



DATA

Implémentez un modèle de scoring

🕒 80 heures

Accéder au contenu

Qu'allez-vous apprendre dans ce projet ?

Ce projet va reprendre un contexte qui vous est désormais familier : un problème de modélisation supervisée via une classification binaire. Toutefois, vous allez vous concentrer sur l'apprentissage d'une nouvelle compétence centrale : **la gestion du cycle de vie d'un modèle de ML**.

En effet, après une très brève analyse exploratoire, vous allez entamer vos travaux de modélisation en utilisant un outil de versionning des modèles comme MLFlow, un outil qui a pris une place centrale dans la communauté Data Science ces dernières années.

Après avoir sécurisé votre modèle le plus performant et avoir rendu transparent l'impact des features sur votre modèle, vous allez le mettre en production dans le cloud. Pour ce faire, vous allez consolider davantage vos acquis du projet précédent en ce qui concerne l'utilisation des APIs. Vous allez concevoir votre propre API d'ailleurs, en utilisant des librairies comme FastAPI ou Flask. Enfin, vous allez gérer le cycle de vie du modèle via du CI/CD et l'étude du drift entre autres.

Pourquoi ces compétences sont-elles importantes pour votre carrière ?

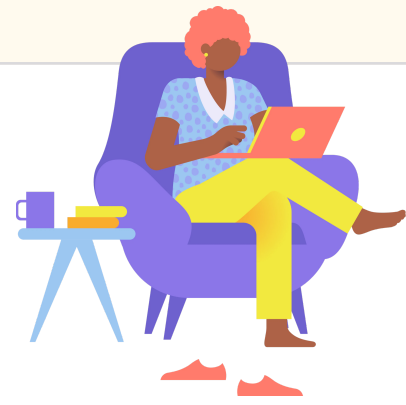
En plus d'être capable de résoudre un problème métier en sécurisant les bonnes features et le bon modèle de ML, il est de plus en plus attendu aujourd'hui qu'un Data Scientist peut déployer son modèle dans le Cloud pour l'exposer au client de son équipe.

Effectivement, si un modèle de ML réussit à avoir un impact métier fort dans votre entreprise, de plus en plus de clients voudront l'utiliser. Ainsi, vous aurez très rapidement besoin d'automatiser l'accès aux prédictions pour que vos clients n'aient plus à vous solliciter pour obtenir les résultats du modèle. Ce processus fait partie de ce que l'on appelle le déploiement à l'échelle et représente une étape clé du cycle de vie d'un modèle de ML.

L'ensemble des pratiques et raisonnements qui impactent le cycle de vie d'un modèle est souvent groupé sous le terme MLOps (Machine Learning Operations) qui se démocratise de plus en plus dans la communauté Data Science et dans les prérequis des fiches de postes. En somme, il ne suffit plus aujourd'hui d'avoir la capacité de créer des modèles, il faut également avoir les compétences techniques pour les rendre accessibles à l'échelle et de les faire vivre dans le temps, faute de quoi votre modèle cessera d'avoir l'impact métier positif qu'il a eu initialement dans votre entreprise.

Objectifs pédagogiques

- 💡 Concevoir un déploiement continu d'un moteur d'inférence sur une plateforme Cloud
- 💡 Définir et mettre en œuvre un pipeline d'entraînement des modèles
- 💡 Définir la stratégie d'élaboration d'un modèle d'apprentissage supervisé
- 💡 Évaluer les performances des modèles d'apprentissage supervisé
- 💡 Mettre en œuvre un logiciel de version de code
- 💡 Suivre la performance d'un modèle en production et en assurer la maintenance



STATUT DU PROJET

Non commencé

Contenu du projet

- 📄 [Mission - Élaborez le modèle de scoring](#)
- 📄 [Mission - Intégrez et optimisez le système MLOps](#)
- 📄 [Livrables et soutenance](#)