



Projet Final : Outil de Credit Scoring et Démarche MLOps

TRAN Delvin - LAUDRAIN Lukas - MOREL Antoine

Fabien LONTI

Table des matières

Préparation et enrichissement des données	2
Modélisation et évaluation.....	2
Optimisation du seuil métier.....	2
Interprétabilité	3
MLOps et déploiement	4
Conclusion	4

Préparation et enrichissement des données

Les données utilisées proviennent de plusieurs tables du jeu *Home Credit Default Risk*, décrivant à la fois les caractéristiques des clients, leur situation financière et leur historique de crédit. Un premier travail de sélection est effectué afin de limiter la dimension du problème et de ne conserver que les variables pertinentes d'un point de vue métier. Dans la table principale, des variables comme le revenu annuel, le montant du crédit, l'annuité, l'âge, l'ancienneté professionnelle ou les scores issus de sources externes sont conservées. À l'inverse, des identifiants secondaires, des variables très spécifiques ou fortement redondantes sont écartés.

Les tables secondaires (historique de crédits, paiements, cartes de crédit, crédits précédents) sont ensuite agrégées au niveau du client. Seules les variables numériques sont retenues et résumées par des statistiques simples telles que le minimum, le maximum, la moyenne et la somme. Par exemple, l'historique des paiements est synthétisé par des indicateurs globaux de montants payés, de soldes moyens ou de retards, ce qui permet de capturer le comportement financier sans conserver l'ensemble des séries temporelles.

En complément, plusieurs features métiers dérivées sont créées afin de mieux représenter la capacité de remboursement. Parmi elles figurent le ratio entre le montant du crédit et le revenu annuel, le ratio entre l'annuité et le revenu, une approximation de la durée du crédit, ainsi que le rapport entre l'ancienneté professionnelle et l'âge du client. Ces variables permettent de normaliser les montants financiers et d'améliorer la comparabilité entre clients.

Les valeurs manquantes sont analysées sans suppression excessive de colonnes, et une optimisation mémoire est appliquée afin de réduire la taille du dataset final. La séparation entre entraînement et validation est réalisée de manière stratifiée afin de conserver la proportion de clients en défaut.

Modélisation et évaluation

Plusieurs modèles de classification sont testés afin de comparer différentes approches. Une attention particulière est portée au déséquilibre des classes, la majorité des clients ne présentant pas de défaut. Pour limiter le biais vers la classe majoritaire, des pondérations de classes sont utilisées lors de l'entraînement.

Les performances sont évaluées à l'aide de métriques classiques comme l'AUC, utilisées comme indicateur global de discrimination, mais surtout à l'aide d'une métrique métier personnalisée. Cette métrique attribue un coût beaucoup plus élevé aux faux négatifs qu'aux faux positifs, afin de refléter le risque financier réel pour l'entreprise. Une validation croisée est mise en place afin d'obtenir des résultats plus stables et comparables entre modèles.

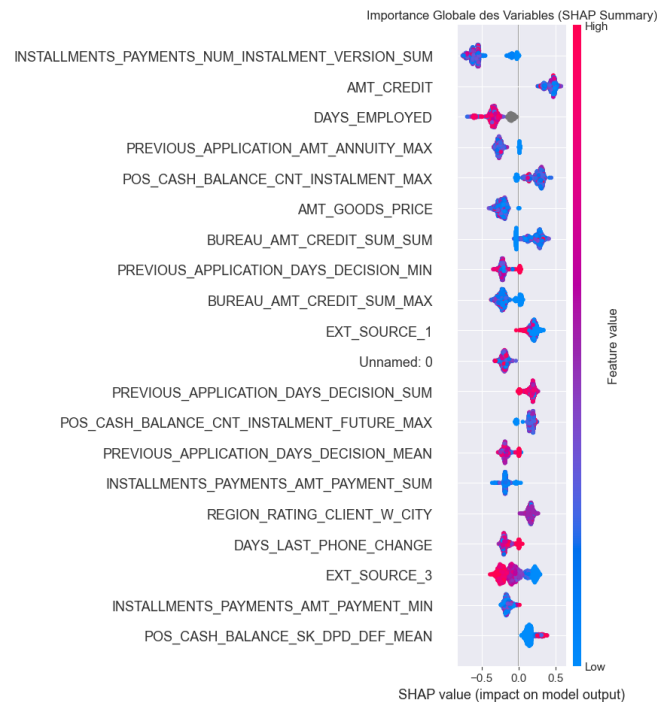
Optimisation du seuil métier

Plutôt que d'utiliser un seuil de décision standard fixé à 0,5, une optimisation explicite du seuil est réalisée. Différents seuils de probabilité sont testés, et pour chacun d'eux, le coût métier total est calculé à partir du nombre de faux négatifs et de faux positifs.

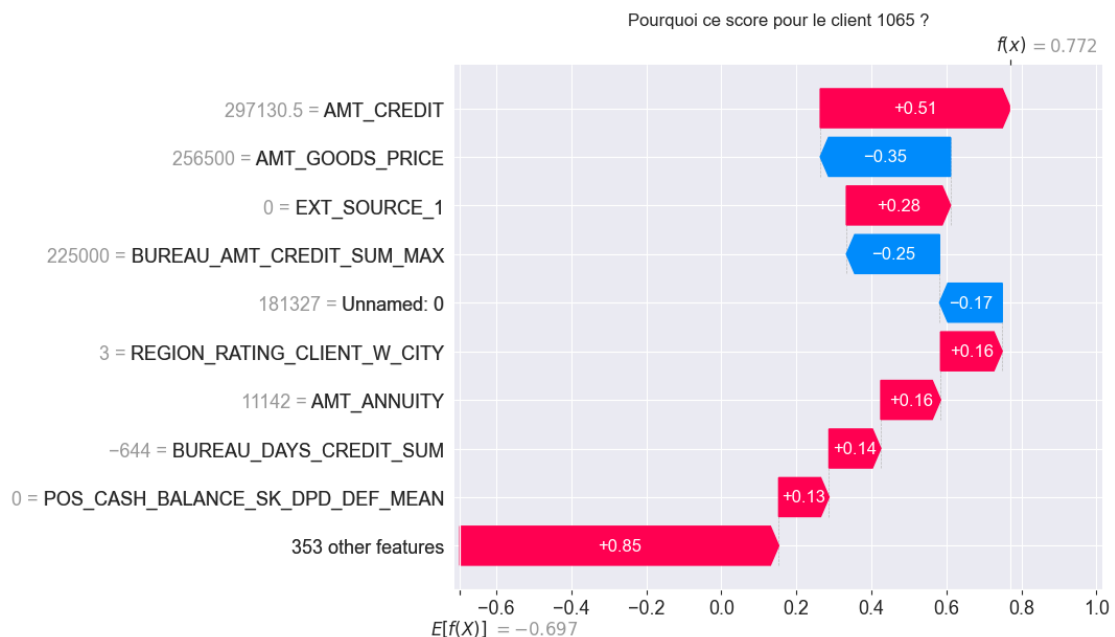
Le seuil minimisant ce coût est retenu comme seuil métier optimal. Il est sauvegardé avec le modèle final et utilisé lors de la phase d'inférence. Cette approche permet par exemple de refuser un crédit pour des probabilités relativement modérées lorsque le coût d'un défaut est jugé trop élevé, tout en acceptant un nombre contrôlé de faux positifs.

Interprétabilité

L'interprétabilité du modèle est abordée à deux niveaux. Une analyse globale permet d'identifier les variables les plus influentes sur l'ensemble du modèle, telles que les scores issus de sources externes, les ratios financiers ou l'ancienneté professionnelle. Cette analyse donne une vision générale des facteurs de risque pris en compte.



Au niveau individuel, une explication locale est fournie pour chaque client à l'aide de SHAP. Pour un dossier donné, il est possible de visualiser les variables qui ont contribué à augmenter ou à diminuer la probabilité de défaut, par exemple un revenu élevé compensant un montant de crédit important, ou au contraire une annuité trop élevée par rapport aux revenus.



MLOps et déploiement

L'ensemble des expérimentations est tracé avec MLflow, qui centralise les paramètres testés, les métriques d'évaluation et les artefacts produits lors de l'entraînement (courbes, figures, modèles). L'interface MLflow permet de comparer facilement les différents essais, d'analyser les compromis entre performance statistique et coût métier, et d'identifier le modèle final retenu. Ce modèle est ensuite enregistré dans le model registry MLflow, ce qui permet de gérer explicitement sa version et de sécuriser son cycle de vie.

Une fois validé, le modèle est déployé via une API FastAPI dédiée, exposant un endpoint de prédiction. Cette API charge le modèle et le seuil métier optimisé au démarrage, puis renvoie pour chaque client la probabilité de défaut ainsi qu'une décision d'octroi ou de refus basée sur ce seuil. Cette séparation claire entre entraînement et inférence garantit la cohérence entre les résultats observés en phase d'expérimentation et ceux obtenus en production.

Un dashboard Streamlit vient compléter le dispositif. Il permet d'interroger l'API sur des clients issus du jeu de test, de visualiser la probabilité de défaut sous forme graphique, la décision associée, et une justification locale basée sur SHAP. Cet outil facilite l'utilisation du modèle par des utilisateurs non techniques et renforce la transparence du système.

L'ensemble de la solution est « encapsulé » avec Docker, ce qui assure la reproductibilité de l'environnement, simplifie le déploiement et constitue une base solide pour un passage en pré-production ou en production.

Conclusion

Ce projet aboutit à un outil de credit scoring réellement exploitable, qui couvre l'ensemble du cycle de vie du modèle, depuis la préparation des données jusqu'à la mise en production. Le choix d'une métrique métier spécifique et l'optimisation du seuil de décision permettent de prendre des décisions cohérentes avec les enjeux économiques réels, en assumant explicitement le coût plus élevé des faux négatifs par rapport aux faux positifs.

La démarche MLOps mise en place apporte une vraie valeur opérationnelle. Le suivi des expérimentations, la comparaison des modèles et la gestion des versions via MLflow rendent le processus plus transparent et reproductible. La séparation claire entre l'entraînement du modèle et son utilisation en inférence via une API dédiée limite les écarts entre les performances mesurées en développement et celles observées en pratique.

L'ajout de moyens d'explicabilité renforce la confiance dans le système. Les équipes métiers peuvent comprendre les facteurs qui influencent le score, aussi bien de manière globale que pour un client donné, ce qui facilite l'appropriation de l'outil et son utilisation au quotidien.

Enfin, l'architecture retenue est suffisamment flexible pour évoluer dans le temps. Elle permet d'envisager sans rupture des améliorations comme le suivi de la dérive des données, la mise à jour régulière du modèle ou l'intégration de nouvelles sources d'information, tout en restant compatible avec des contraintes de déploiement réelles.