Entraînement d’un modèle prédisant la capacité de remboursement d’un client.

Contexte

Des données bancaires et annexes ont été collectées sur de nombreux clients et sont disponibles ici.

L’objectif est de les utiliser pour établir un modèle supervisé permettant de prédire la capacité de remboursement d’un nouveau client venant faire une demande de crédit chez « prêt à dépenser ».

Dans cette note, je fais synthèse :

- de l’exploration des données et du problème inhérent de déséquilibre des classes ;
- de mes réflexion sur la construction d’une fonction coût métier adaptée au problème, et des choix des métriques plus classiques que j’ai décidé de monitorer ;
- du développement d’outils réalisés pour trouver les meilleurs hyper-paramètres de différents modèles (au sens de pipeline comprenant éventuellement des étapes de pré-traitements) afin d’optimiser ladite fonction coût métier ;
- de l’analyse des résultats et de la sélection du/des meilleurs modèles ;
- de l’interprétabilité du modèle retenu ;
- de l’analyse du data-drift ;
- et enfin, je pointerai les limites et améliorations possibles concernant le modèle retenu.

Exploration des données.