

Projet 7

Modèle de scoring crédit

Décision métier, explicabilité & industrialisation

OpenClassrooms – Parcours Data Scientist



Anderson
Carbeti

SOMMAIRE

1. Contexte business et enjeux du scoring crédit
 2. Problématique et objectifs métier
 3. Données utilisées et périmètre d'analyse
 4. Préparation des données et feature engineering
 5. Séparation train / test et cadre méthodologique
 6. Modèles évalués et critères de comparaison
 7. Pourquoi l'AUC n'est pas suffisante
 8. Définition du score métier
 9. Optimisation du seuil de décision
 10. Choix du modèle champion
 11. Traçabilité et MLOps avec MLflow
 12. Interprétabilité globale du modèle (SHAP)
 13. Interprétabilité locale et justification client
 14. Analyse de dérive des données (Evidently)
 15. Pipeline d'inférence et industrialisation
 16. Résultats finaux et performance métier
 17. Limites du projet
 18. Perspectives et améliorations possibles
 19. Synthèse et messages clés
 20. Conclusion
-

CONTEXTE

Faux négatif	Faux positif
<p>Client risqué accepté → Perte financière élevée</p>	<p>Bon client refusé → Perte d'opportunité</p>

Problématique

1

Automatiser une décision de crédit :

- **fiable,**
- **explicable,**
- **et économiquement optimale**

Data → Model → Decision



DONNÉES

➤ Application client

application

➤ Historique crédit

bureau / previous

➤ Comportement de remboursement

POS / installments

➤ Données multi-tables



~300 000 clients

features client

Feature engineering

Ratios financiers

Historique de défaut

Comportement de paiement

Agrégations par client

Séparation des données

Train / validation / test

Pas de fuite d'information

Test jamais utilisé pour le seuil

Train → OOF (seuil) → Test (évaluation finale)

Modèles évalués

Logistic Regression

Random Forest

XGBoost

LightGBM

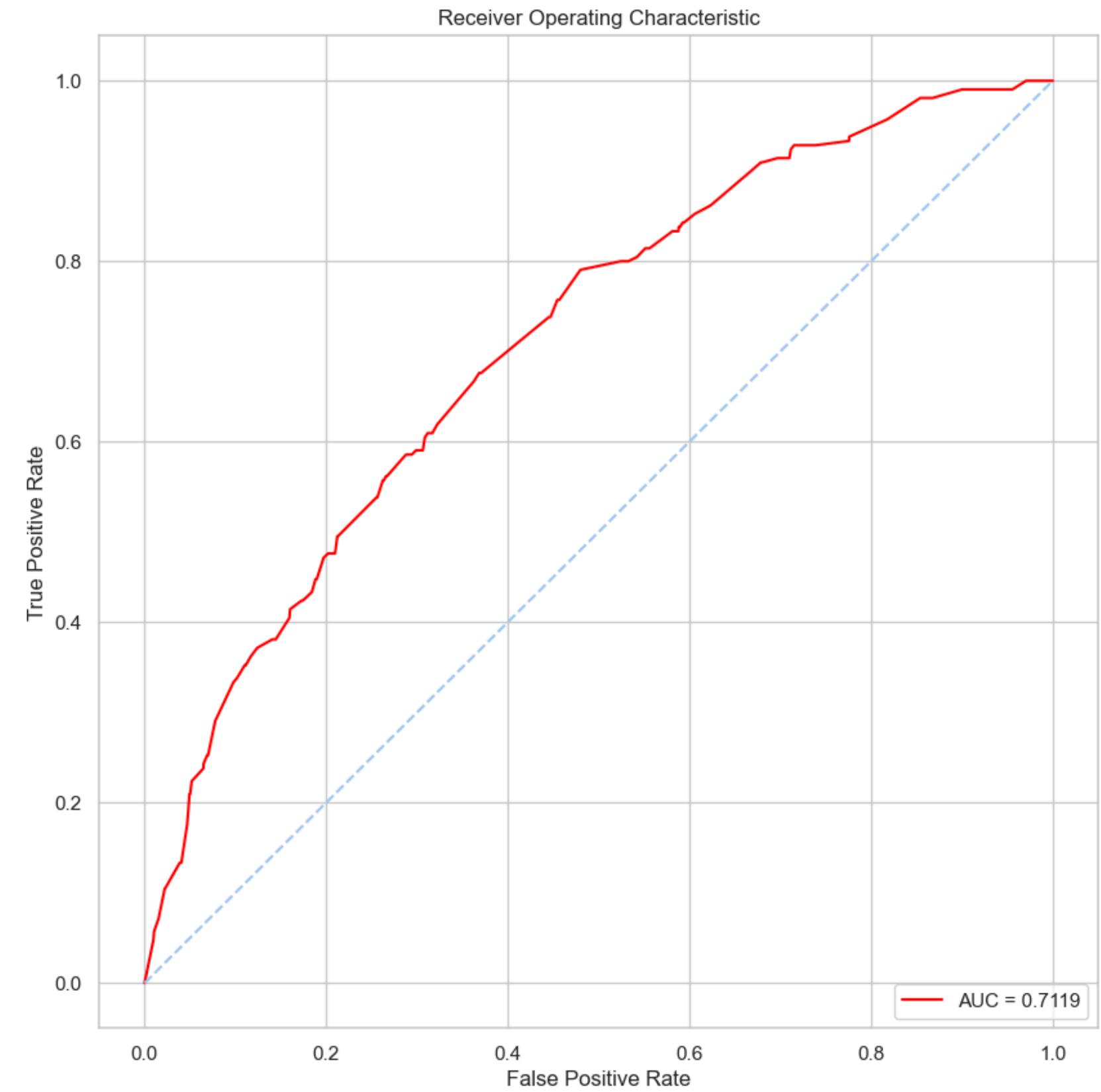
Un seul critère : le score métier

Pourquoi l'AUC n'est pas suffisante

Ignore le seuil

Suppose un coût symétrique

Pas alignée business



Score métier

Faux négatif : coût élevé

Faux positif : coût plus faible

Objectif : minimiser le coût global

$$\text{Coût} = \text{FN} \times 10 + \text{FP} \times 1$$

Objectif : minimiser le coût global

Optimisation du seuil

- Seuil optimisé sur OOF train : 0.402
- Minimisation du coût métier (FN ×10, FP ×1)
- Seuil ≠ 0.5 par construction métier

Modèle champion : XGBoost

Sélection et versionnement du modèle champion

Model name	Status	Created ↓	Logged from	Source run	Registered models	Dataset
model	Ready	1 day ago	Projet7_scoring_jury_ready_	LightGBM-RandomizedSearch	home_credit_default_model	-
model	Ready	1 day ago	Projet7_scoring_jury_ready_	LightGBM	-	-
model	Ready	1 day ago	Projet7_scoring_jury_ready_	XGBoost	-	-
model	Ready	1 day ago	Projet7_scoring_jury_ready_	LogisticRegression	-	-
model	Ready	1 day ago	Projet7_scoring_jury_ready_	RandomForest	-	-

- Tous les modèles sont enregistrés dans MLflow
- Comparaison basée sur le coût métier
- Le modèle final est versionné dans le registry
- Alias unique pour l'inférence : champion

Traçabilité & MLOps

Tracking des runs

Comparaison des modèles

Registry & alias champion

SHAP global

Interprétabilité globale

- **Variables dominantes :**
 - **Revenu**
 - **Montant du crédit**
 - **Historique de défaut**
- **Méthode : SHAP**

Explicabilité Locale

Interprétabilité locale

- **Justification individuelle des décisions**
- **Utilisable pour conformité et relation client**

Data Drift

Surveillance des dérives

- Comparaison train / nouvelles données
- Détection de drift

Pipeline d'inférence

- Pipeline final
 - Chargement du champion
 - Preprocessing
 - Seuil métier
 - Décision finale

JSON → Preprocessing → Model → Seuil → Décision

Résultats

Élément	Valeur
Modèle	XGBoost
AUC test	0.73
Coût métier test	1616
Seuil	0.402
FN	pénalisés ×10

Perspectives

- Next steps
- Monitoring continu
- Recalibrage
- Analyse de biais

Conclusion

Décider mieux plutôt que prédire mieux.



Merci de votre
attention

Avez-vous des
questions ?