

# Problématique



Mise en œuvre d'un outil de « scoring crédit » indiquant la probabilité de faillite d'un client

Transparence du Modèle

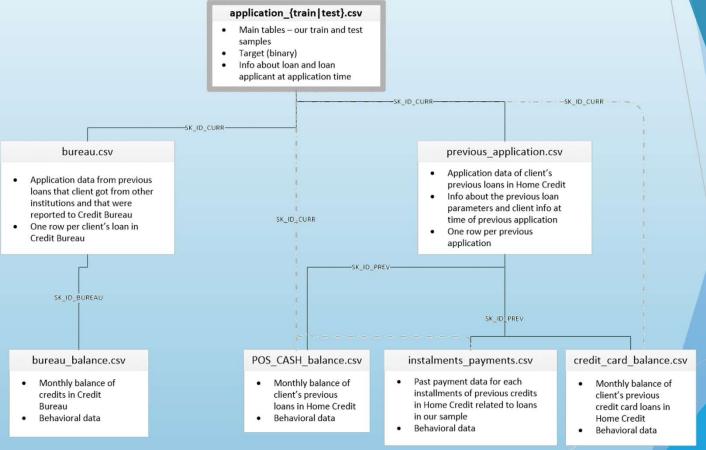
Pouvoir surveiller la dérive du modèle



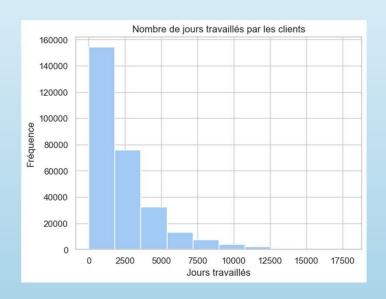
## Le jeu de données

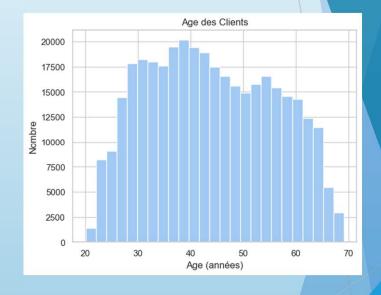


#### Données Kaggle de la compétition Home Credit Default Risk



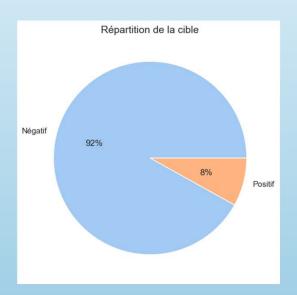




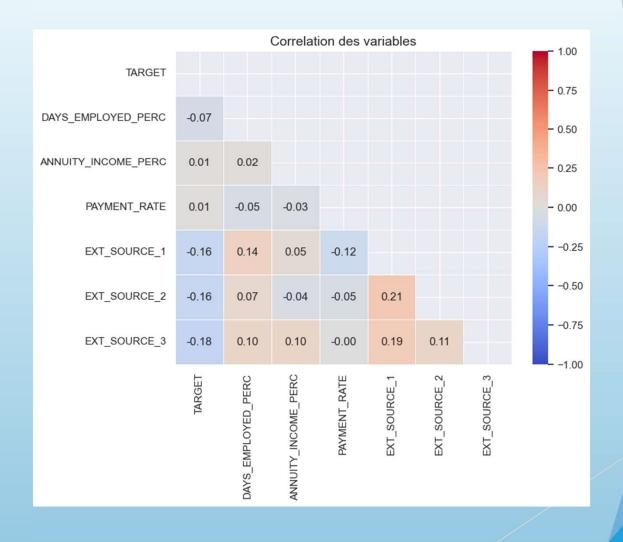




Cible : Capacité du client a rembourser son crédit (variable Target : binaire 1 faillite )









Données initiales (350 000 x 800)





Données sans Target → Application (50 000 x 800)

Données avec Target → Modélisation (300 000 x 800)







Données de Test (60 000 x 800)

Données de Train non équilibrées (240 000 x 800)

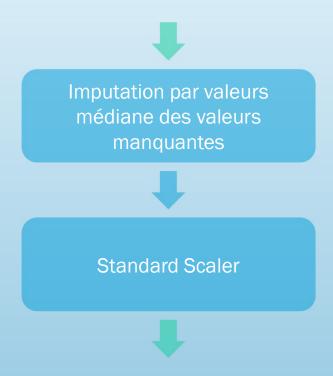
Données de Train équilibrées avec SMOTE (450 000 x 800)



### La Modélisation



#### Pipe d'entrée des données





## Le choix des mesures



#### **AUROC**

Score Métier = 
$$\frac{FN*10+FP}{(TP+TN+FP+FN)*10}$$

Durée



## La démarche



#### Etape 1 – Balayage de différents modèles

Base Line Logistic Regression LGBM Classifier

Dataset équilibré (SMOTE) Logistic Regression LGBM Classifier Rééquilibrage des classes intégré Logistic Regression LGBM Classifier



#### Etape 2 – Recherche des hypers paramètres pour le modèle choisi

GridSearchCV

n estimators colsample bytree max depth learning rate



#### Etape 3 – Affinage du modèle

Calcul du meilleur seuil de décision

Dernière évaluation du modèle



# MlFlow



Centralisation et Sauvegarde de toutes les modélisations

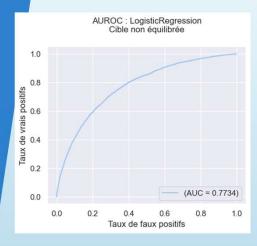
Affinage & compréhension des paramètres

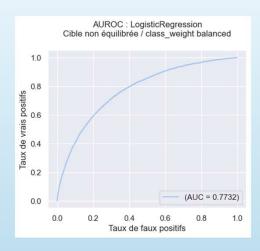
Validation du modèle final

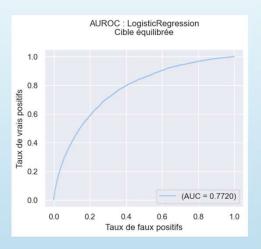


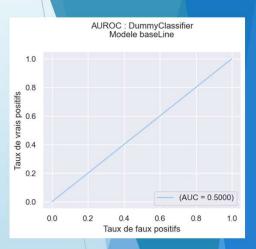
### Les résultats

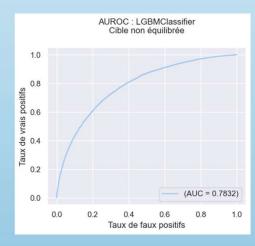


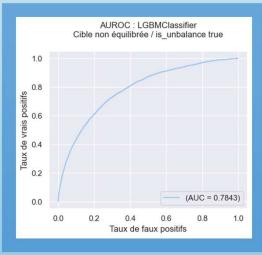


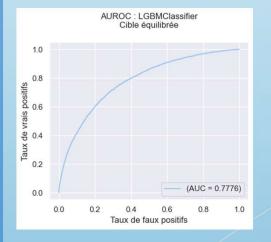








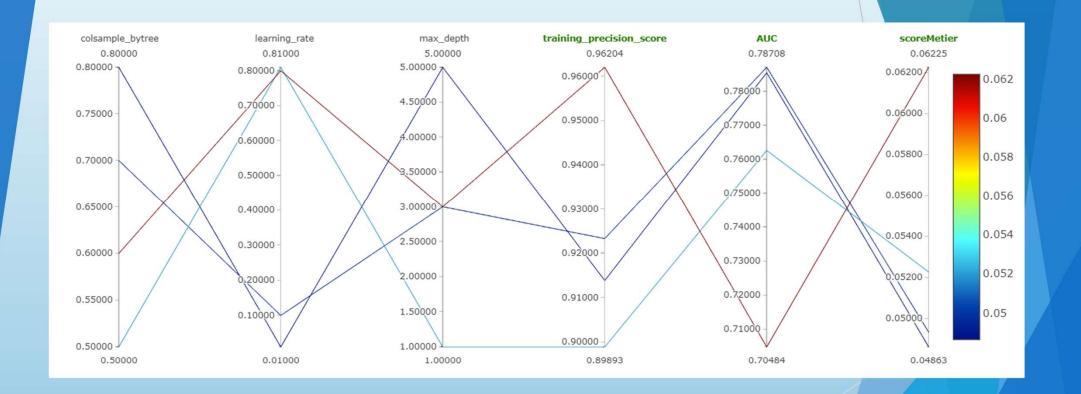






	Modèle	Description	AUC	Accuracy	Score Metier	Durée
0	DummyClassifier	Modele baseLine	0.500000	0.919271	0.080729	93.202768
1	LogisticRegression	Cible non équilibrée	0.773441	0.919840	0.078331	109.341892
2	LGBMClassifier	Cible non équilibrée	0.783152	0.919921	0.077796	106.850640
3	LogisticRegression	Cible non équilibrée / class_weight balanced	0.773172	0.709245	0.050704	153.382084
4	LGBMClassifier	Cible non équilibrée / is_unbalance true	0.784334	0.724220	0.049046	108.007541
5	LogisticRegression	Cible équilibrée	0.771995	0.712416	0.050636	169.095033
6	LGBMClassifier	Cible équilibrée	0.777598	0.919791	0.078760	136.291062



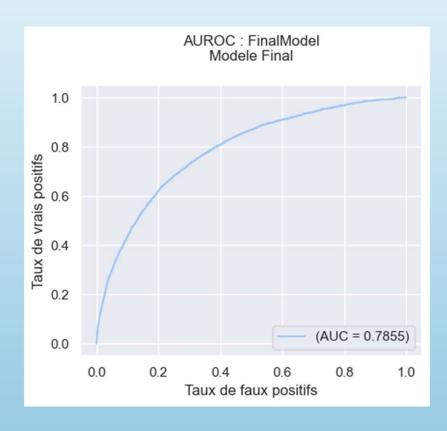




	Modèle	Description	AUC	Accuracy	Score Metier	Durée
0	DummyClassifier	Modele baseLine	0.500000	0.919271	0.080729	88.705140
1	LogisticRegression	Cible non équilibrée	0.773441	0.919840	0.078331	137.189040
2	LGBMClassifier	Cible non équilibrée	0.783152	0.919921	0.077796	104.017422
3	LogisticRegression	Cible non équilibrée / class_weight balanced	0.773172	0.709245	0.050704	105.532030
4	LGBMClassifier	Cible non équilibrée / is_unbalance true	0.784334	0.724220	0.049046	105.933751
5	LogisticRegression	Cible équilibrée	0.771995	0.712416	0.050636	152.704010
6	LGBMClassifier	Cible équilibrée	0.777598	0.919791	0.078760	151.099868
7	LGBMClassifier	Cible non équilibrée / best params	0.762638	0.701668	0.052267	773.148352
8	LGBMClassifier	Cible non équilibrée / best params	0.785474	0.742724	0.048629	439.063881
9	LGBMClassifier	Cible non équilibrée / best params	0.787079	0.765861	0.049345	302.474519
10	LGBMClassifier	Cible non équilibrée / best params	0.704837	0.835534	0.062250	306.400169



#### Modèle Final



Seuil affiné : 0.52763

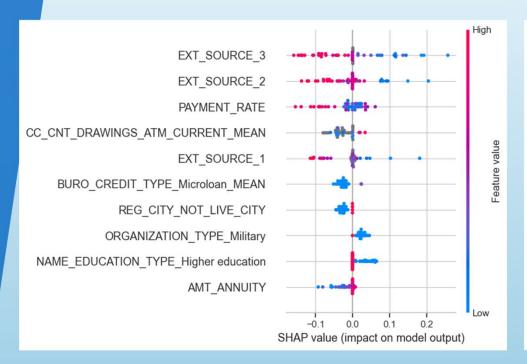
Score Métier: 0.48600

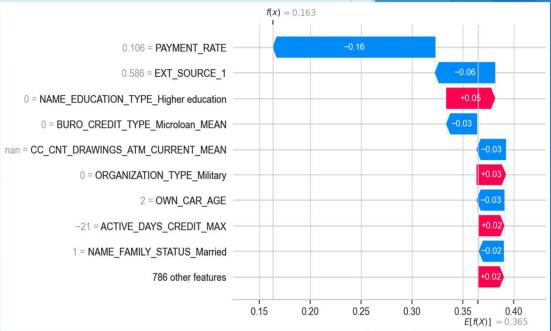
AUC : 0.78547



# Analyse feature importance



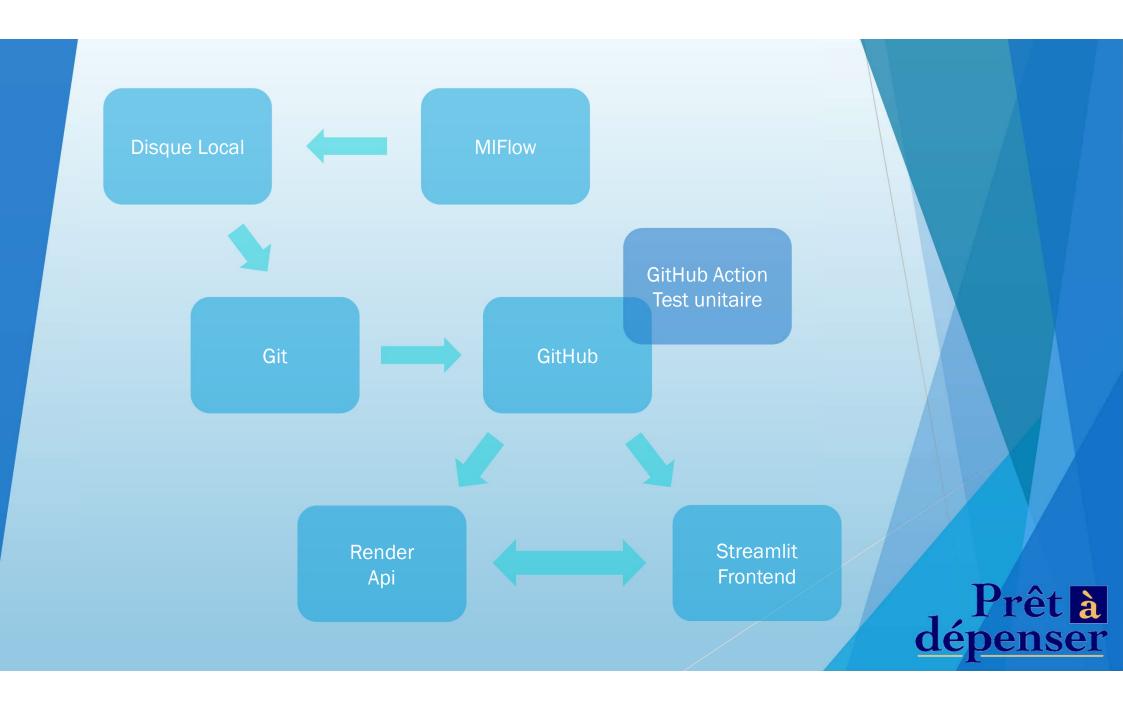


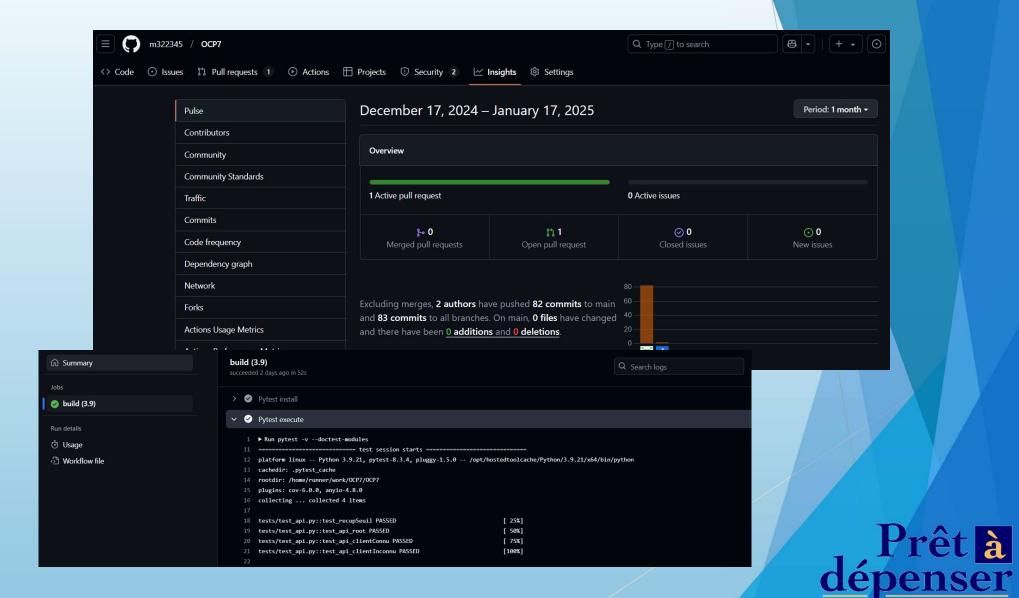




## Le déploiement

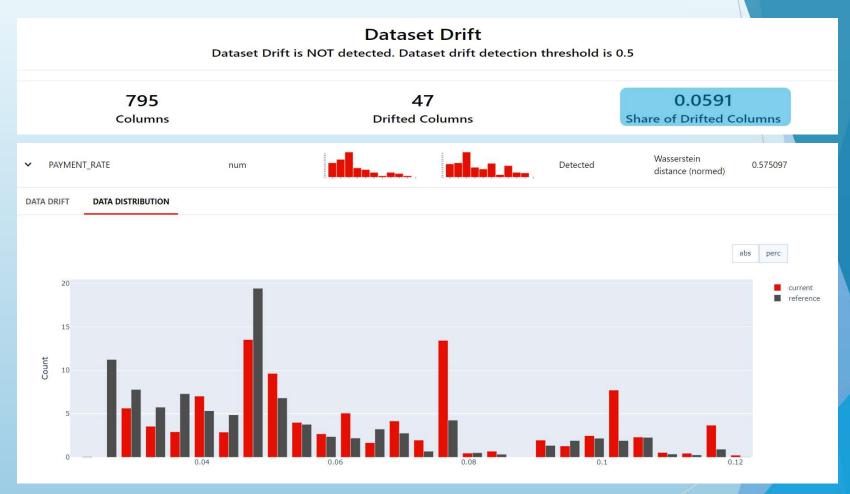






# Analyse data drift







# Démonstration de l'application



#### URL de l'application : https://ocp7-froidure.streamlit.app/



