

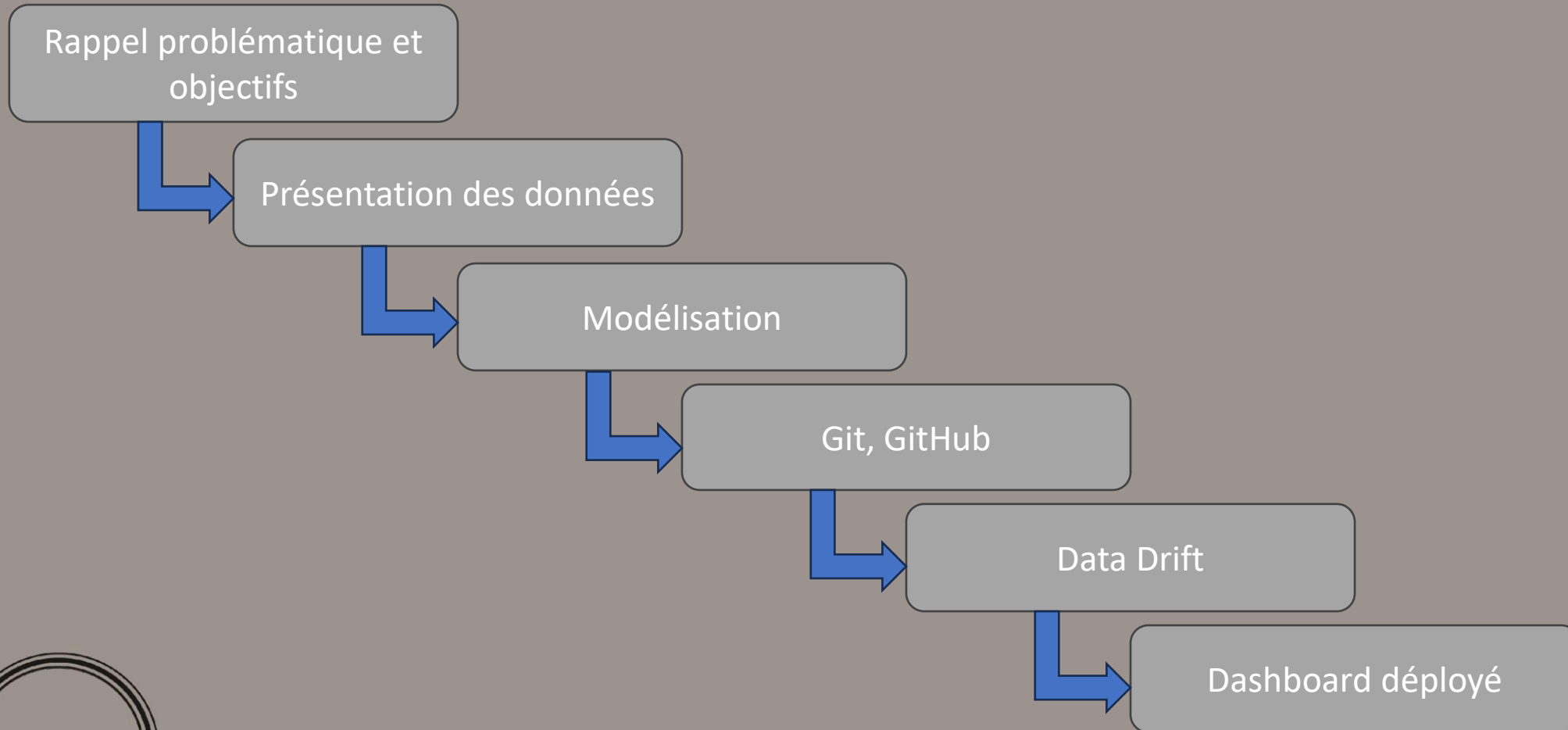
Implémentez un modèle de Scoring

Projet 7 – OpenClassrooms

Stéphane LUBIN



Sommaire



Problématique et objectifs

Problématique

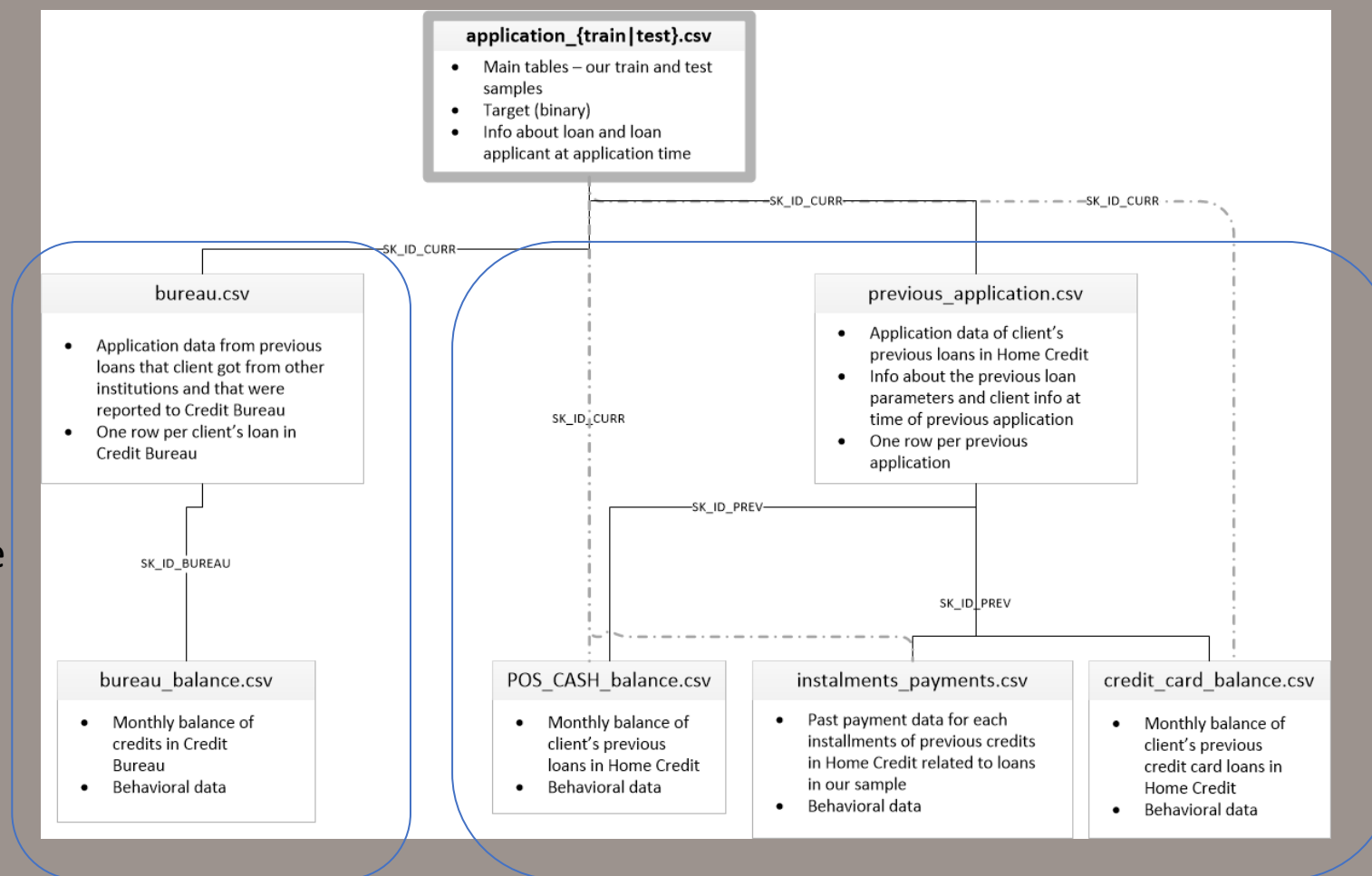
La société « Prêt à dépenser » souhaite mettre en place un modèle de scoring pour la prise de décision sur l'attribution ou le refus d'un prêt

Objectifs

- Modèle automatique de scoring
- Dashboard interactif
- Mettre place MLOps
- Test Data Drift



Présentation jeu de données

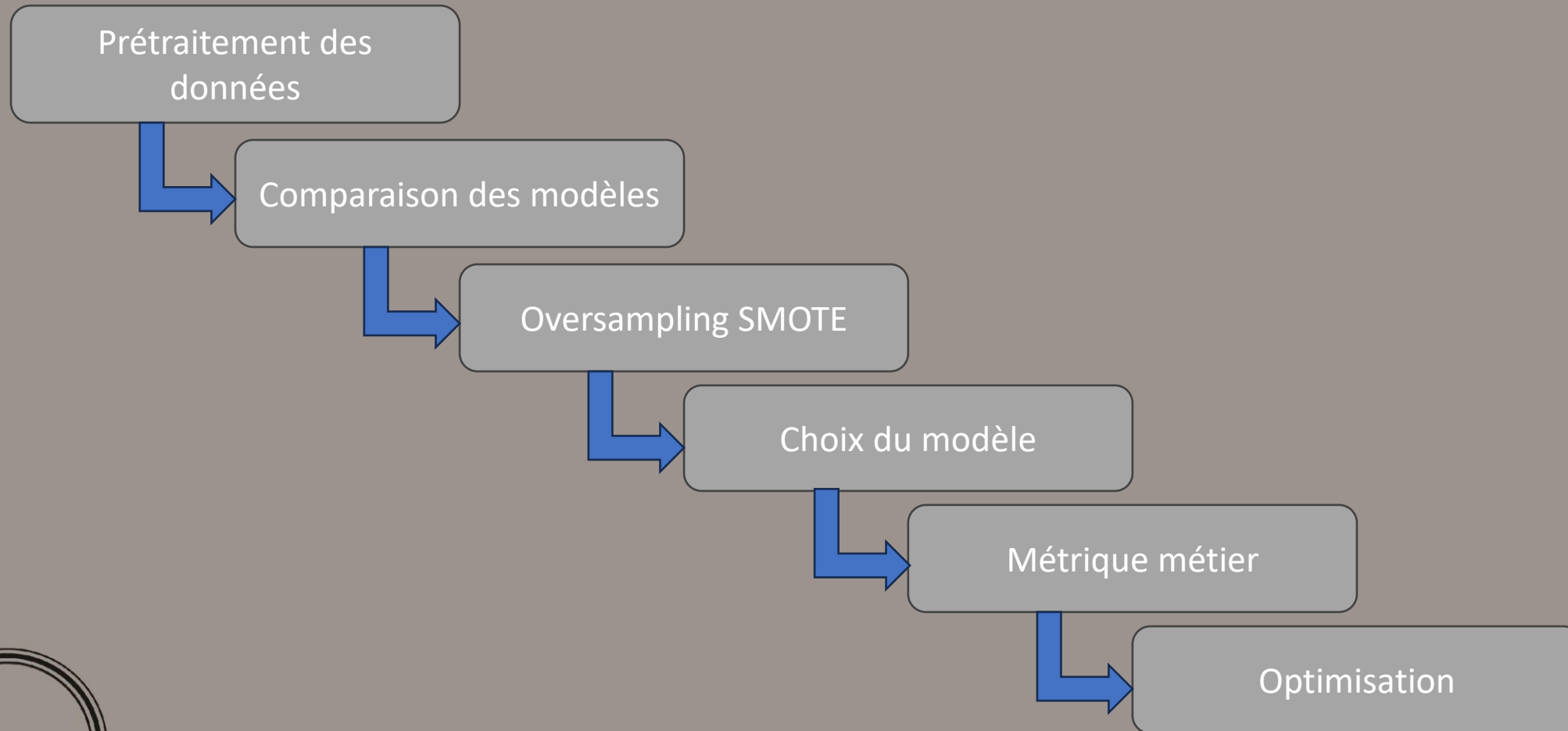


Données organisme extérieur

Données antérieur, prêt, paiement et balance

Prêt à dépenser

Modélisation



Prétraitement des données

- Kernel Kaggle
- Traitement des valeurs aberrantes
- Création de nouvelles variables tels que le taux de paiement
- Création des caractéristiques spécifiques pour les demandes approuvées et refusées
- Création des caractéristiques spécifiques pour les demandes ouvertes et fermées
- Encodage One hot pour les variables de type catégorie
- Fusion des fichiers

	Lignes	Colonnes	Col type float	Col type int	Col type obj	NaN (%)	Double (%)
bureau.csv	1716428	17	8	6	3	13.5	0.0
bureau_balance.csv	27299925	3	0	2	1	0.0	0.0
previous_application.csv	1670214	37	15	6	16	17.98	0.0
POS_CASH_balance.csv	10001358	8	2	5	1	0.07	0.0
installments_payments.csv	13605401	8	5	3	0	0.01	0.0
credit_card_balance.csv	3840312	23	15	7	1	6.65	0.0
application_train.csv	307511	122	65	41	16	24.4	0.0

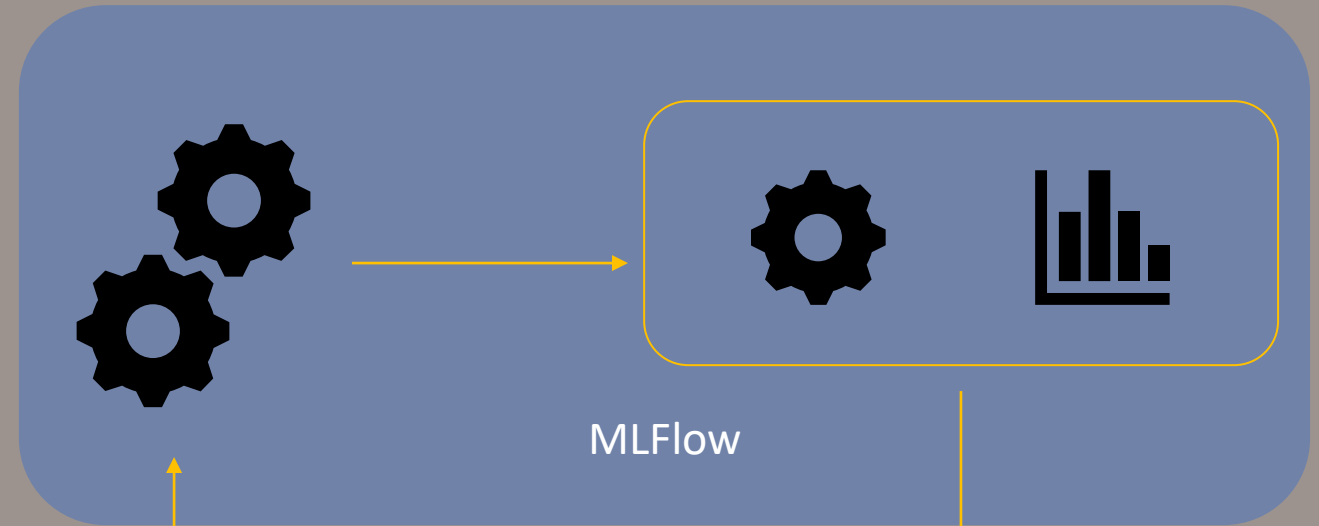
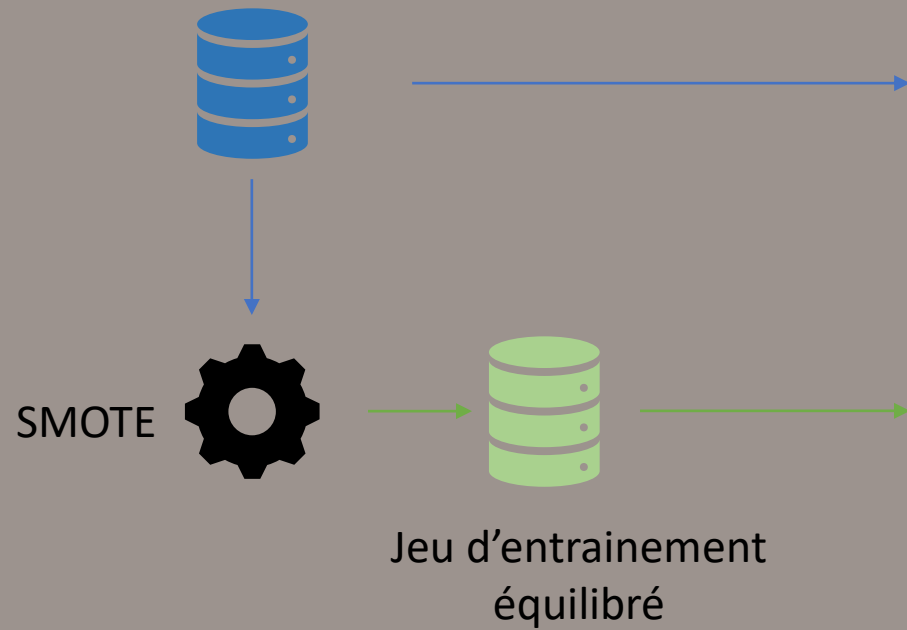
Before Preprocessing. Train samples: (307511, 122), test samples: (48744, 121)

After Preprocessing. Train shape: (307507, 798), test shape: (48744, 798)



SGD Classifier
Random Forest Classifier
XGB Classifier
LGBM Classifier

Jeu d'entraînement



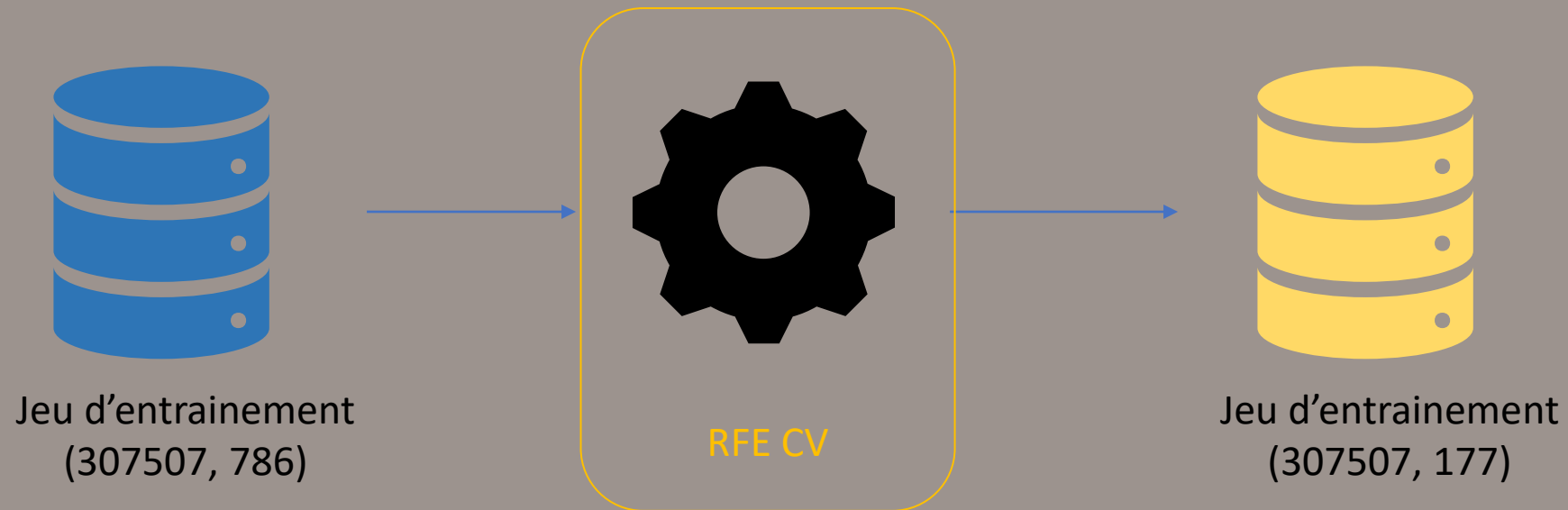
Jeu de validation



MLFlow Tracking UI



Features Sélection



Métrique métier

$$TOT = VP * TVP + VN * TVN - FP * TFP - FN * 20 * TFN$$

$$GM = (VP + FP) * TVP + (VN + FN) * TVN$$

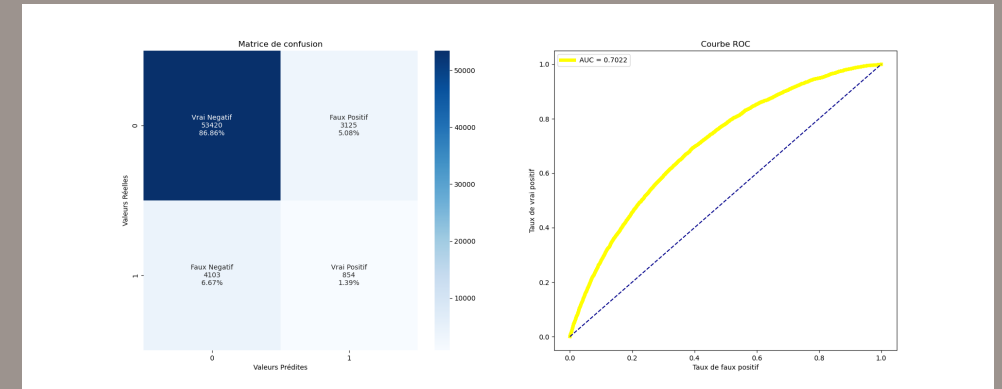
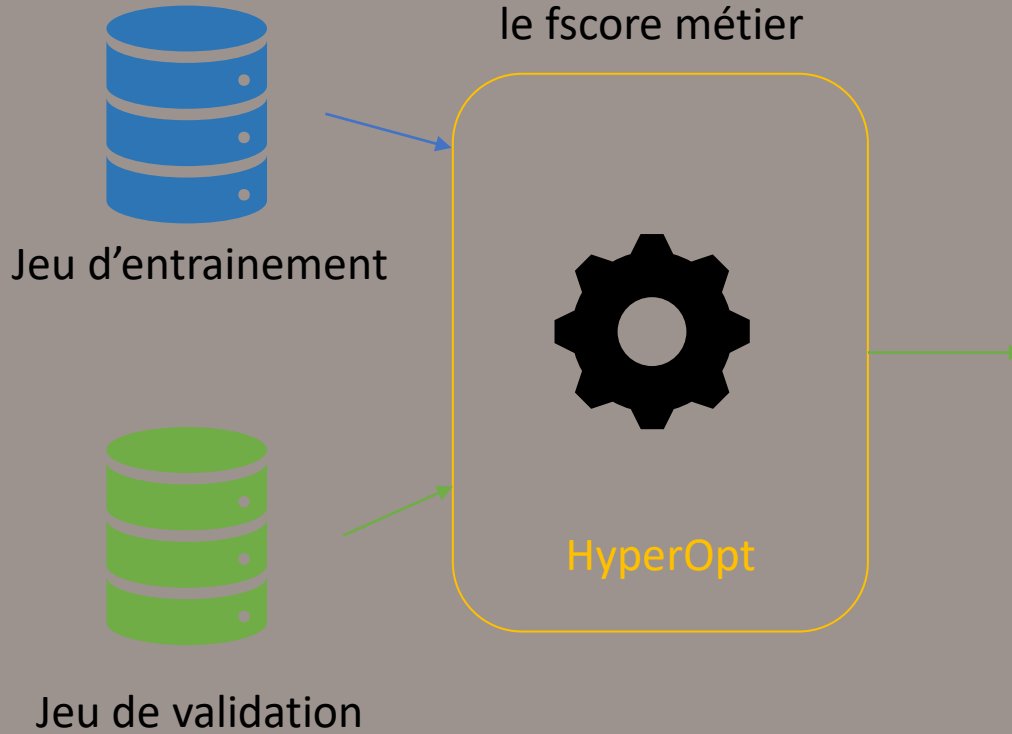
$$PM = -(FP + VP) * TFP - (FN + VN) * 20 * TFN$$

$$Score = (TOT - PM) / (GM - PM)$$

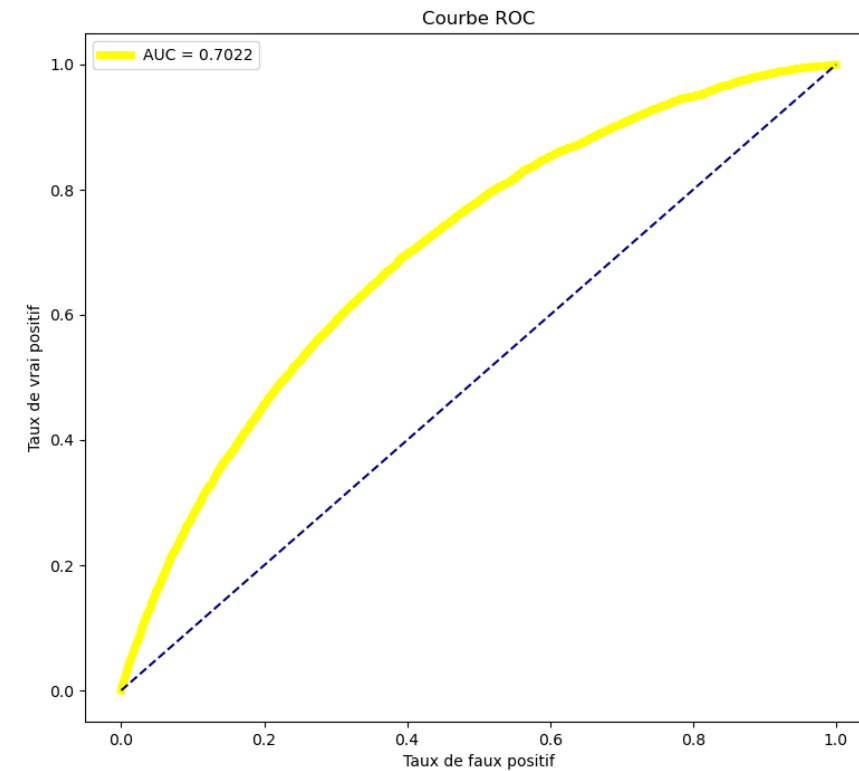
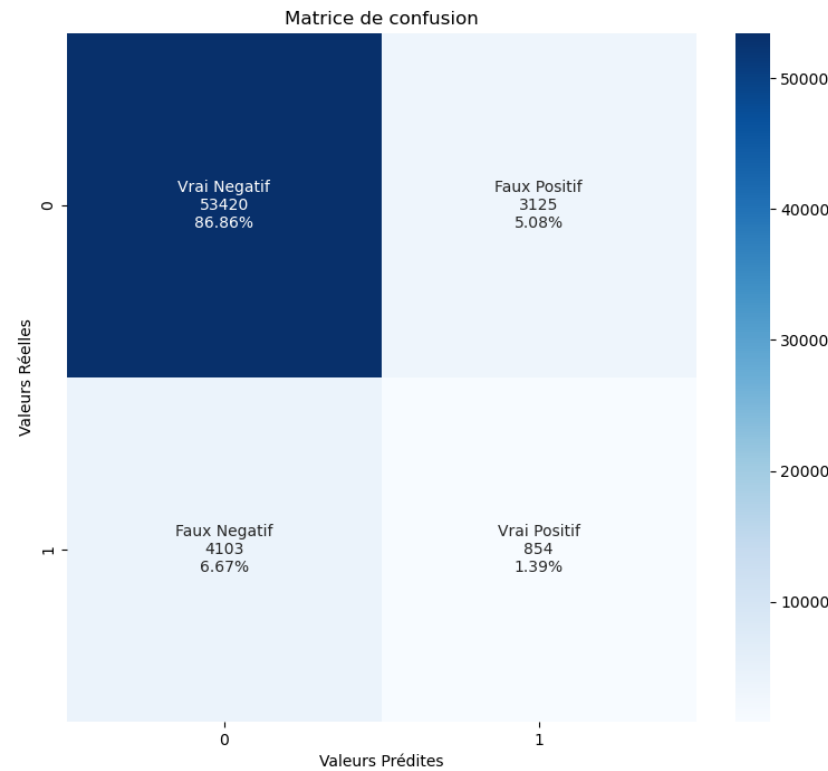


Optimisation

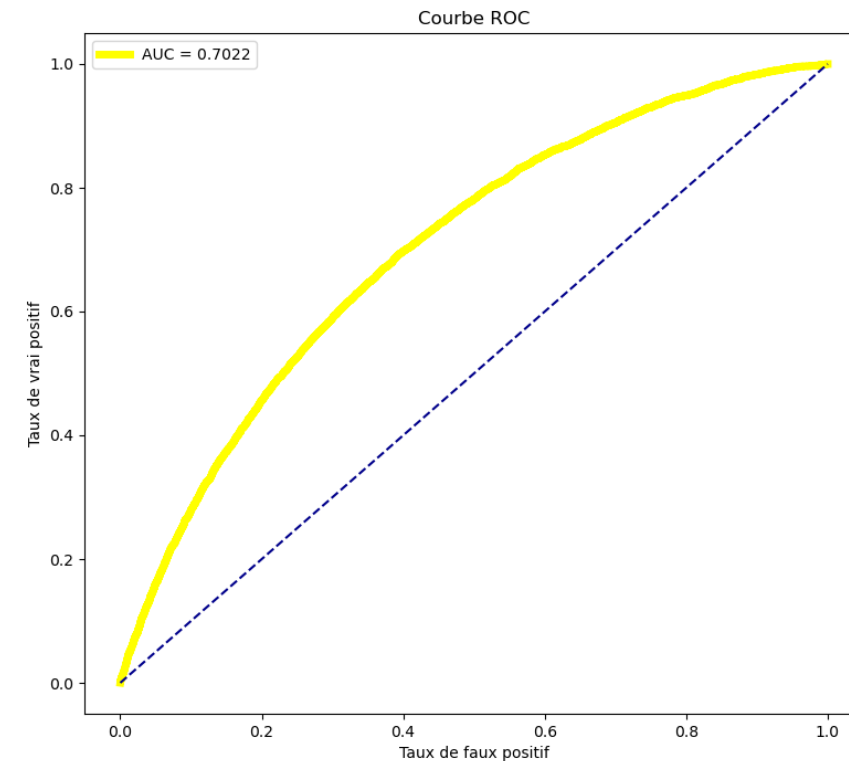
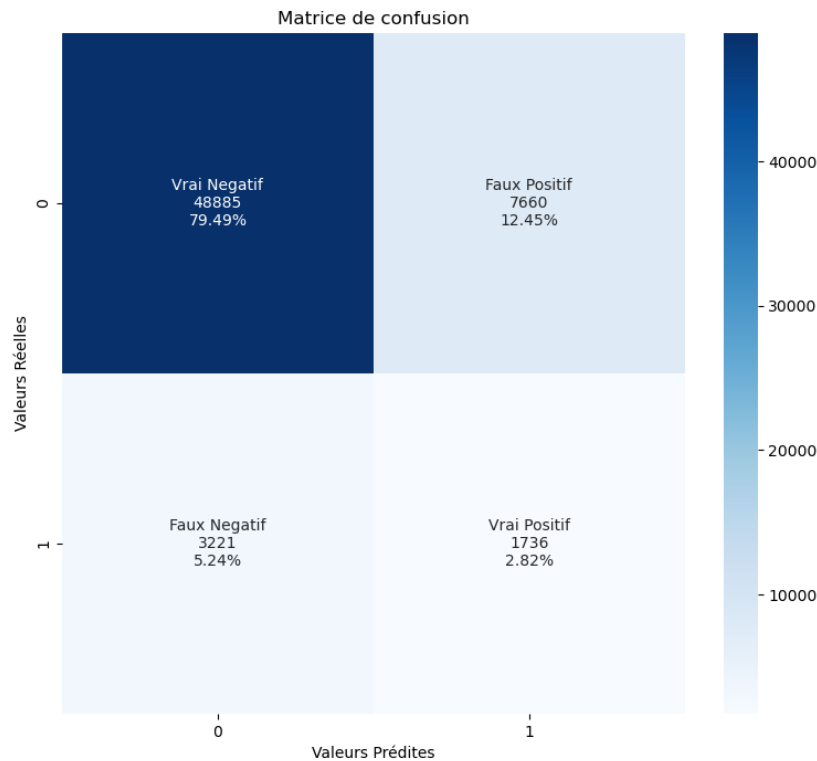
On cherche à minimiser
le fscore métier



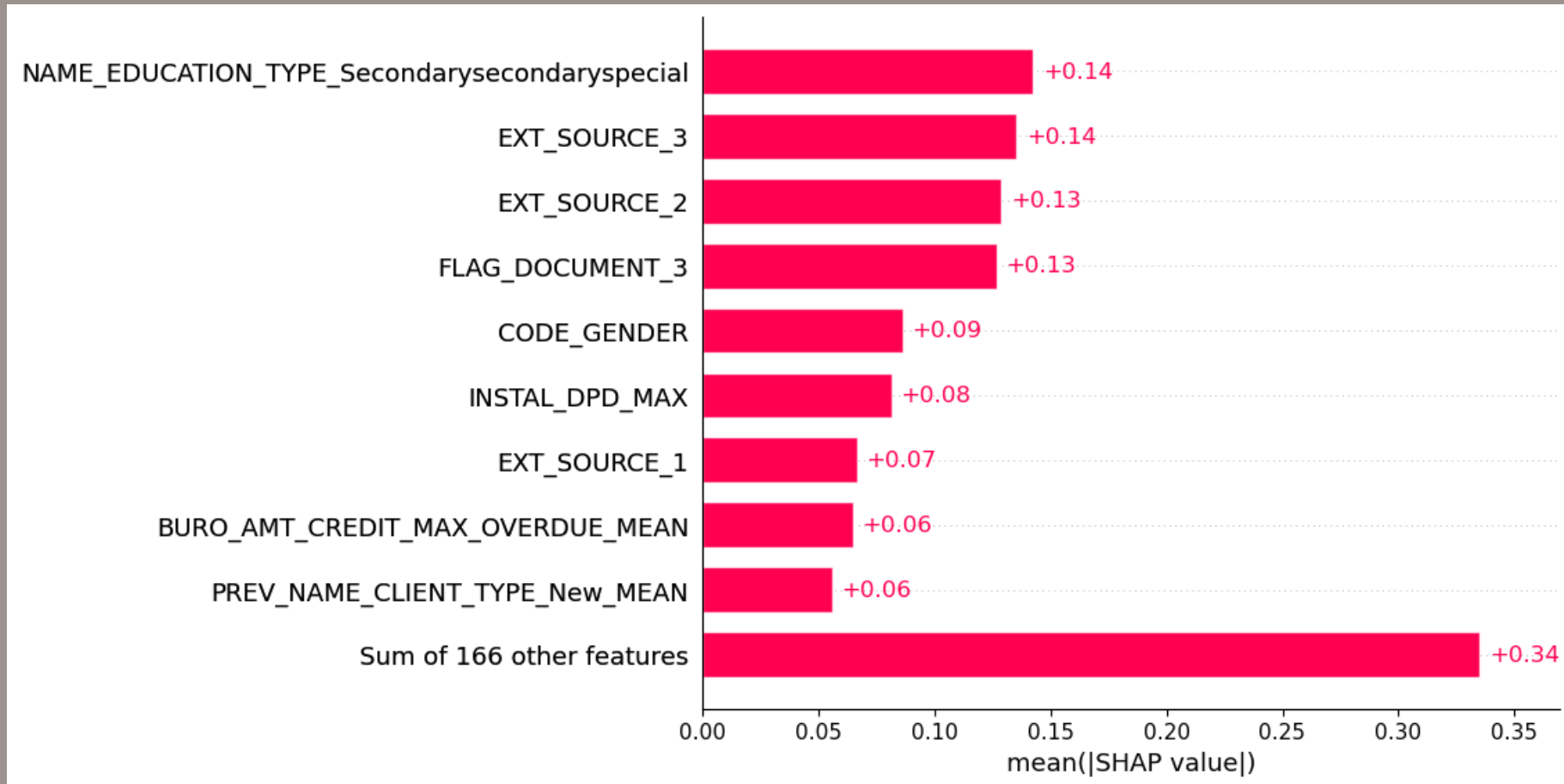
LightGBM Classifieur optimisé

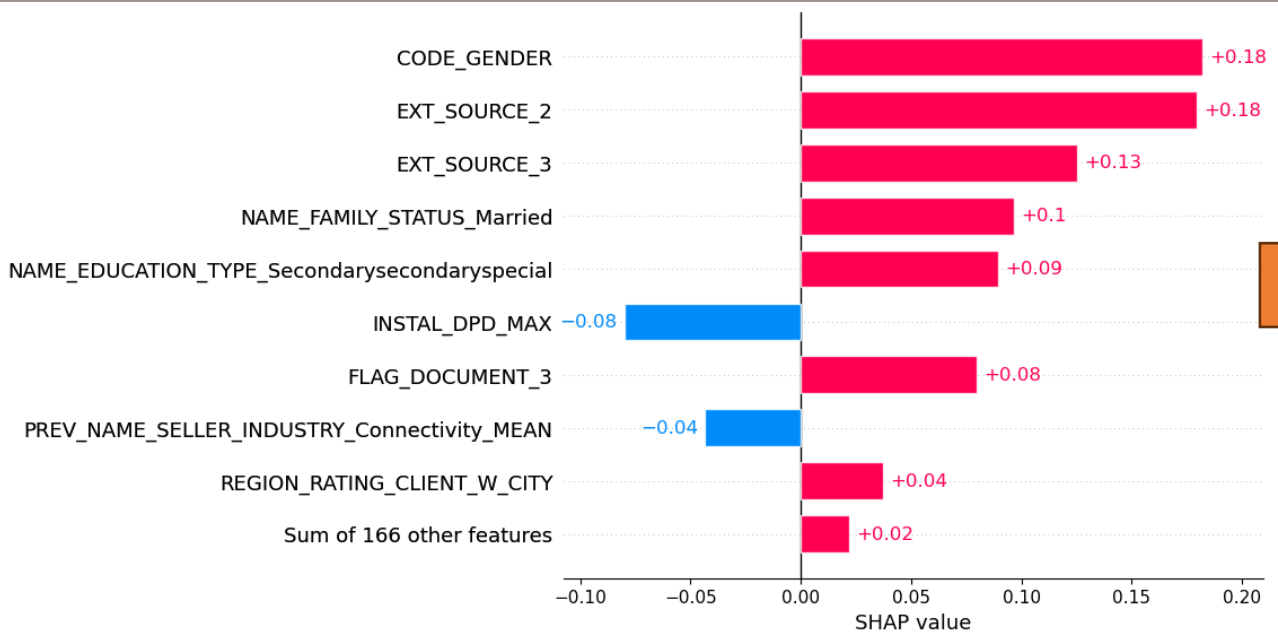


LightGBM avec le seuil



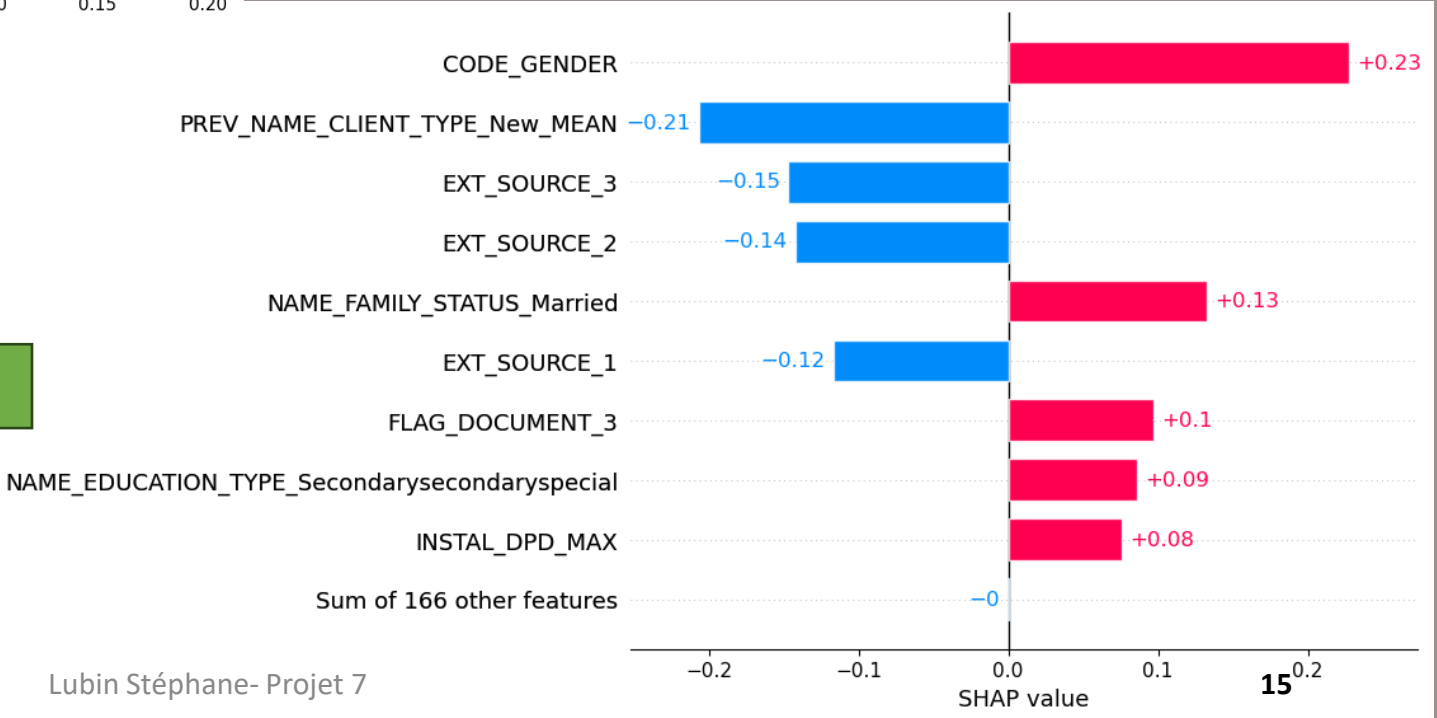
Features Importance



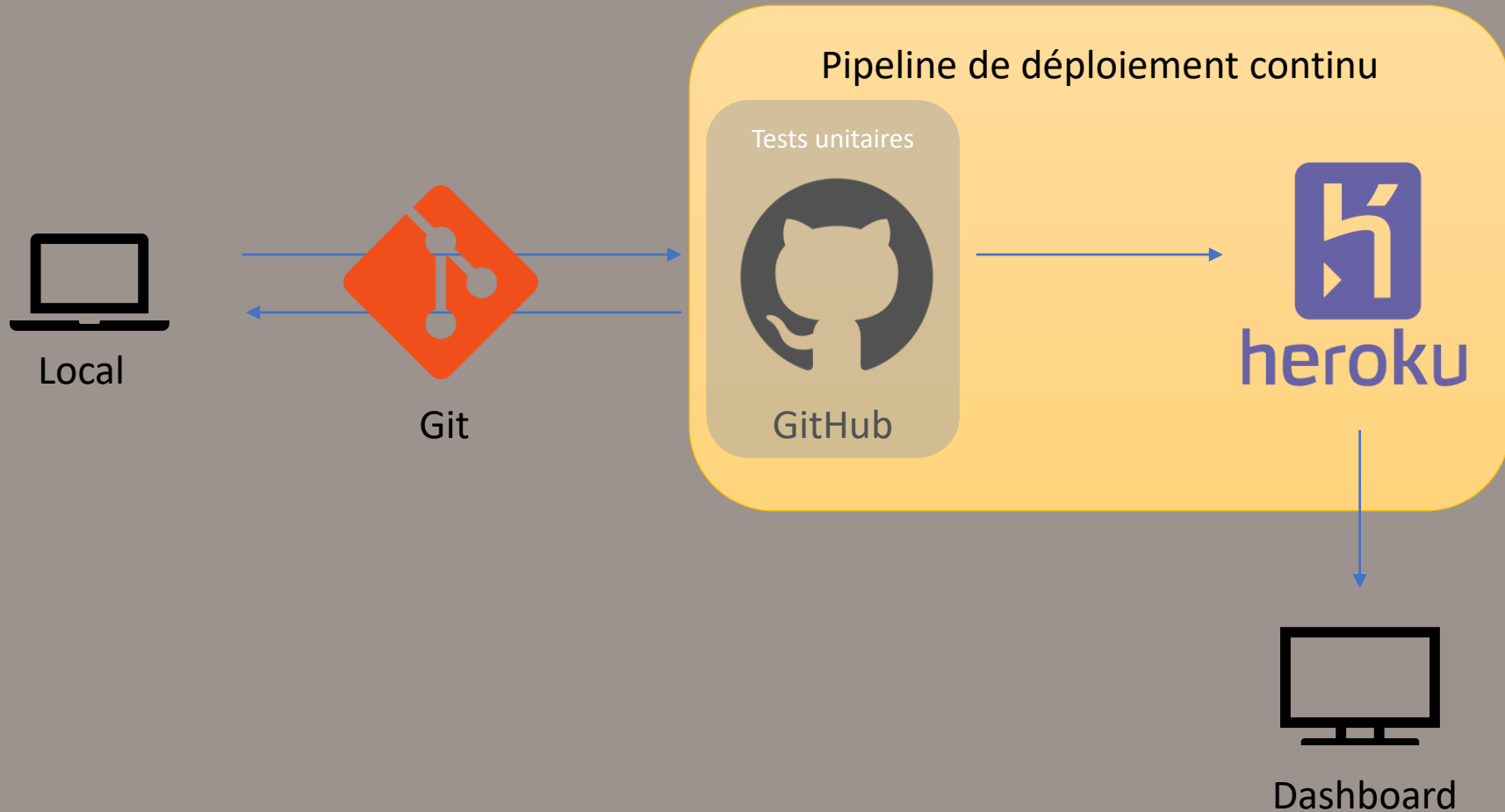


Cas Refusé

Cas accepté



GIT, GITHUB et conception



main 2 branches 0 tags

Go to file Add file <> Code

BlackStylz	Update dashboard.py	ba2b466 4 hours ago	96 commits
.github/workflows	Update python-app.yml		14 hours ago
Data	Ajout fichier dash		yesterday
images	Ajout des images		3 days ago
.gitattributes	'gitattr'	Ajout des images	2 days ago
Procfile	Update Procfile	Ajout des images nécessaires pour le dashboard	yesterday
README.md	Update README.md		14 hours ago
dashboard.py	Update dashboard.py		4 hours ago
requirements.txt	Update requirements.txt		2 days ago
runtime.txt	Create runtime.txt		3 days ago
setup.sh	Create setup.sh		3 days ago

Heroku Git
Use Heroku CLI

GitHub
Connected

Container Registry
Use Heroku CLI

Disable Automatic Deploys

Environments 3

prediction-api
 Active

dashboardsl
 Active

Commits

main

Commits on Jun 14, 2023

Update dashboard.py
BlackStylz committed 5 hours ago

Update dashboard.py
BlackStylz committed 5 hours ago

Actions

New workflow

All workflows

Python application

Management

Caches

Deployments

All workflows

Showing runs from all workflows

5 workflow runs

Update dashboard.py
Python application #9: Commit ba2b466 pushed by BlackStylz

main

Update dashboard.py
Python application #8: Commit f150b9d pushed by BlackStylz

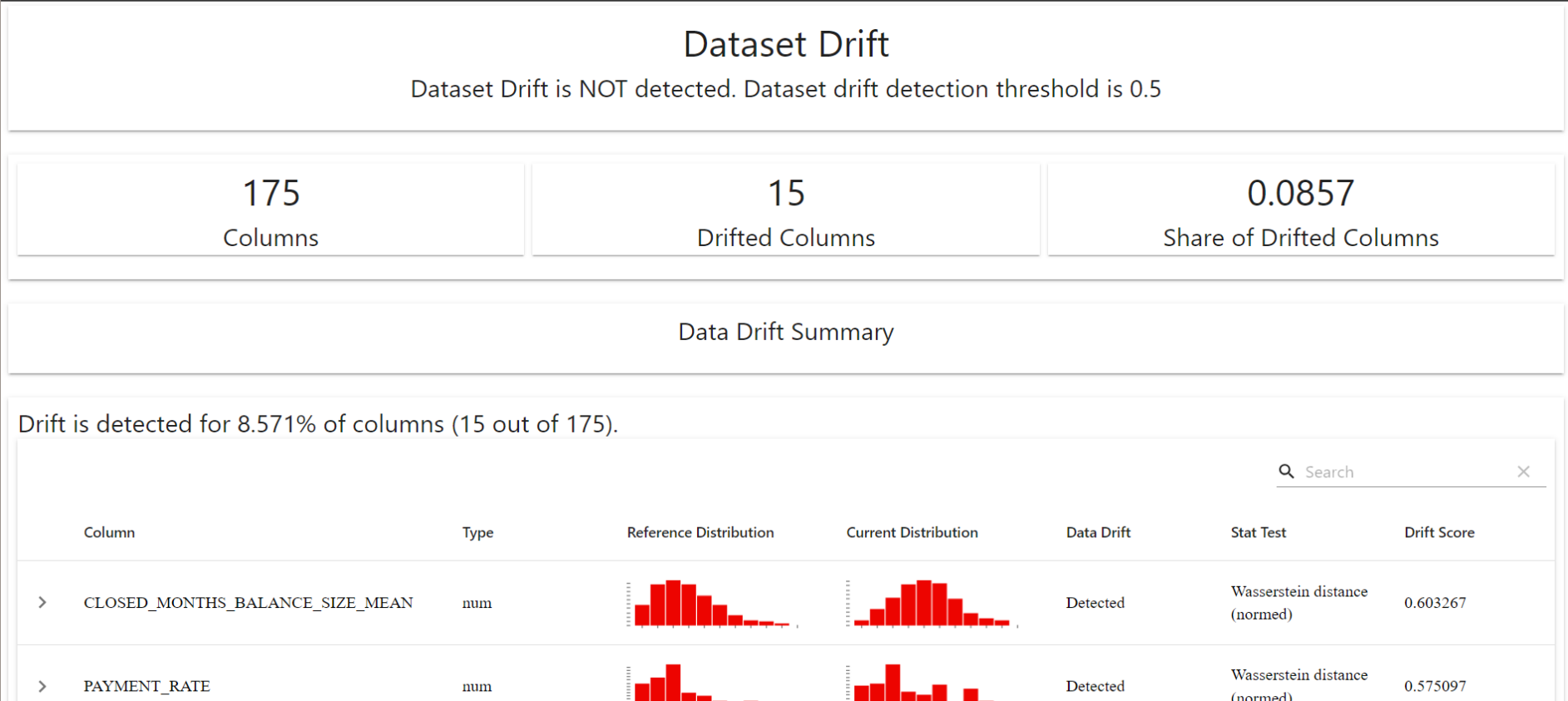
main

Prêt à dépenser

Dashboard déployé



Data Drift via Evidently



Conclusion

- Les limites de mes connaissances
- Les axes d'amélioration
- L'aspect du dashboard
- Les retours des service clientèle

MERCI POUR VOTRE ATTENTION

