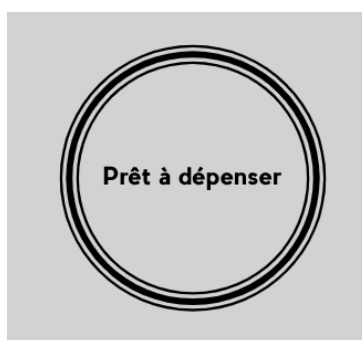


Note méthodologique



Sommaire

1	Introduction	1
2	Classification	1
2.1	Jeu de données	1
2.2	”Feature engeneering”	1
2.3	Déséquilibre des valeurs cible	1
2.4	Modèles	1
3	Évaluation, optimisation et coût métier	1
4	Interprétabilité	1
4.1	Globale	1
4.2	Locale	1
5	Limites et améliorations	1
6	Conclusion	1

1 Introduction

L'entreprise Prêt à dépenser souhaite utiliser un outil de "scoring" afin de calculer la probabilité qu'un client fasse ou non défaut lors du remboursement de son crédit. Pour cela nous devons entrainer un modèle de classification sur des données variées (comportementales, autres institutions financières, etc).

2 Classification

2.1 Jeu de données

Afin de mieux comprendre les données nous avons procédé à une analyse exploratoire des données sur le jeu `application_train.csv`. Pour l'entraînement du modèle nous utiliserons uniquement ces données car l'utilisation des fichiers supplémentaires demande beaucoup de ressources tant en temps d'analyse et d'exploration des données qu'en capacités de calculs. Nous utiliserons le fichier `application_test.csv` pour le dashboard.

2.2 "Feature engeneering"

2.3 Déséquilibre des valeurs cible

2.4 Modèles

3 Évaluation, optimisation et coût métier

4 Interprétabilité

4.1 Globale

4.2 Locale

5 Limites et améliorations

6 Conclusion