**Conseil à l'attribution de crédits**

**Méthode d'intelligence artificielle de l'application**

**"Home Credit AI advice"**

**Sommaire**

Conception et entraînement du modèle 2p

Données nécessaires

Modèles testés

Evaluation des modèles

Un modèle adapté aux assurances : fonction coût métier 1p

Interpréter les résultats du modèle 1p

Interprétation globale

Interprétation locale

Limites 1p

Perspectives et axes d'amélioration

Annexe : données nécessaires à l'algorithme

**Préambule**

L'application "Home Credit AI advice" a pour but d'apporter une aide à la décision dans l'attribution des crédits aux clients. L'utilisation de l'intelligence artificielle permet de se baser sur les résultats (remboursement ou défaut) de plus de 300 000 clients. En utilisant cette grande quantité d'informations, un algorithme a appris à évaluer la probabilité de remboursement d'un client à partir de ses informations personnelles, du type de crédit demandé et de son historique bancaire.

Les principales fonctions de l'application sont :

- Un chargement direct des données clients au format excel

- Un résumé des principales informations sur votre client

- La présentation des résultats de l'algorithme d'intelligence artificielle

- Des informations complémentaires pour comprendre le résultat de l'algorithme et situer votre client au sein de l'ensemble des clients de Home Credit

Les prédictions d'un algorithme d'IA comportent toujours une part d'incertitude.

Votre expertise métier reste indispensable pour correctement évaluer la capacité de remboursement de vos clients.

1. Conception et entraînement du modèle

1.1. Problématique

La problématique posée est de définir si le client va rembourser son prêt ou faire défaut. Il y a donc 2 réponses possibles, on appelle ce type de problème une classification binaire. Pour prendre en compte l'incertitude sur le résultat, on préfèrera exprimer le résultat en termes de probabilités :

- probabilité que le client appartienne au groupe de clients qui remboursent

- et probabilité qu'il appartienne au groupe de clients qui font défaut.

Dans l'application, on ne garde que cette seconde probabilité multipliée par 100. Le chiffre donné représente ainsi le pourcentage de risque que le client fasse défaut.

1.2. Données nécessaires

La base de données de Home Credit comporte initialement plus de 300 000 clients et leur données personnelles récoltées lors de leur application, leur historique de remboursement, l'historique de leur précédents crédits à Home Credit et dans d'autres établissement bancaires. Une sélection a été effectuée afin de ne conserver que les informations disponibles pour au moins 80% des clients (donc faciles à obtenir). Les clients ayant fourni trop peu d'informations ont ensuite été écartés. La liste des informations (variables) utilisées par le modèle est disponible en annexe.

1.3. Modèles testés

Plusieurs types de modèles ont été testés : des modèles linéaires (régression logistique, SVC linéaire) et ensemblistes (Random Forest, LightGBM). Les réseaux de neurones ont été écartés. Ce type de modèles calculent de très nombreux paramètres (architecture du réseau de neurone, poids affectés aux informations qui transitent de neurones en neurones, etc…) qui sont ensuite combinés de façon complexe. Bien que performants, leur fonctionnement ne permet pas d'avoir une explicabilité suffisante pour répondre aux exigences de transparence de Home Credit et de ses clients.

1.4. Evaluation des modèles

Pour sélectionner le meilleur type de modèle, plusieurs mesures ont été effectuées sur des données d'entraînement, de validation et de test : accuracy, balanced accuracy, précision, rappel, F-mesure, spécificité, ROC AUC. Ce sont les mesures sur les données de test (qui n'ont pas servies à l'entraînement ni à l'optimisation des modèles et permettent une évaluation la plus indépendante possible) qui sont utilisée pour déterminer le meilleur algorithme. La mesure privilégiée dans un premier temps est l'aire sous la courbe ROC (ROC AUC). Celles-ci est construite à parti du rappel et de la spécificité et permet une meilleure évaluation des modèles qu'avec une mesure seule. L'aire sous la courbe sera comprise entre 0,5 et 1. Une valeur de 0,5 correspond à un modèle aléatoire, une valeur de 1 à un classifieur parfait.