

Implémentez un modèle de scoring

MISSION

Vous êtes Data Scientist au sein d'une société financière, nommée "**Prêt à dépenser**", qui propose des crédits à la consommation pour des personnes ayant peu ou pas du tout d'historique de prêt.



L'entreprise souhaite **mettre en œuvre un outil de "scoring crédit"** pour **calculer la probabilité** qu'un client rembourse son crédit, puis classifie la demande en crédit accordé ou refusé. Elle souhaite donc développer un **algorithme de classification** en s'appuyant sur des sources de données variées (données comportementales, données provenant d'autres institutions financières, etc.).

De plus, les chargés de relation client ont fait remonter le fait que les clients sont de plus en plus demandeurs de **transparence** vis-à-vis des décisions d'octroi de crédit. Cette demande de transparence des clients va tout à fait dans le sens des valeurs que l'entreprise veut incarner.

Prêt à dépenser décide donc de **développer un dashboard interactif** pour que les chargés de relation client puissent à la fois expliquer de façon la plus transparente possible les décisions d'octroi de crédit, mais également permettre à leurs clients de disposer de leurs informations personnelles et de les explorer facilement.

Les données

Voici [les données](#) dont vous aurez besoin pour réaliser le dashboard. Pour plus de simplicité, vous pouvez les télécharger à [cette adresse](#). Vous aurez sûrement besoin de joindre les différentes tables entre elles.

Votre mission

1. Construire un modèle de scoring qui donnera une prédiction sur la probabilité de faillite d'un client de façon automatique.
2. Construire un dashboard interactif à destination des gestionnaires de la relation client permettant d'interpréter les prédictions faites par le modèle, et d'améliorer la connaissance client des chargés de relation client.

Michaël, votre manager, vous incite à sélectionner un kernel Kaggle pour vous faciliter la préparation des données nécessaires à l'élaboration du modèle de scoring. Vous analyserez ce kernel et l'adapterez pour vous assurer qu'il répond aux besoins de votre mission.

Vous pourrez ainsi vous focaliser sur l'élaboration du modèle, son optimisation et sa compréhension.

Spécifications du dashboard

Michaël vous a fourni des spécifications pour le dashboard interactif. Celui-ci devra contenir au minimum les fonctionnalités suivantes :

- Permettre de visualiser le score et l'interprétation de ce score pour chaque client de façon intelligible pour une personne non experte en data science.
- Permettre de visualiser des informations descriptives relatives à un client (via un système de filtre).
- Permettre de comparer les informations descriptives relatives à un client à l'ensemble des clients ou à un groupe de clients similaires.

Livrables

- Le **dashboard** interactif répondant aux spécifications ci-dessus et l'API de prédiction du score, déployées chacune sur le cloud.
- Un **dossier** sur un outil de versioning de code contenant :
 - Le code de la modélisation (du prétraitement à la prédiction)
 - Le code générant le dashboard
 - Le code permettant de déployer le modèle sous forme d'API
- Une **note méthodologique** décrivant :

- La méthodologie d'entraînement du modèle (2 pages maximum)
- La fonction coût métier, l'algorithme d'optimisation et la métrique d'évaluation (1 page maximum)
- L'interprétabilité globale et locale du modèle (1 page maximum)
- Les limites et les améliorations possibles (1 page maximum)
- Un **support de présentation** pour la soutenance, détaillant le travail réalisé.

Pour faciliter votre passage au jury, déposez sur la plateforme, dans un dossier nommé "*P7_nom_prenom*", tous les livrables du projet. Chaque livrable doit être nommé avec le numéro du projet et selon l'ordre dans lequel il apparaît, par exemple "*P7_01_dashboard*", "*P7_02_dossier*", et ainsi de suite.

Soutenance

Pendant la soutenance, l'évaluateur jouera le rôle de Michaël, à qui vous présentez votre travail.

- **Présentation (20 minutes)**
 - Rappel de la problématique et présentation du jeu de données (5 minutes)
 - Explication de l'approche de modélisation (10 minutes)
 - Présentation du dashboard (5 minutes)
- **Discussion (5 minutes)**
 - L'évaluateur, jouant le rôle de Michaël, vous challengera sur vos choix.
- **Débriefing (5 minutes)**
 - À la fin de la soutenance, l'évaluateur arrêtera de jouer le rôle de Michaël pour vous permettre de débriefer ensemble.

Compétences évaluées

- Présenter son travail de modélisation à l'oral
- Réaliser un dashboard pour présenter son travail de modélisation
- Rédiger une note méthodologique afin de communiquer sa démarche de modélisation
- Utiliser un logiciel de version de code pour assurer l'intégration du modèle
- Déployer un modèle via une API dans le Web