

Contexte du projet et objectifs

Vous êtes consultant data scientist dans une grande entreprise de conseil française. Votre client est une entreprise qui génère beaucoup de données mais pour le moment aucun modèle de Machine Learning n'a été mis en place pour les exploiter. Elle a donc naturellement fait appel à vous.

L'entreprise en question exploite plus de 3 000 magasins dans 7 pays européens. Actuellement, les directeurs de magasin sont chargés d'estimer leurs ventes quotidiennes jusqu'à six semaines à l'avance. Les ventes des magasins sont influencées par de nombreux facteurs, notamment les promotions, la concurrence, les vacances scolaires, la saisonnalité et la localité. Avec des milliers de gestionnaires individuels prédisant les ventes en fonction de leur situation particulière, la précision des résultats peut être très variable.

Avec votre équipe de consultant data scientists, vous récupérez les données disponibles et allez mener un projet complet pour les exploiter et répondre à la problématique. Vous devez préparer et soutenir, pour le 21 janvier 2021, un rendu sous forme de slides qui tirera les conclusions de votre travail, incluant une démonstration présentant les résultats du modèle de Machine Learning de manière visuelle à destination des utilisateurs finaux. Vous aurez 12 minutes pour présenter votre travail à votre donneur d'ordre client.



Les directeurs de magasin et vous lorsque vous leur montrerez vos résultats.

Les aspects à traiter dans votre projet



1

Analyse Statistique des données & Validation des données

Réaliser une analyse statistique du jeu de données et tirer des conclusions qui permettraient d'aider les équipes métiers. L'objectif de cette étape est d'extraire de l'information à partir d'indicateurs statistiques.

2

Modélisation par machine learning

Réaliser un modèle de Machine Learning qui répond à la problématique du cas d'usage à partir du jeu de données. Toutes les décisions prises concernant le choix de la métrique d'évaluation, la stratégie de validation ainsi que le choix de l'algorithme d'apprentissage, devront être justifiés.

Remarque : l'objectif de cette étape n'est pas d'optimiser d'un centième de point la loss de l'algorithme d'apprentissage mais d'établir une méthodologie argumentée et de la mettre en pratique.

3

Présentation des résultats & Recommandations

Réaliser une analyse des performances du modèle d'un point de vue métier pour convaincre les décideurs de la pertinence de l'industrialisation du modèle. Des recommandations sur l'industrialisation (ou non) devront être présentées ainsi qu'une application simple de démonstration présentant les résultats du modèle de façon visuelle.

L'application doit utiliser le modèle de Machine Learning que vous avez développé précédemment et être à destination des utilisateurs finaux.

4

Industrialisation

Considérer et analyser les éventuels problèmes de mise en production mis en œuvre dans le cadre de votre projet : maintenance, interprétabilité, sécurité.

Zoom sur le rendu du projet



Présentation des travaux en 12 minutes lors de la séance 5 (le 21 janvier) sur Teams.



Support de présentation au format .PDF ou .PPTX à déposer sur Moodle le 21 janvier **avant 14h.**



Inclure la démonstration ou un lien vers la démonstration dans le support, ou bien déposer la démonstration sur Moodle avec le support si cela est pertinent.



1 rendu et 1 présentation par *project group* (4 à 5 personnes).