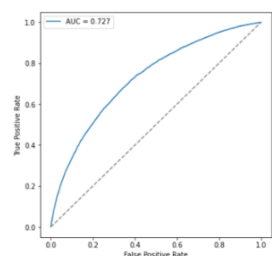
0 = (pas de difficulté de paiement) 282686 (91.93%) / 1=(difficulté de paiement) = nombre de 24825 (8.07%)



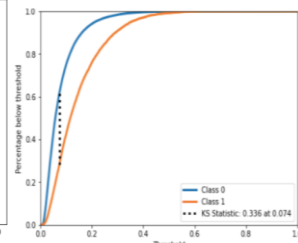
Model Registry :

« Credit Scoring Tool Baseline »

ROC CURVE



KS Statistic Plot



Experiment Name :

« Credit_Scoring_Tool_FeatureEng »

=== RÉSULTATS VALIDATION CROISÉE ===

AUC moyen : 0.711 ± 0.002

AUC global : 0.711

Score métier moyen : 0.656

Score métier global : 0.656

Accuracy global : 0.616

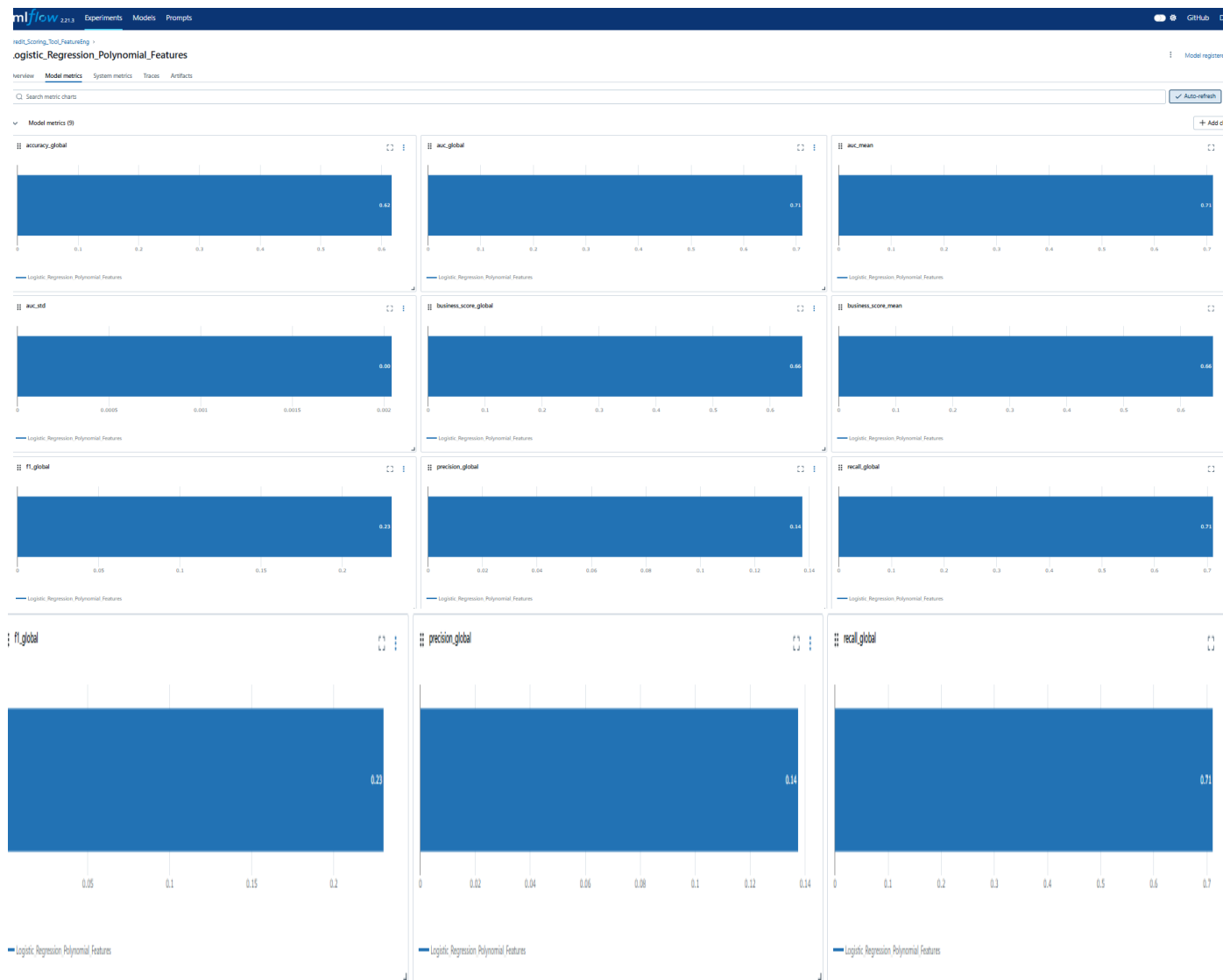
Precision global : 0.137

Recall global : 0.711

F1-score global : 0.230

avec feature Engineering – caractéristiques polynomiales

Résultat de la modélisation avec l'algorithme de classification de la Régression Linéaire.



SCORE METIER
> 0



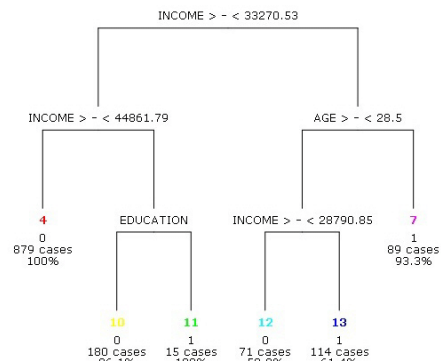
Minimisation des
erreurs coûteuses

- AUC assez élevé et le score métier supérieur à 0.50
- Le modèle reste simple mais il est efficace



Model Registry :

« Credit Scoring Tool Baseline »



Experiment Name :

« Credit_Scoring_Tool_FeatureEng »

=== RÉSULTATS VALIDATION CROISÉE ===

AUC moyen : 0.734 ± 0.002

AUC global : 0.734

Score métier moyen : 0.534

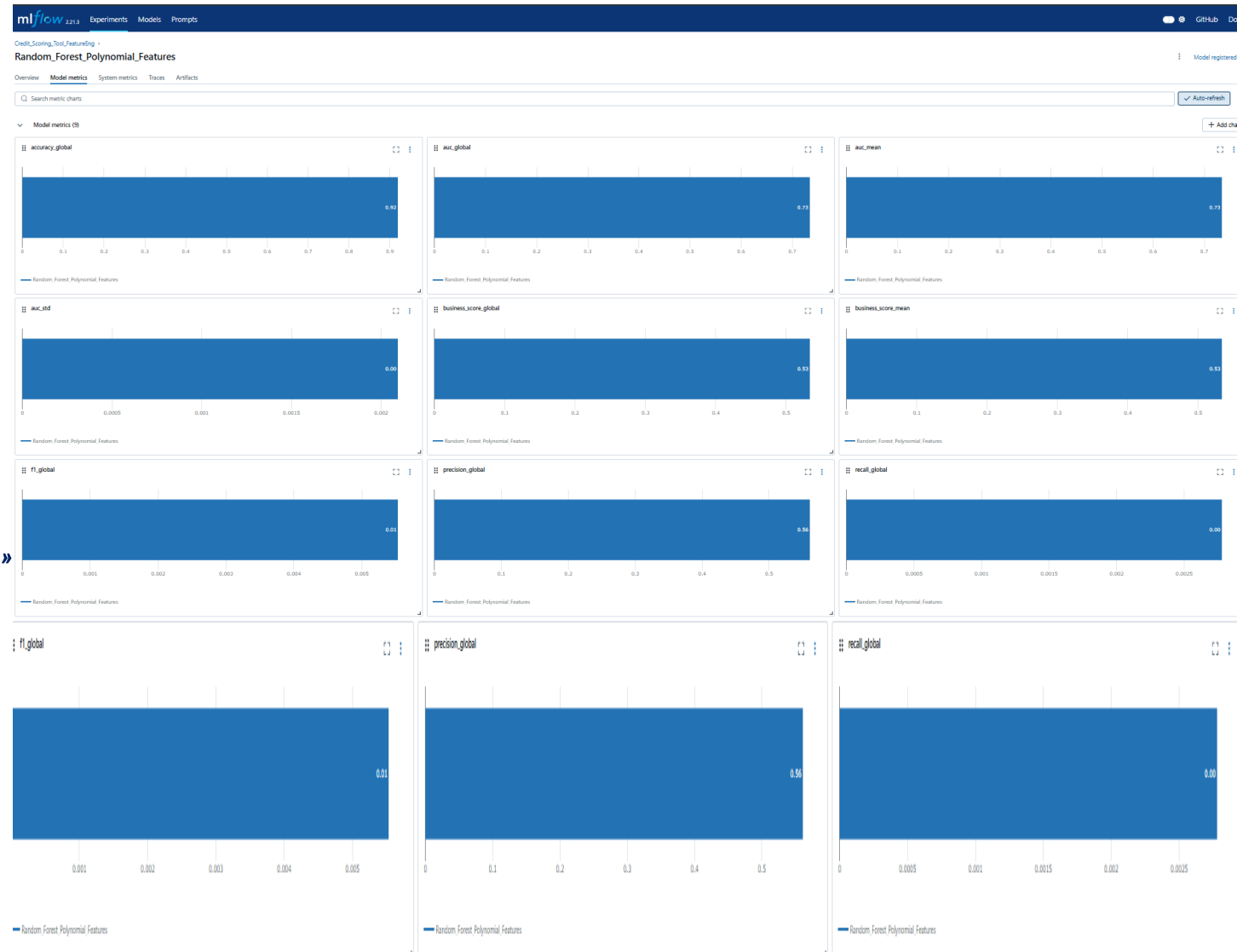
Score métier global : 0.534

Accuracy global : 0.919

Precision global : 0.561

Recall global : 0.003

F1-score global : 0.006

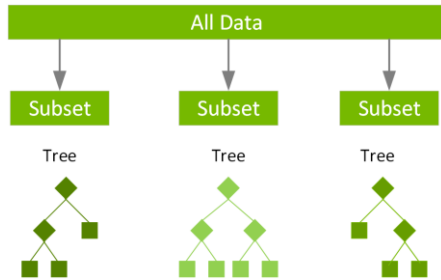


- Très bonne précision (0.561) mais faible rappel.
- Le modèle est conservateur, et effectue peu de détections.



Model Registry :

« Credit Scoring Tool Baseline »



Experiment Name :

« Credit_Scoring_Tool_FeatureEng »

=== RÉSULTATS VALIDATION CROISÉE ===

AUC moyen : 0.752 ± 0.001

AUC global : 0.752

Score métier moyen : 0.533

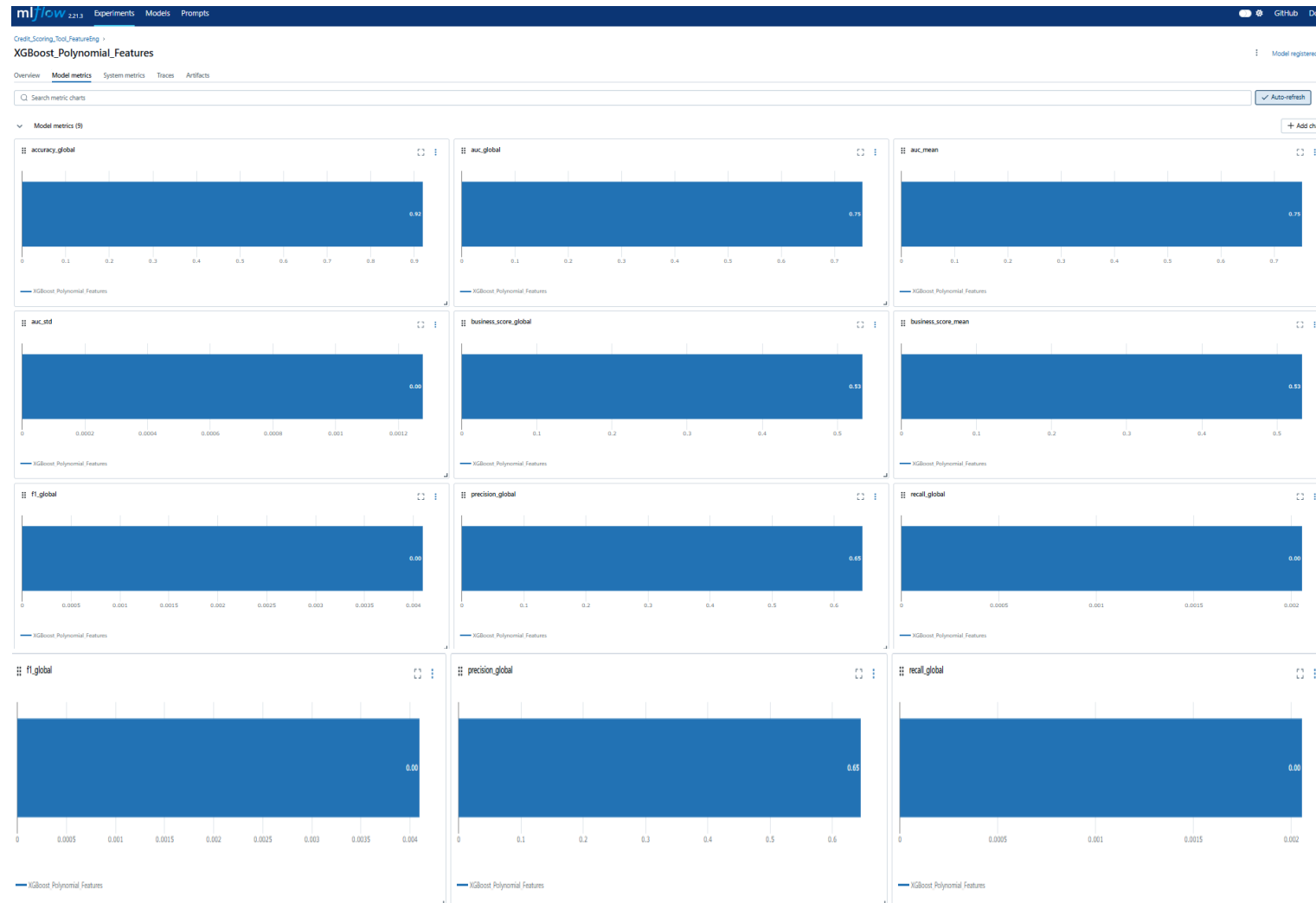
Score métier global : 0.533

Accuracy global : 0.919

Precision global : 0.646

Recall global : 0.002

F1-score global : 0.004



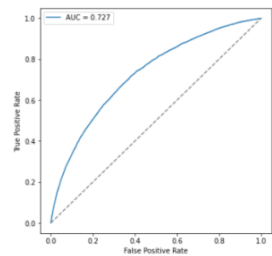
- Meilleure AUC (0.752), bon équilibre global.
- Le modèle est précis mais le score de rappel est encore faible.
- Nous n'avons toujours pas appliqué la gestion du déséquilibre des classe avec SMOTE à ce stade



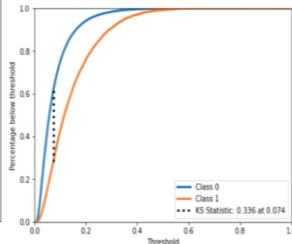
Model Registry :

« Credit Scoring Tool Baseline »

ROC CURVE



KS Statistic Plot



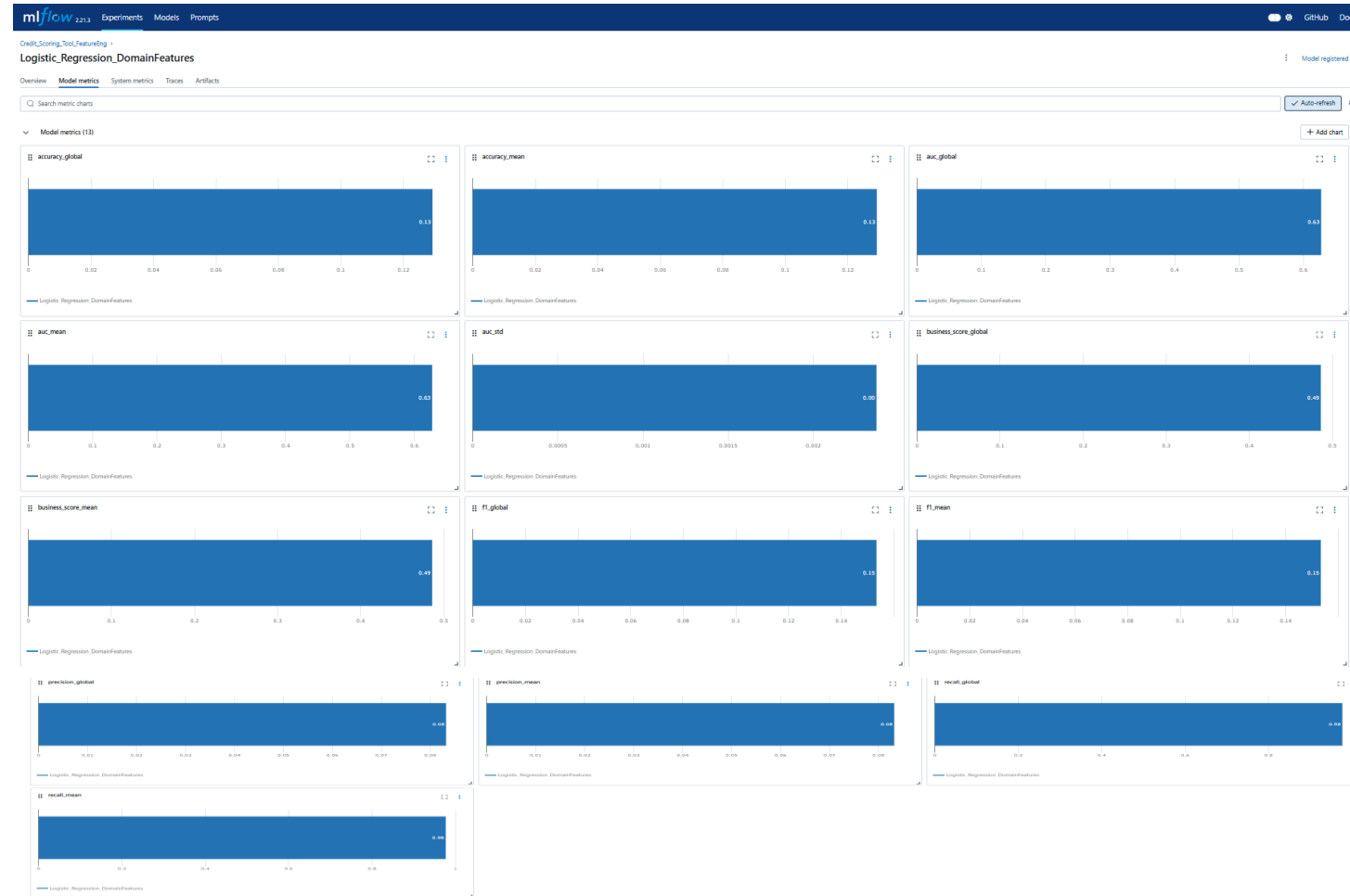
Experiment Name :

« Credit_Scoring_Tool_FeatureEng »

AUC moyen	: 0.628 ± 0.002
AUC global	: 0.628
Score métier moyen	: 0.486
Score métier global	: 0.486
Accuracy moyen	: 0.129
Accuracy global	: 0.129
Précision moyenne	: 0.083
Précision globale	: 0.083
Rappel moyen	: 0.977
Rappel global	: 0.977
F1-score moyen	: 0.153
F1-score global	: 0.153

avec feature Engineering – caractéristiques basées sur le domaine

Résultat de la modélisation avec l'algorithme de classification de la Régression Linéaire.



SCORE METIER
> 0



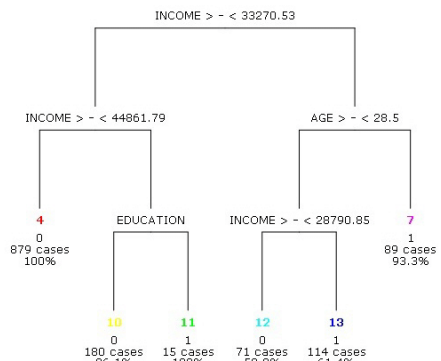
Minimisation des
erreurs coûteuses

- Rappel très fort (0.97) mais précision très faible.
- Beaucoup de faux positifs. opportunité perdue.



Model Registry :

« Credit Scoring Tool Baseline »



Experiment Name :

« Credit_Scoring_Tool_FeatureEng »

=== RÉSULTATS VALIDATION CROISÉE ===

AUC moyen : 0.736 ± 0.001

AUC global : 0.736

Score métier moyen : 0.533

Score métier global : 0.533

Accuracy moyen : 0.919

Accuracy global : 0.919

Précision moyenne : 0.606

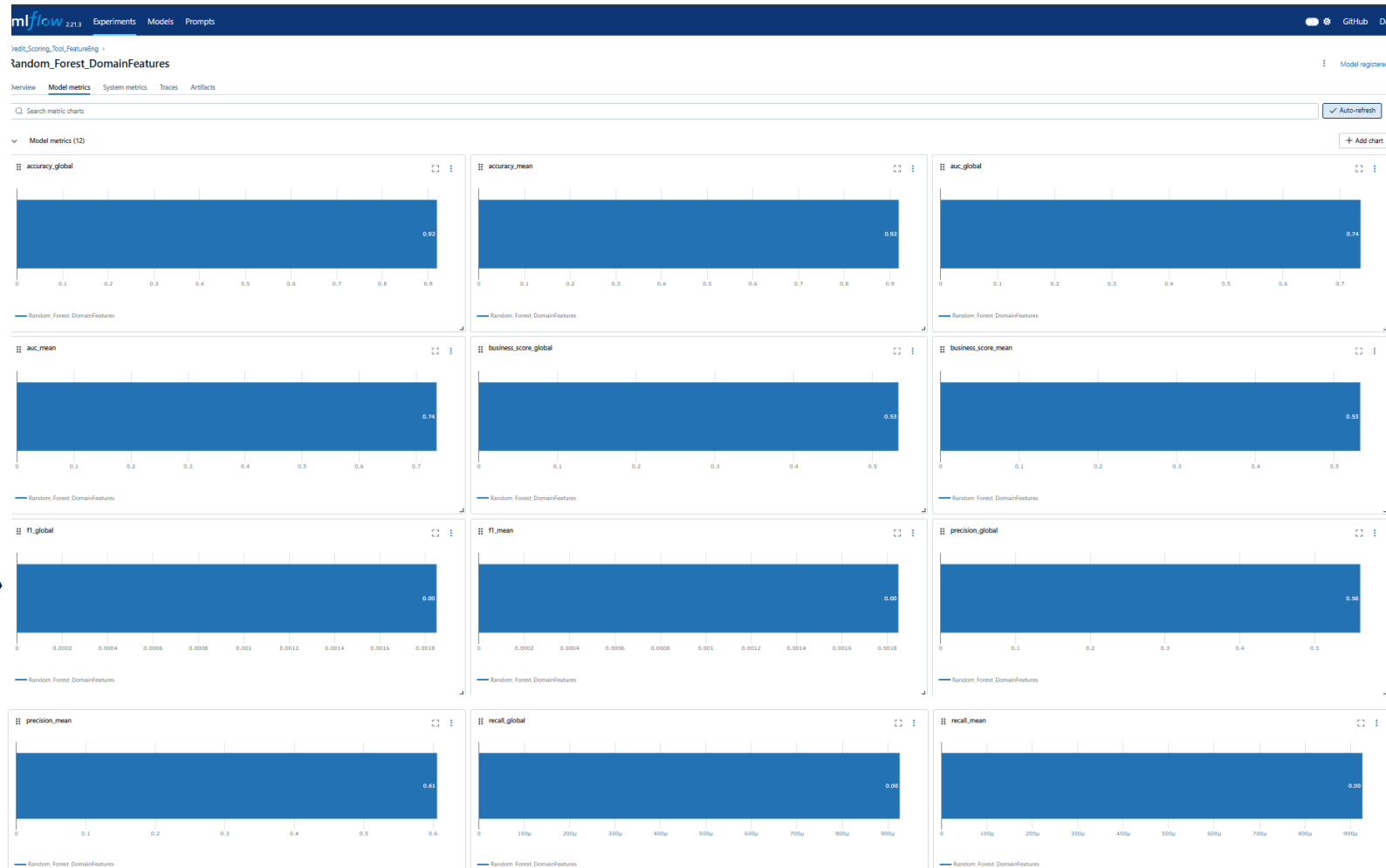
Précision globale : 0.561

Rappel moyen : 0.001

Rappel global : 0.001

F1-score moyen : 0.002

F1-score global : 0.002

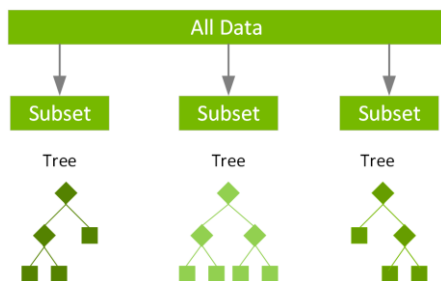


- Score AUC (0.736) correct et précision correcte.
- Rappel trop faible, peu de détections.



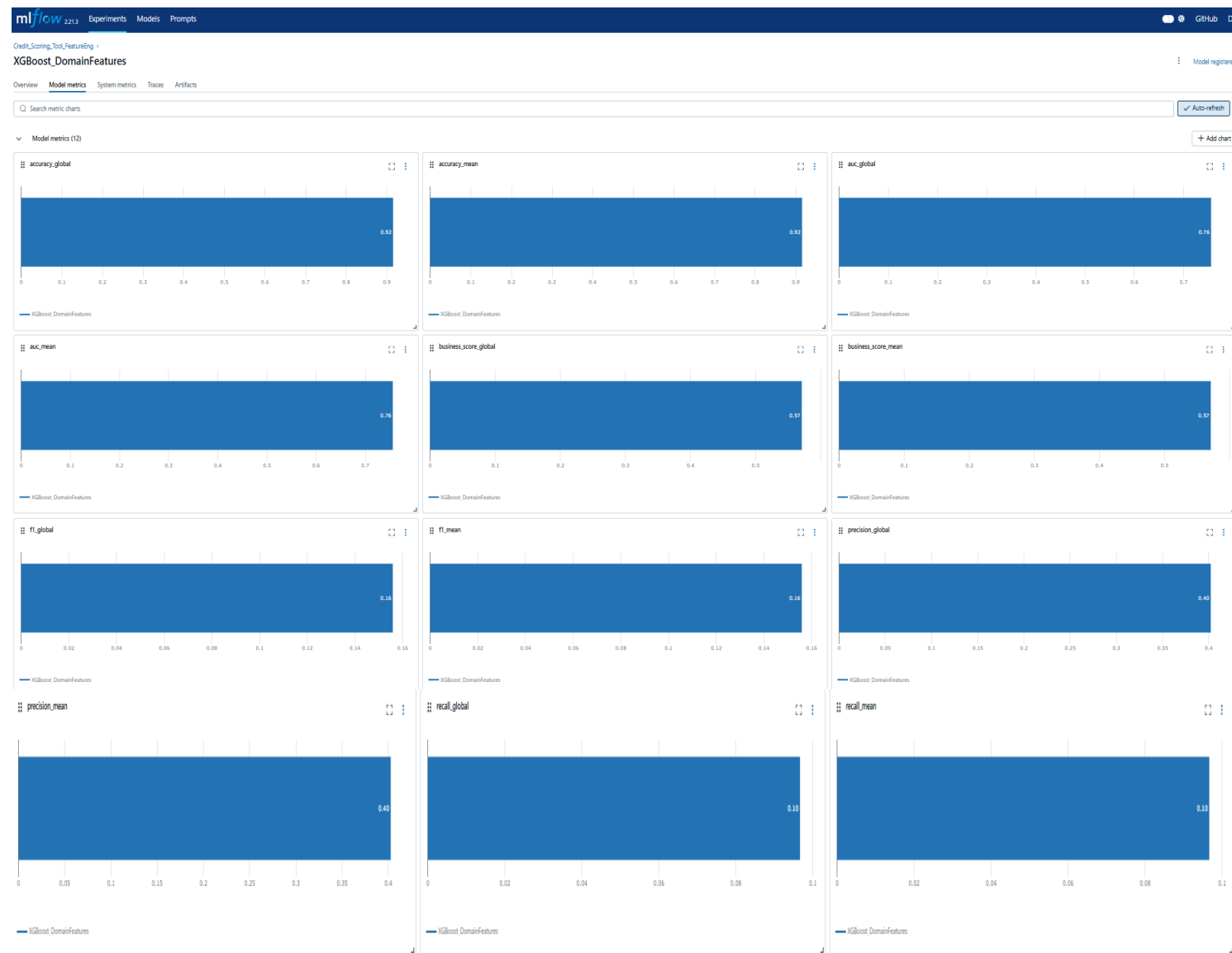
Model Registry :

« Credit Scoring Tool Baseline »



Experiment Name :

« Credit_Scoring_Tool_FeatureEng »



- Meilleure AUC (0.756) avec bon équilibre global.
- Modèle robuste mais le rappel reste encore limité.

=== RÉSULTATS VALIDATION CROISÉE ===

AUC moyen : 0.756 ± 0.002

AUC global : 0.756

Score métier moyen : 0.572

Score métier global : 0.572

Accuracy moyen : 0.915

Accuracy global : 0.915

Précision moyenne : 0.403

Précision globale : 0.403

Rappel moyen : 0.100

Rappel global : 0.100

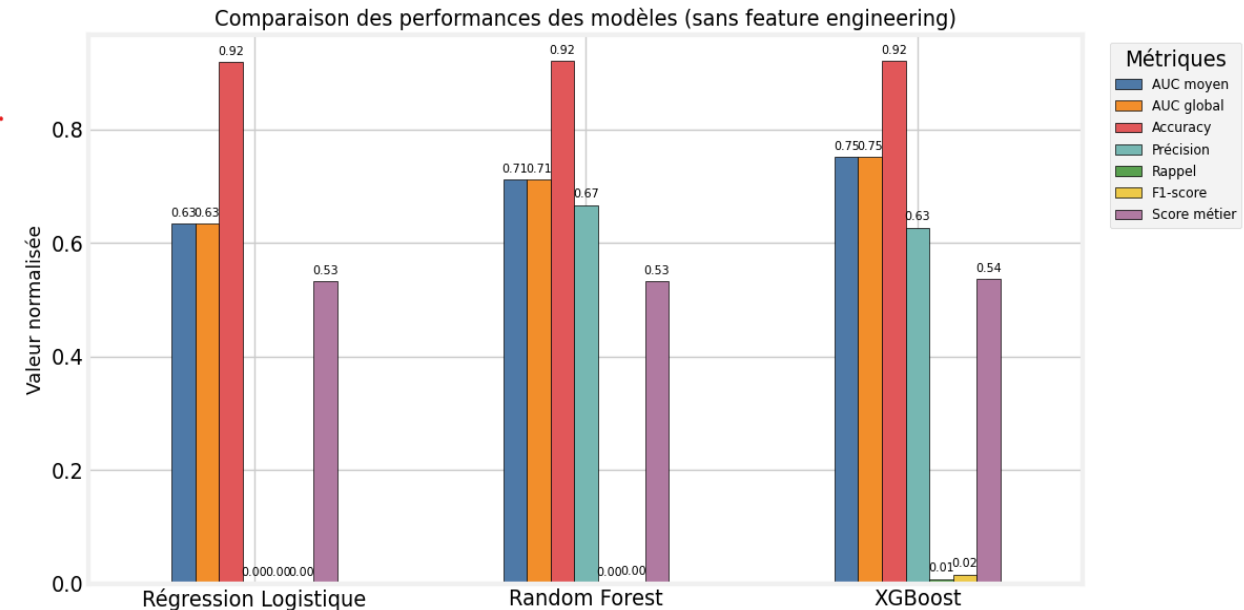
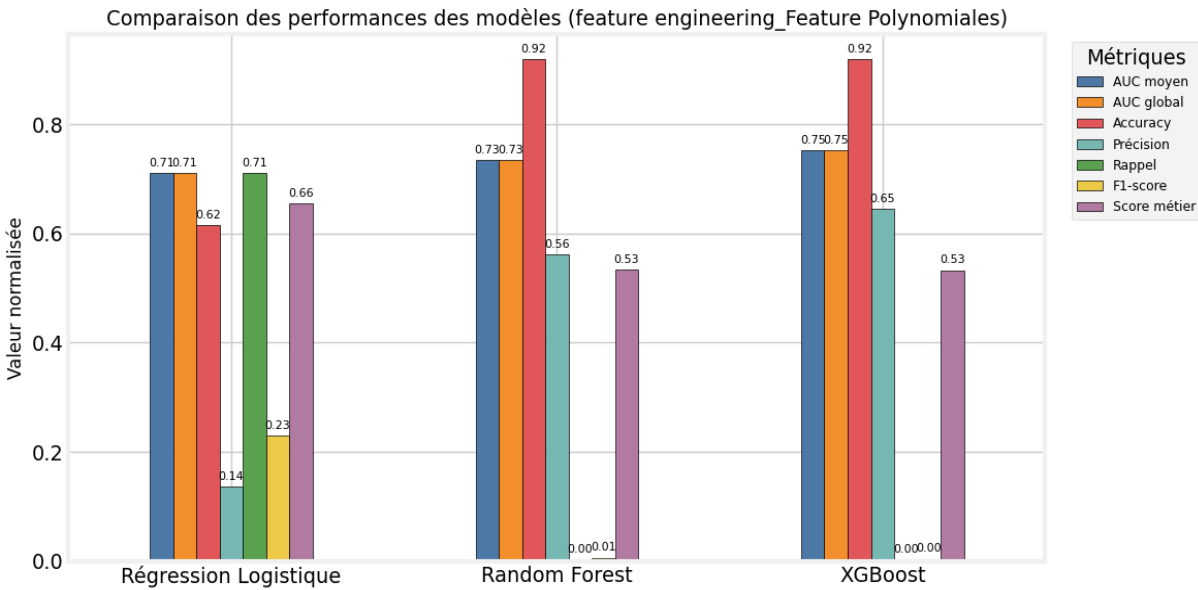
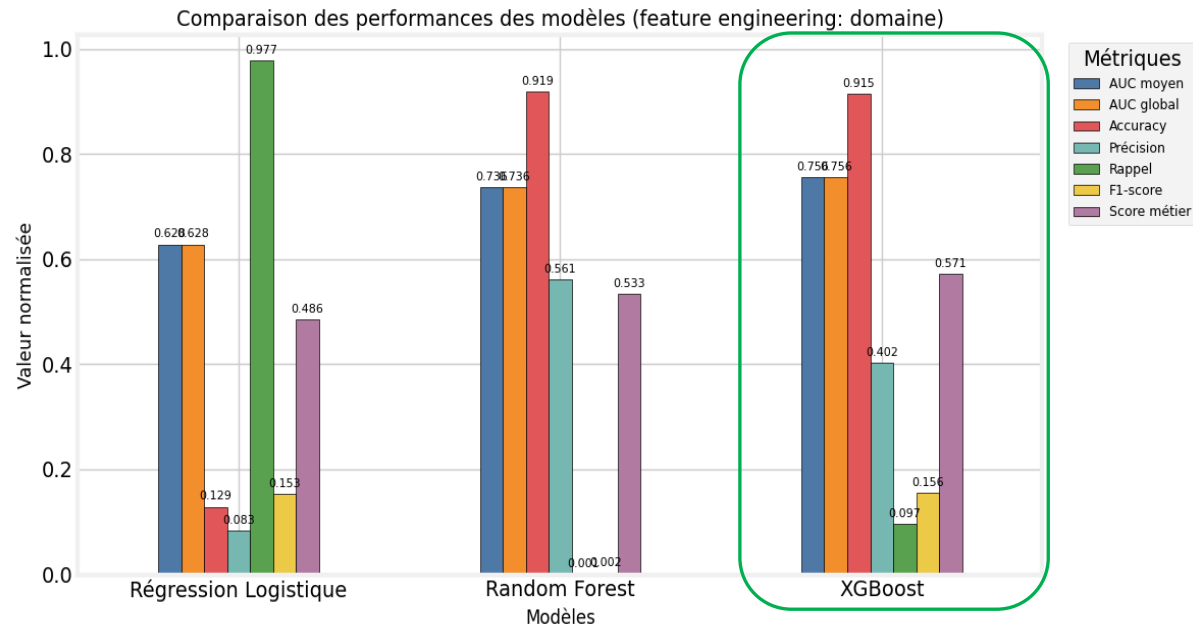
F1-score moyen : 0.161

F1-score global : 0.161

SYTHESE DES RÉSULTATS

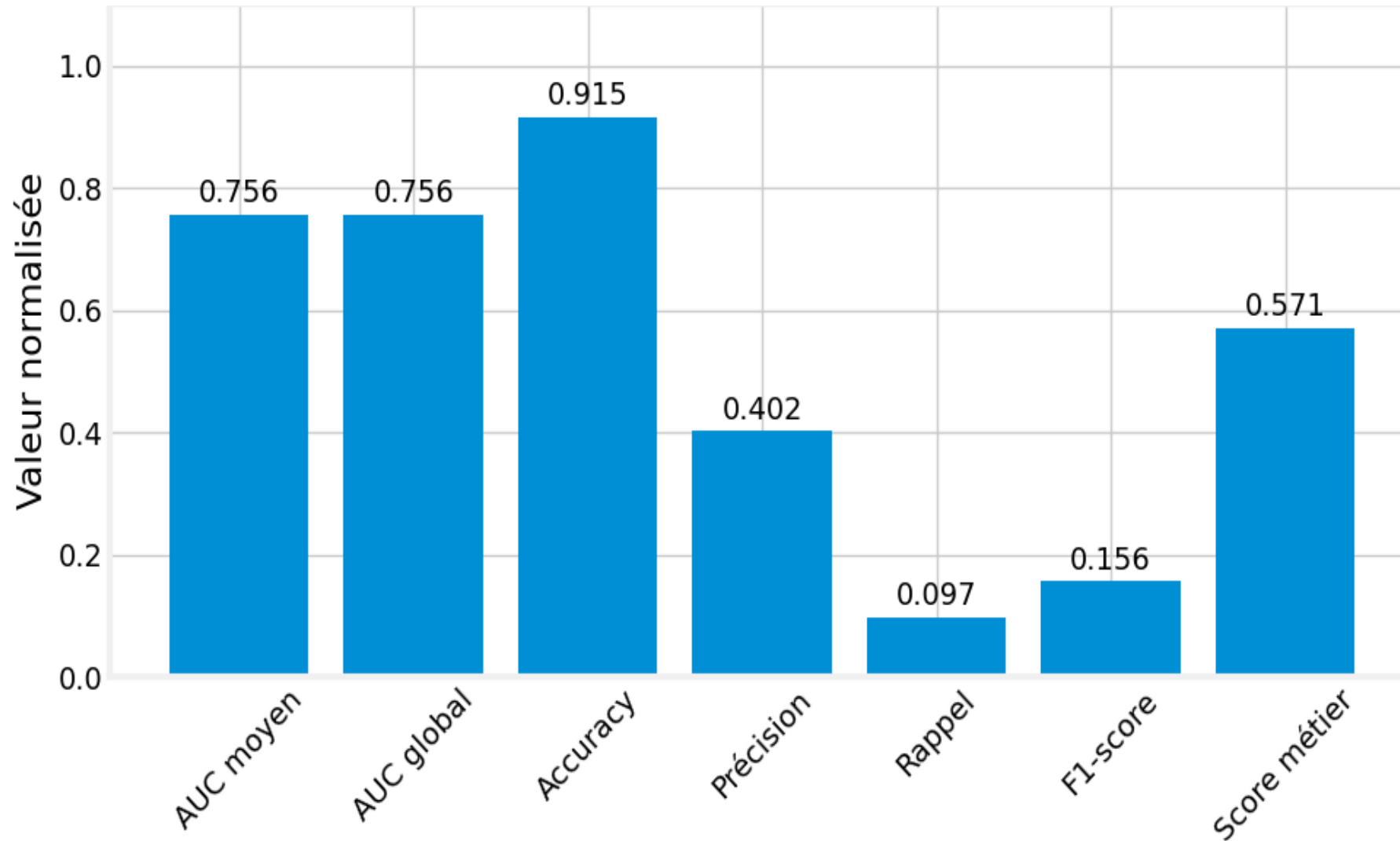


- Le modèle **XGBoost** est le plus performant (score métier le plus performant).
- Les **caractéristiques basées** sur le domaine améliorent surtout le rappel.



SYTHESE DES RÉSULTATS

Le modèle XGBoost avec les caractéristiques basées sur le domaine offre un meilleur compromis global.

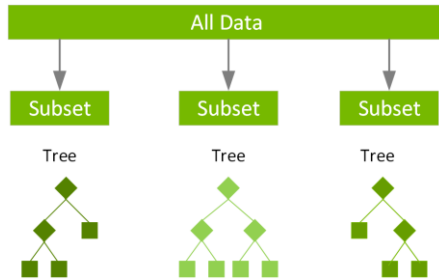


MODÉLISATION DU MEILLEURE MODELE SUR 20 CARACTÉRISTIQUES



Model Registry :

« Credit Scoring Tool Baseline »



Experiment Name :

« Credit_Scoring_Tool_FeatureEng_Production_XgBoost »

=== MÉTRIQUES DE PERFORMANCE ===

AUC : 0.691

Accuracy : 0.670

Précision : 0.140

Recall : 0.602

F1-score : 0.227

Score métier : 0.641

=== Classification Report ===

	precision	recall	f1-score	support
Non Défaut	0.95	0.68	0.79	282686
Défaut	0.14	0.60	0.23	24825
accuracy			0.67	307511
macro avg	0.55	0.64	0.51	307511
weighted avg	0.89	0.67	0.74	307511

Credit_Scoring_Tool_FeatureEng_Production_XgBoost

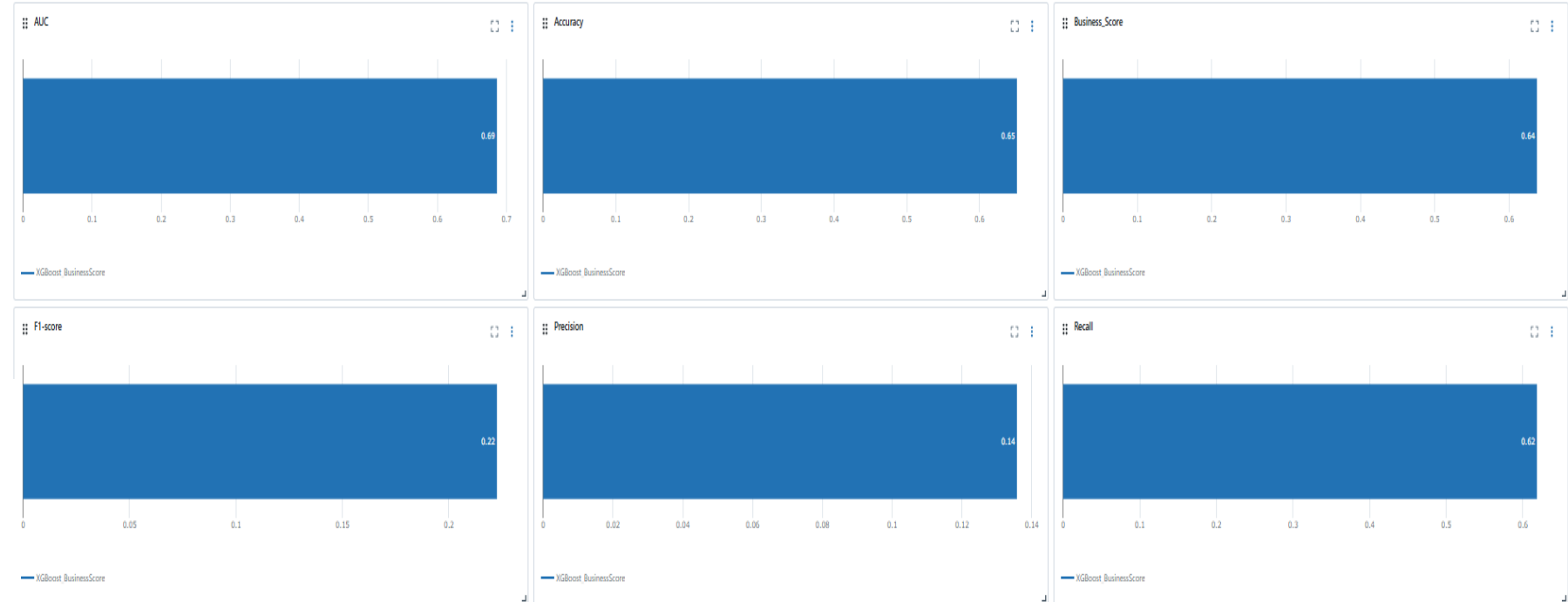
XGBoost_BusinessScore

Overview Model metrics System metrics Traces Artifacts

Search metric charts

Model registered

Model metrics (6)



+ Add section

SYTHESE DES RÉSULTATS

MODÉLISATION DU MEILLEUR MODELE 20 FEATURES

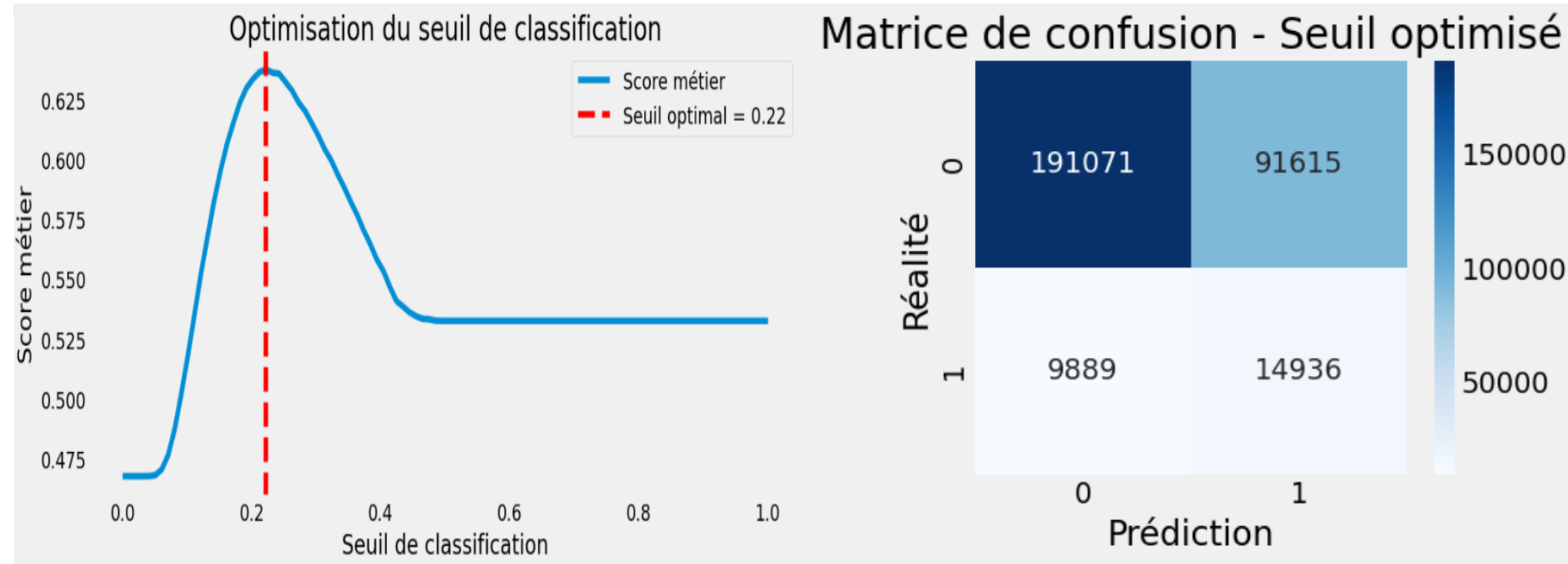


=== MÉTRIQUES DE PERFORMANCE ===

AUC : 0.691
Accuracy : 0.670
Précision : 0.140
Recall : 0.602
F1-score : 0.227
Score métier : 0.641

=== Classification Report ===

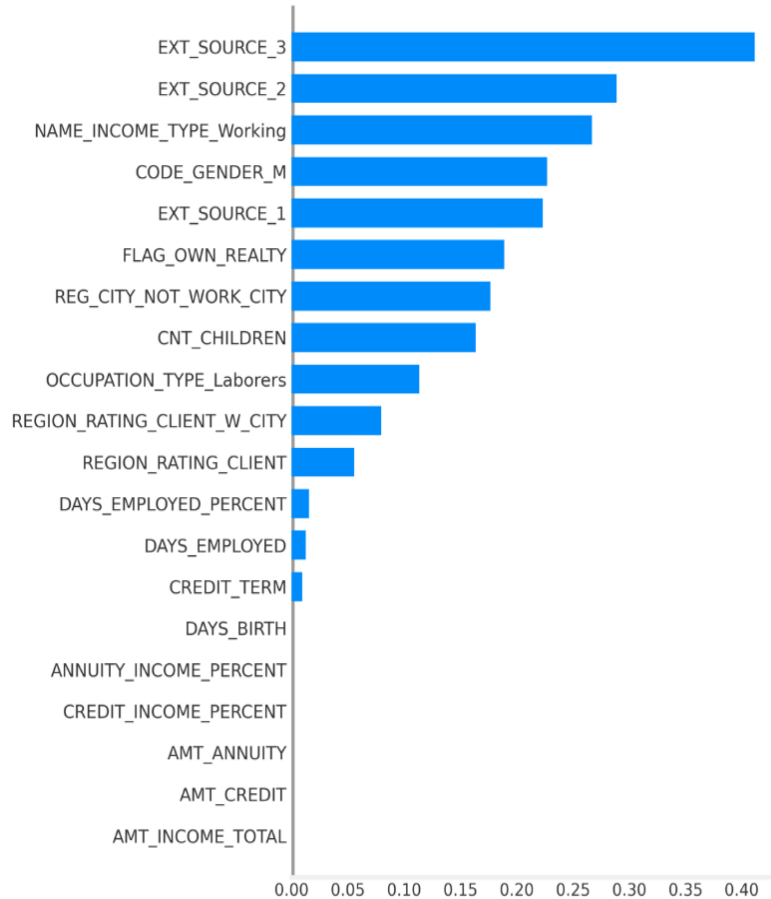
	precision	recall	f1-score	support
Non Défaut	0.95	0.68	0.79	282686
Défaut	0.14	0.60	0.23	24825
accuracy			0.67	307511
macro avg	0.55	0.64	0.51	307511
weighted avg	0.89	0.67	0.74	307511



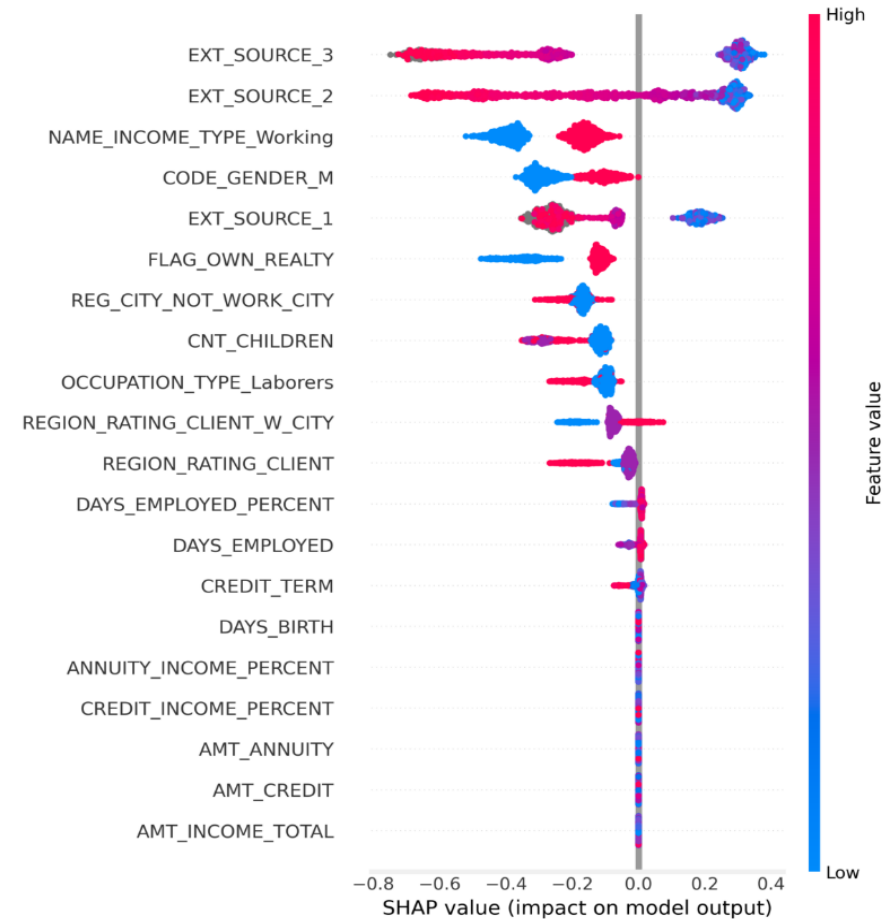
- Le seuil optimal (0.22) maximise le score métier à 0.641.
- XGBoost détecte bien les défauts (rappel = 0.60), malgré une précision faible (0.14).
- Bon compromis entre performance globale et impact métier.



Impact global moyen des 20 variables



Distribution des impacts SHAP

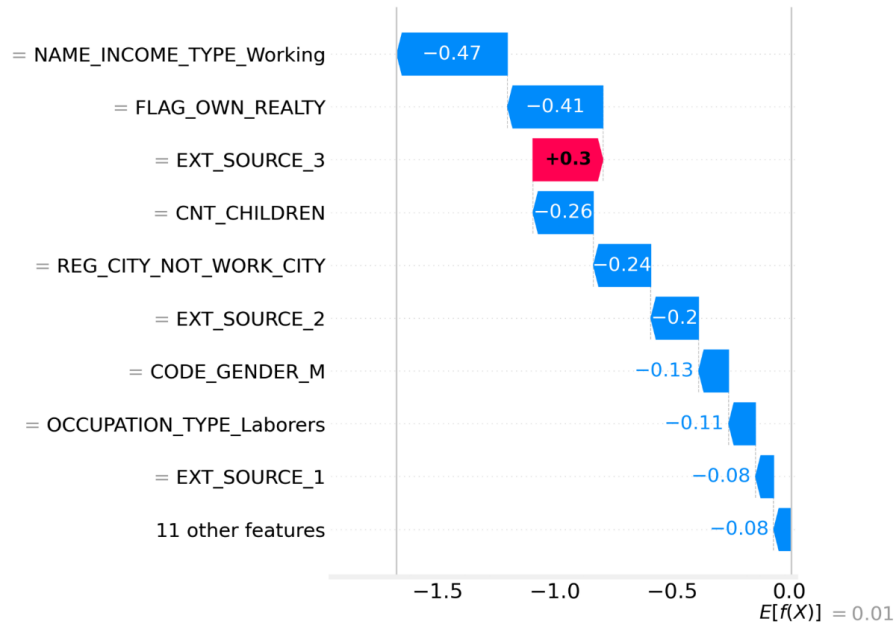


- Les variables EXT_SOURCE_3, EXT_SOURCE_2 et le type de revenu ont le plus fort impact sur le modèle.
- Les caractéristiques **socio-économiques** influencent fortement la prédiction.
- Les effets SHAP sont cohérents et bien séparés selon la valeur des features.



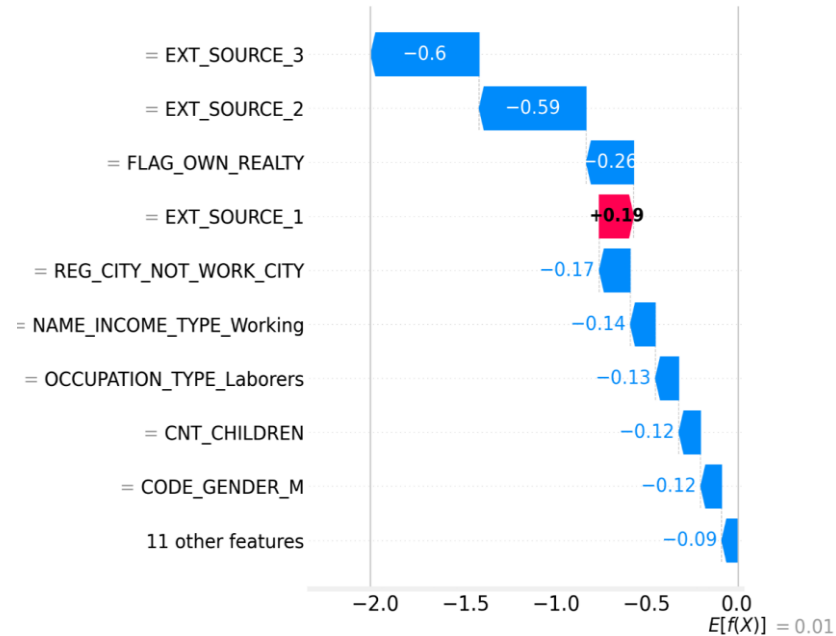
Explication locale - observation 0

= -1.674674



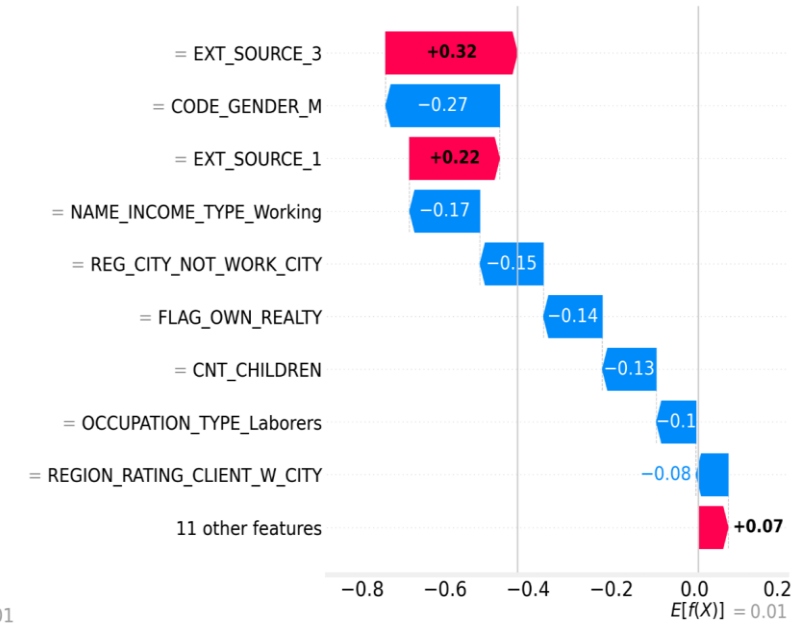
Explication locale - observation 42

= -2.026026

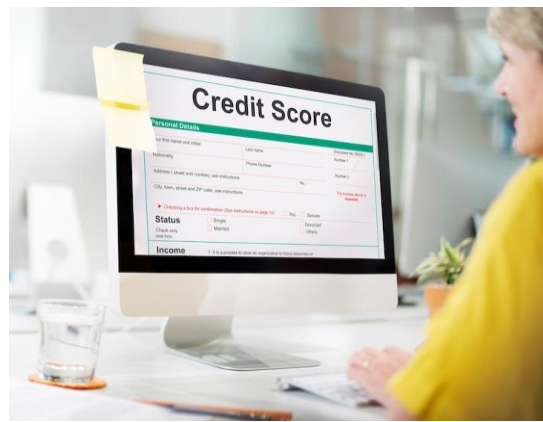


Explication locale - observation 123

= -0.426426



- Les 11 premiers facteurs influencent différemment chaque individu.
- Les scores de solvabilités externes, le revenu et logement ont un **fort impact**, positif ou **négatif**.
- Ces 3 **exemples** nous permettent d'expliquer chaque prédiction de façon personnalisée.



DÉPLOIEMENT

PIPELINE DE DÉPLOIEMENT



DÉVELOPPEMENT DE L'API



Gestion des « endpoints », logique de scoring...

CREATION DE L'INTERFACE UTILISATEUR (UI)



Création du front end pour tester l'API

VERSIONNING



Stockage des versions

DÉPLOIEMENT API



Mise en ligne de l'API (Backend)

DÉPLOIEMENT UI



Mise en ligne de l'interface utilisateur (Frontend)

<https://github.com/012024j/streamlit-fastapi-app>



TEST UNITAIRE

L'objectif du test unitaire est de tester l'authenticité d'une portion d'un programme.
Ici, il s'agit de notre modèle XGBoost que nous avons mis en production.

Test #1



Vérification que le modèle
est entraîné sur *les 20*
features présélectionnés



Test #2



Vérification que le modèle
peut *s'exécuter sur un petit*
lot



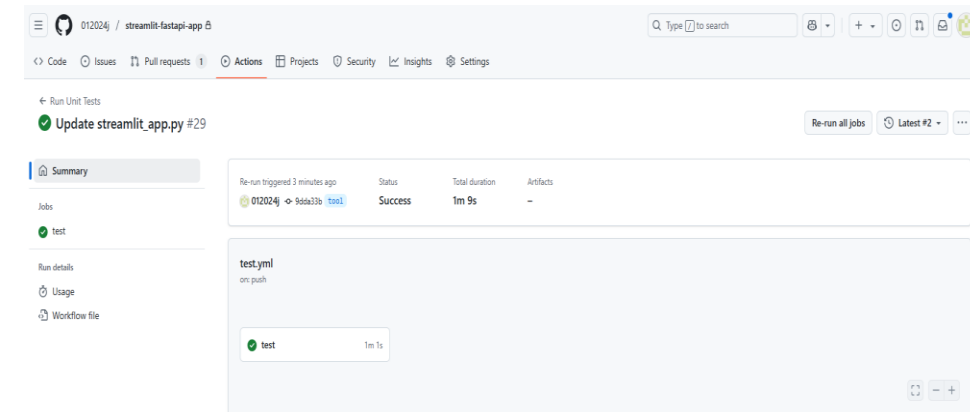
Test #3



Vérification que les
probabilités défaut client
prédites *sont bien situées*
entre 0 et 1



Sortie attendue



DATA DRIFT



rapport_drift_evidently.html

20
Columns

6
Drifted Columns

0.3
Share of Drifted Columns

Data Drift Summary

Data Drift Summary

Drift is detected for 30.0% of columns (6 out of 20).

Search ×						
Column	Type	Reference Distribution	Current Distribution	Data Drift	Stat Test	Drift Score
> DAYS_BIRTH	num			Detected	Wasserstein distance (normed)	7.356832
> DAYS_EMPLOYED_PERCENT	num			Detected	Wasserstein distance (normed)	2.400971
> CREDIT_TERM	num			Detected	Wasserstein distance (normed)	0.575103
> CREDIT_INCOME_PERCENT	num			Detected	Wasserstein distance (normed)	0.293721
> AMT_CREDIT	num			Detected	Wasserstein distance (normed)	0.207334
> AMT_ANNUITY	num			Detected	Wasserstein distance (normed)	0.161102

Le Data Drift signifie (dérive des données) ce qui correspond à un changement de la distribution statistique des données d'entrée (features) entre l'entraînement d'un modèle et son usage réel en production.

Une dérive est détectée sur 30 % des variables (soit 6 colonnes), notamment sur l'âge, l'emploi et le crédit, indiquant un changement significatif du profil des clients.

EXEMPLE DE SCORING CLIENT



Credit Score

Personal Details

Sur first name and initial: _____ Last name: _____ Document No. 2023-1

Nationality: _____ Phone Number: _____ Number 1: _____

Address (street and number), see instructions: _____ City, town, street and ZIP code, see instructions: _____ No. _____ Number 2: _____

► Checking a box for confirmation (see instructions on page 12) ☐ Yes ☐ No ☐ Statute

Status

Check only one box.

☐ Single ☐ Married ☐ Divorced ☐ Others

Income

1. It is a process to allow an organization to focus resources on



EXEMPLE D'UN SCORING CLIENT

Prêt **à** dépendre



Merci de votre attention
-
Question(s) – Réponse(s)