

Projet REG08 : Credit Scoring ML

Prédire et expliquer le risque de défaut client

Intaset : Home Credit Default Risk



CHAKOURNE FREDERIC

ATSUVIA Marc Komi Christophe

BA Mohamed

BAH Mamoudou

Objectifs du Projet : Une Approche Systémique



Modélisation

Calculer une probabilité de défaut précise et robuste.



Décision

Transformer un score probabiliste en décision métier (Accord/Refus).



Interprétabilité

Expliquer les causes (Global & Local/SHAP) pour la transparence.



Interface

Dashboard interactif Streamlit pour les chargés de clientèle.



API

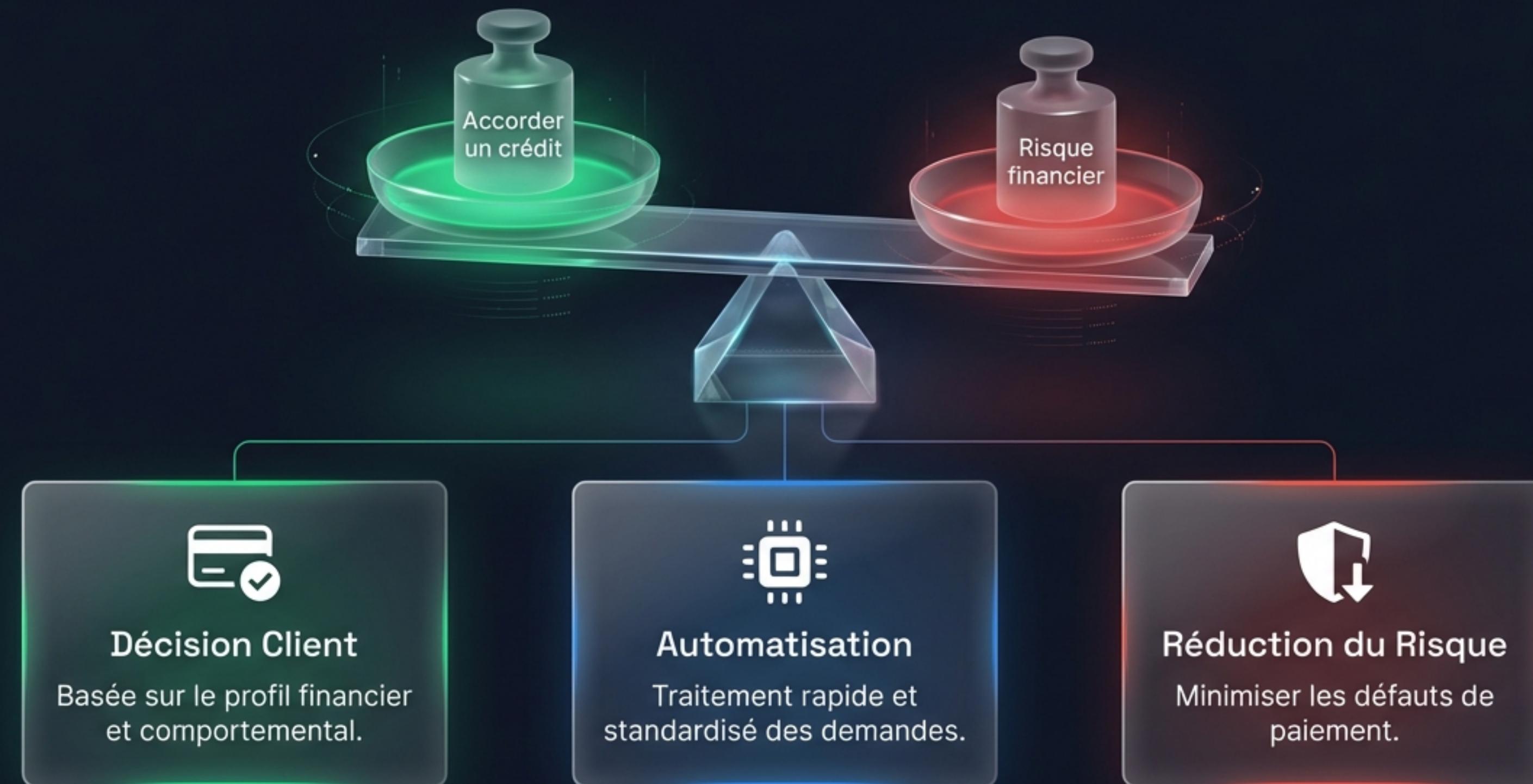
Architecture découpée pour des réponses temps réel.



Tracking

Suivi des expérimentations et reproduction via MLflow.

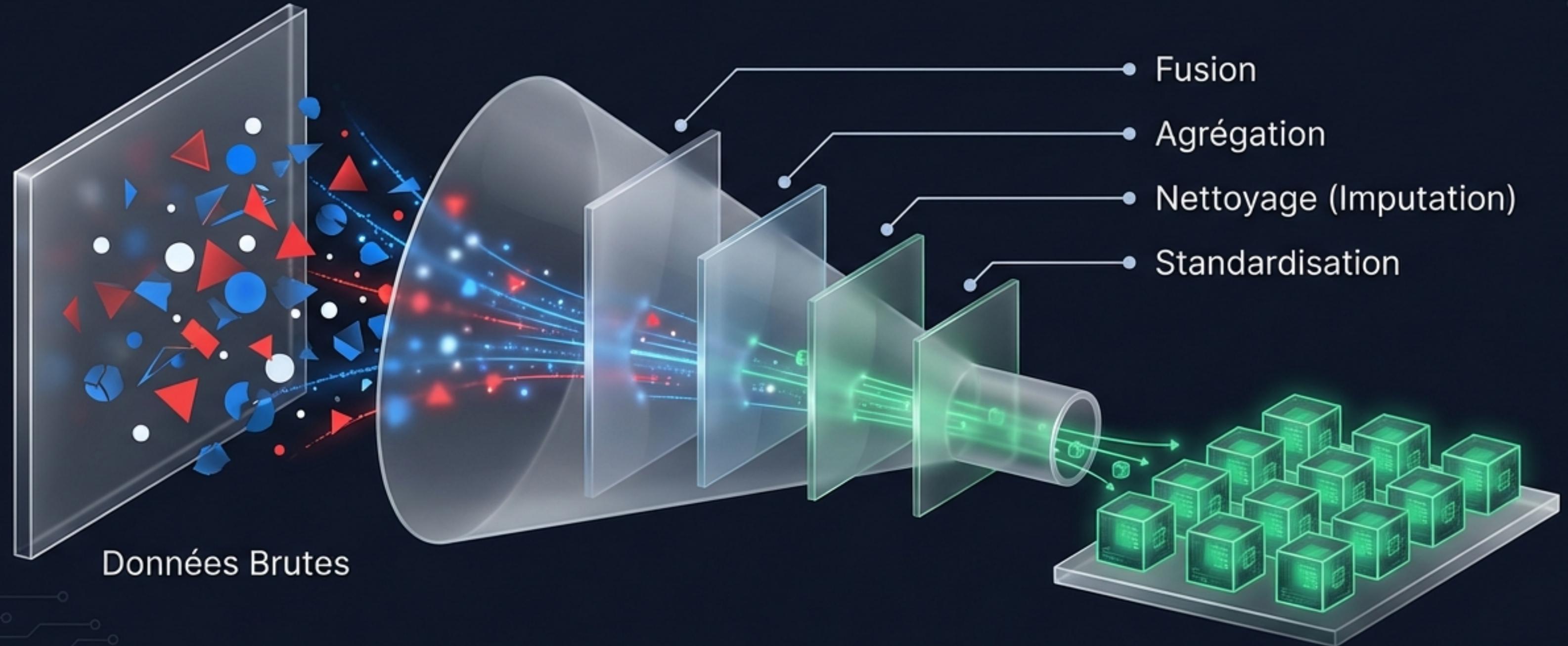
La Problématique Métier : Équilibre Risque vs Opportunité



Écosystème de Données : Une Vue 360°



Préparation et Feature Engineering

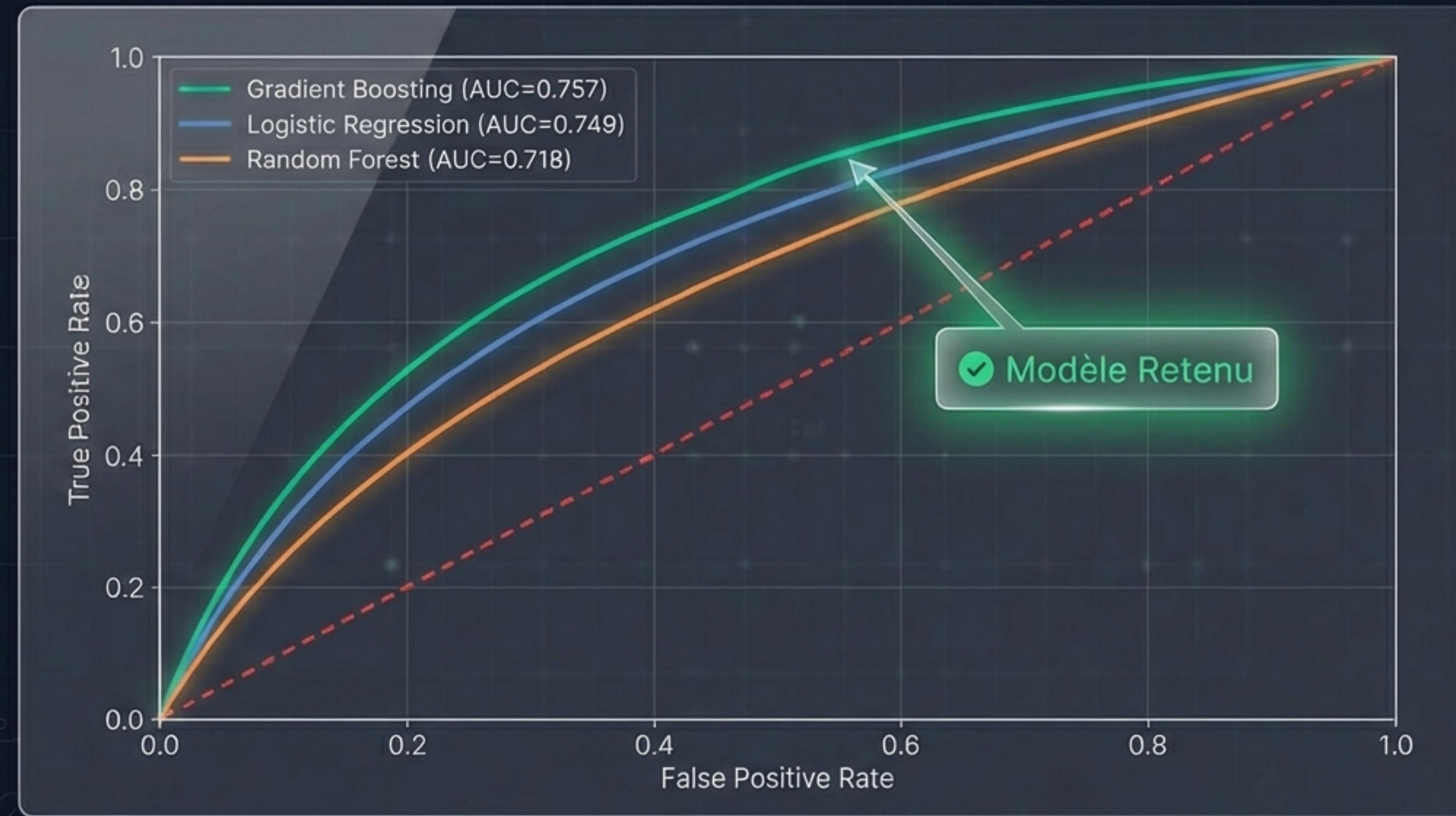


Features du Modèle

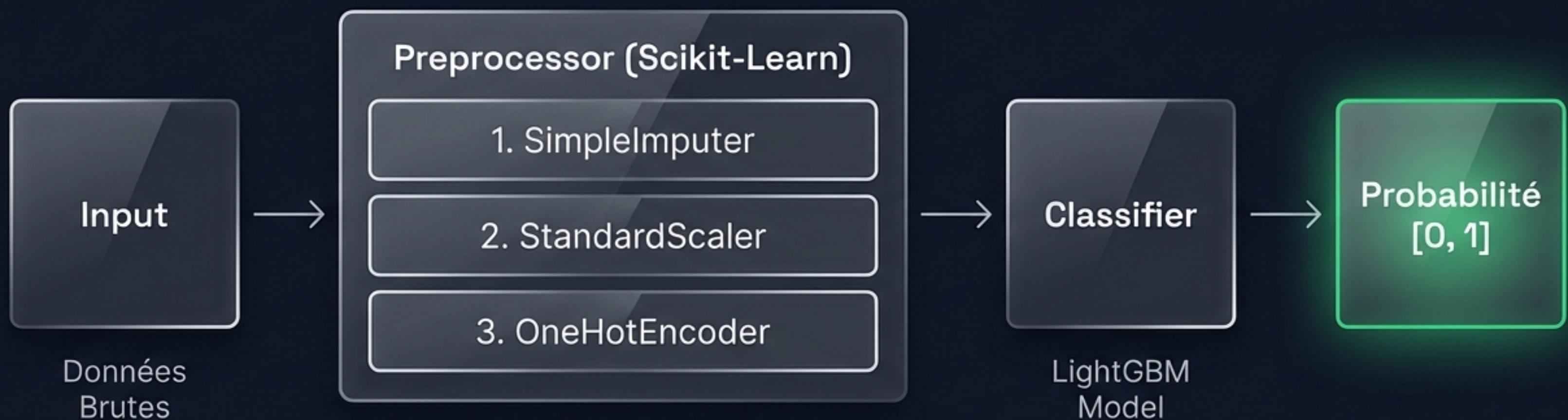
Sélection du Modèle : La Stratégie du Champion



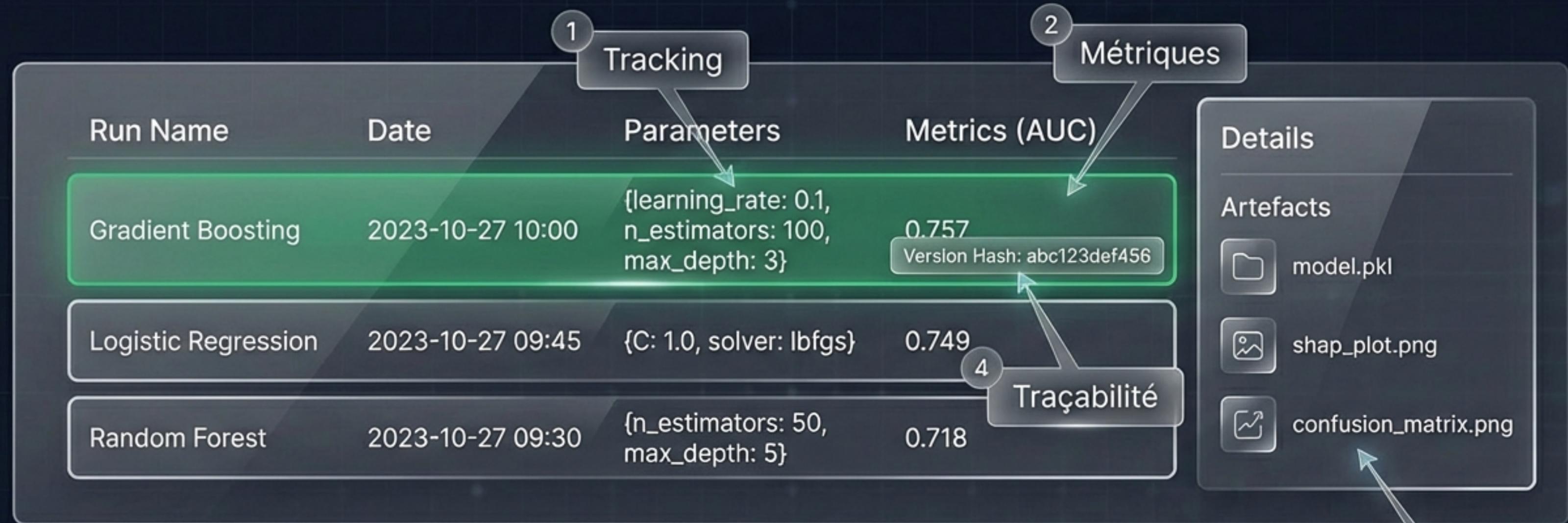
Performance : Analyse des Courbes ROC AUC



Pipeline Machine Learning : Flux Automatisé

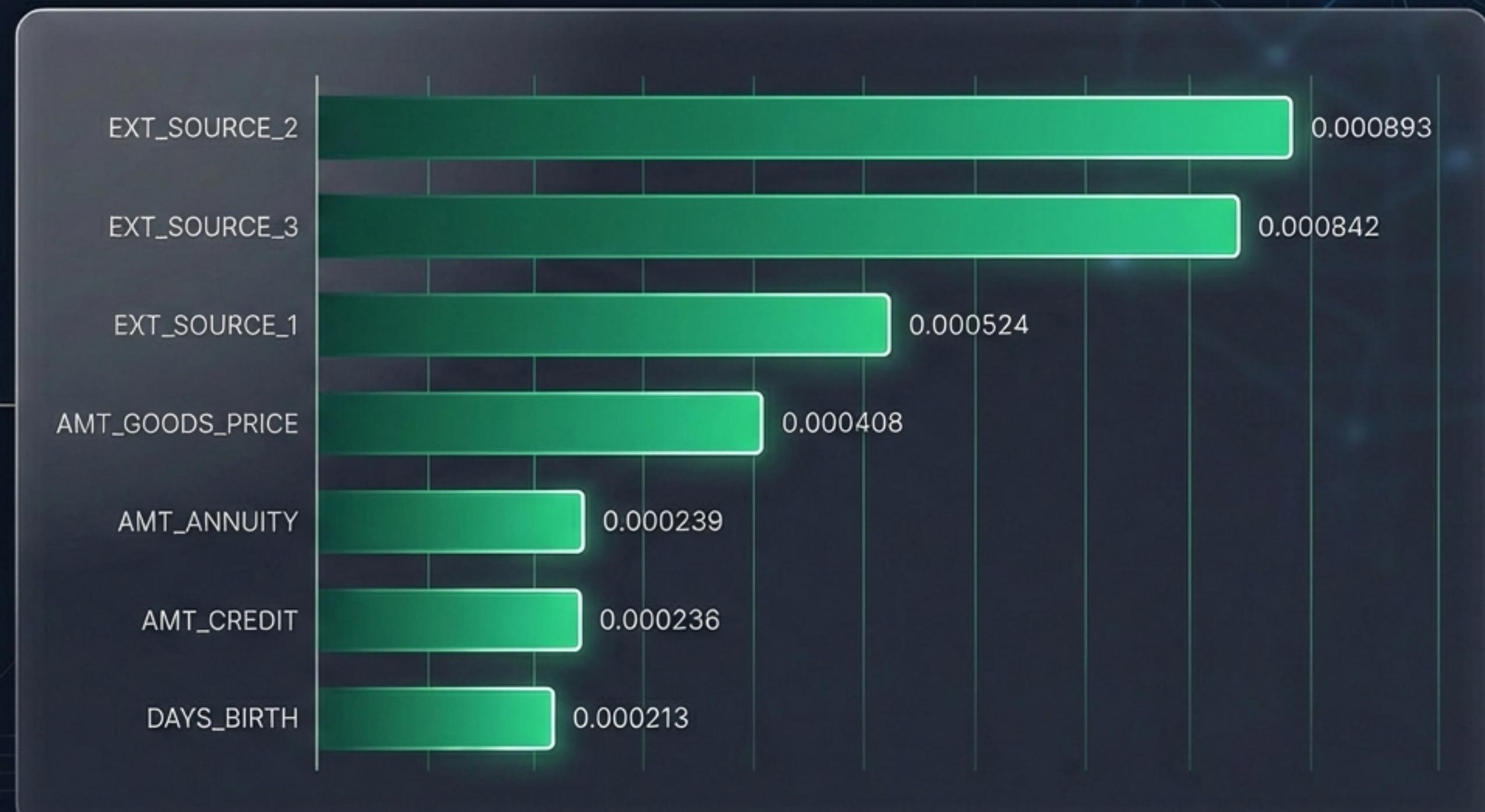


Suivi des Expérimentations avec MLflow



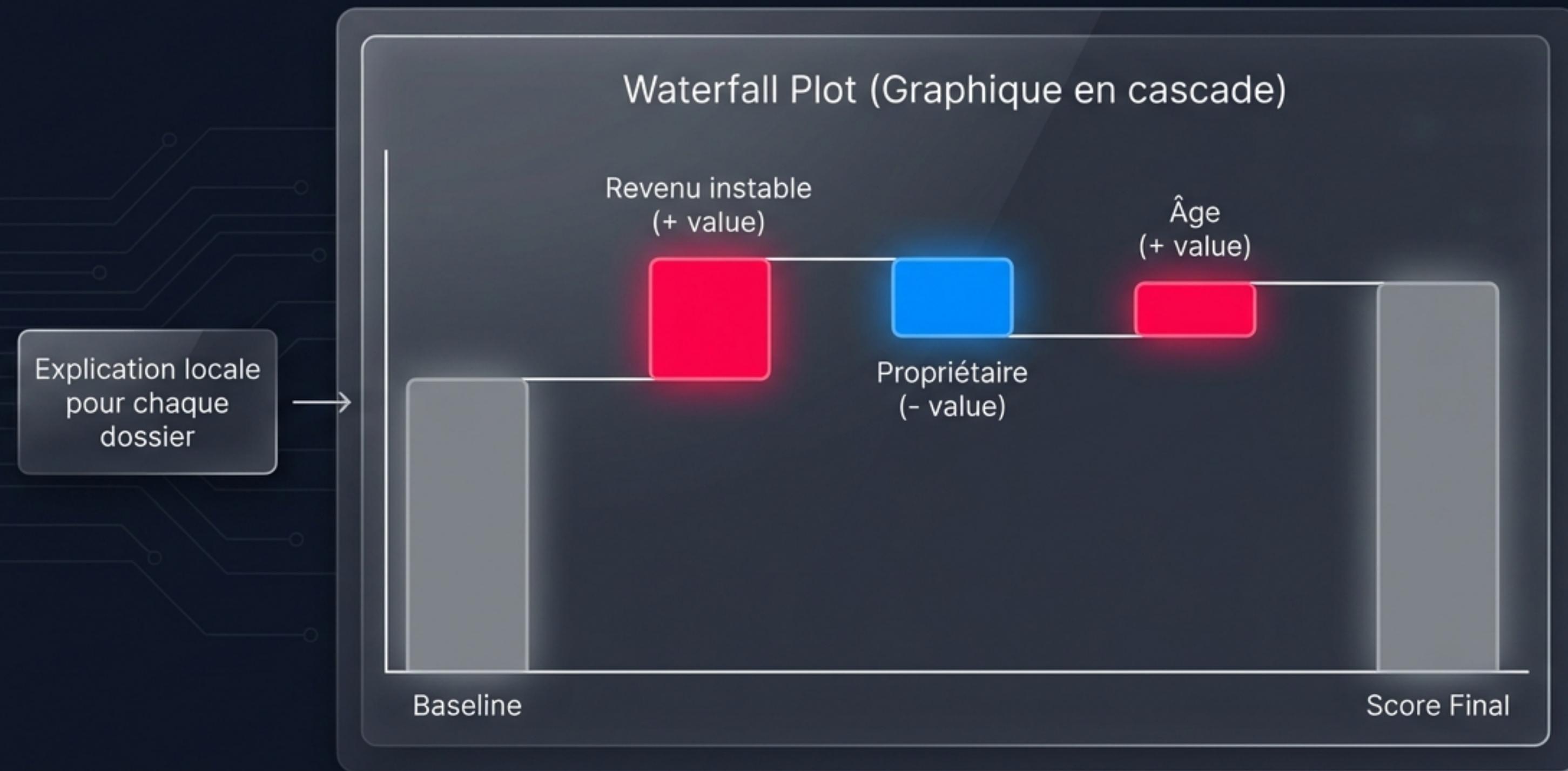
Feature Importance (Globale)

Les scores externes (EXT_SOURCE) et le montant du financement sont les leviers principaux du risque.

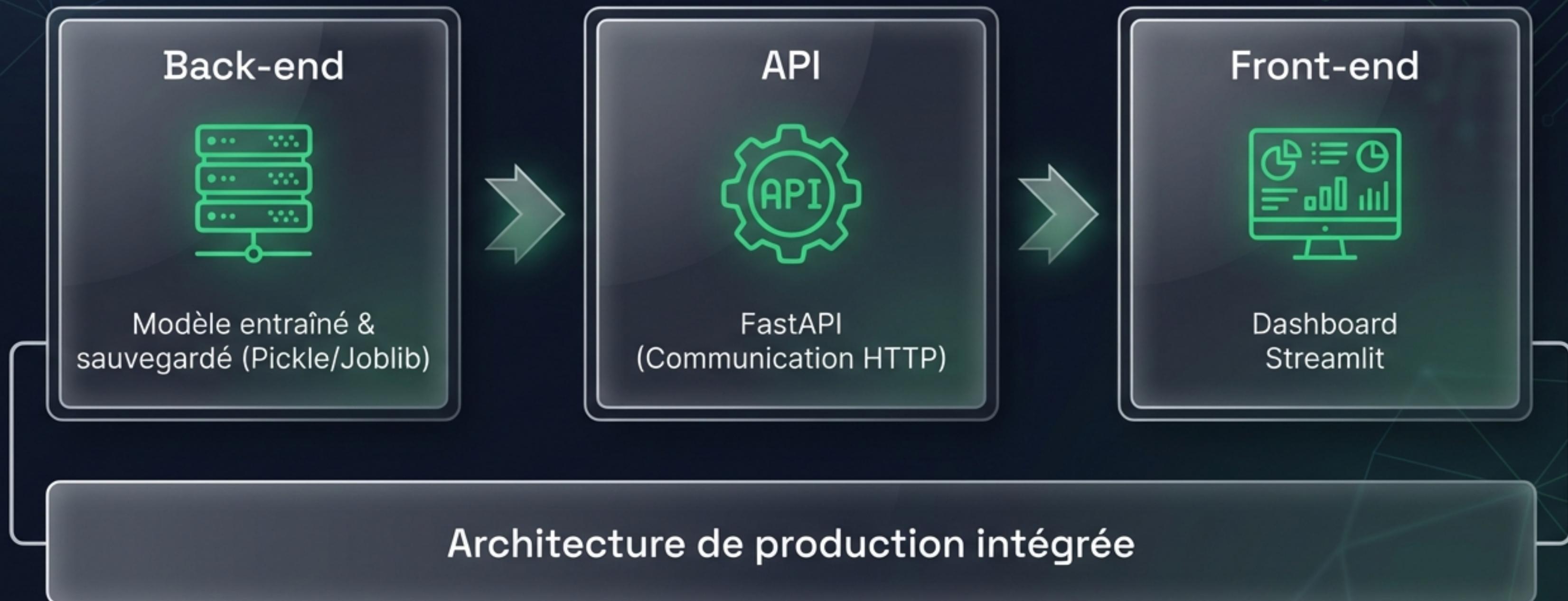


Transparence Décisionnelle : SHAP

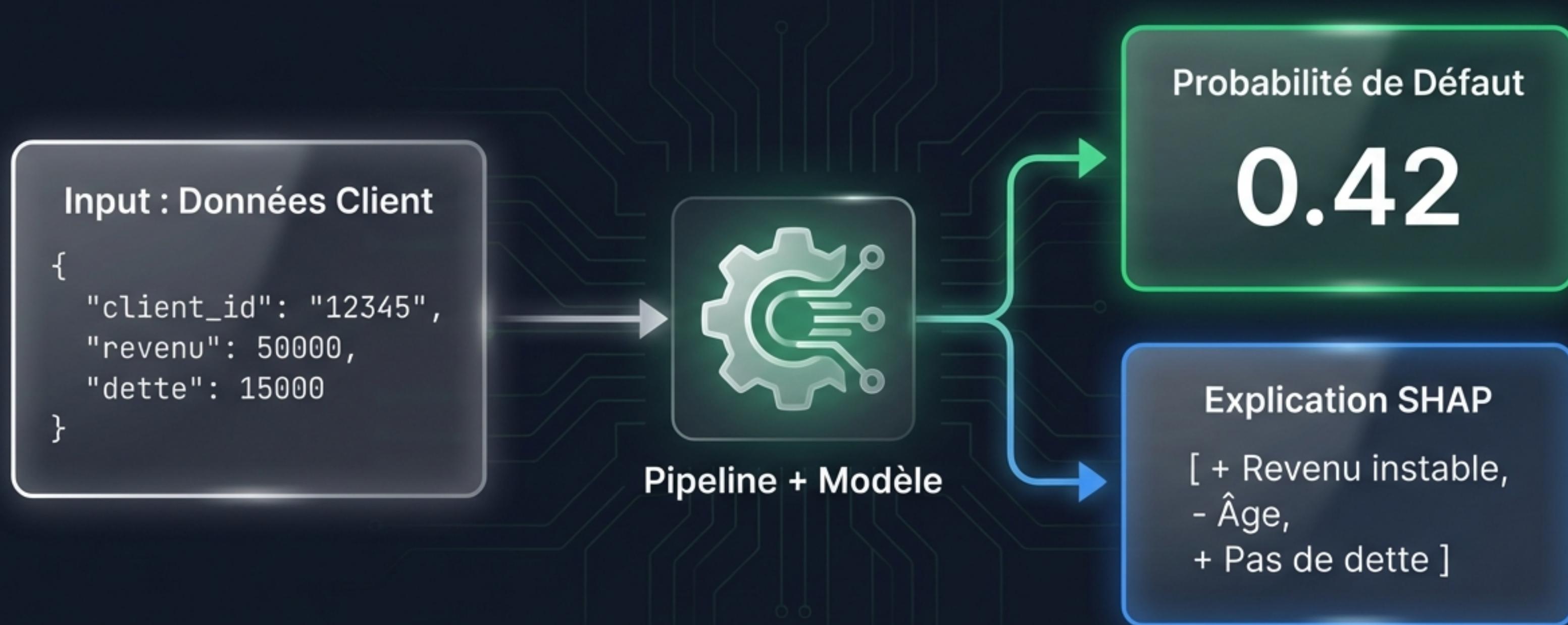
Pourquoi ce client a-t-il ce score ?



Architecture Système & MLOps



Le Moteur : API Credit Scoring



Traitement en temps réel pour une décision instantanée.

Logique de Décision et Interface

Seuil : 0.5

ACCORD

REFUS



Demande de crédit refusée

Risque estimé : 58.1%

Niveau de risque : Élevé

Pourquoi cette décision ?

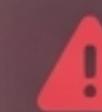
- ⚠️ Fiabilité de paiement externe augmente le niveau de risque
- ⚠️ Historique financier externe augmente le niveau de risque
- ⚠️ Score de solvabilité externe augmente le niveau de risque

Bilan & Conclusion



Points Forts

- Système complet de bout en bout (Data -> Dashboard).
- Performance validée (AUC 0.757).
- Interprétabilité locale (SHAP) pour chaque client.



Limites & Vigilance

- Biais historiques potentiels dans les données.
- Sensibilité aux données manquantes.
- Nécessité de ré-entraînement régulier (Data Drift).