



MACHINE LEARNING – DE LA CONCEPTUALISATION A L'INDUSTRIALISATION D'UN MODELE DE MACHINE LEARNING

Ce n'est plus un secret pour personne, depuis ces dernières années la quantité de données a explosé et va continuer d'augmenter de manière exponentielle dans les années à venir. Des nouveaux profils sont apparus pour exploiter cette masse d'informations ; on peut notamment citer les data scientists, récemment déclarés par Glassdoor meilleur poste 2019 aux Etats unis. Le rôle de ces profils est d'implémenter des modèles statistiques pour répondre à des enjeux métiers. Cependant la plupart de ces profils réalisent le projet qui répond aux objectifs fixés par les métiers mais qui peine à être intégré dans une chaîne de traitement des données. Une nouvelle fonction se crée pour pallier à la difficulté de mise en production des modèles de machine learning : le machine learning engineer.

Objectif du stage

L'objectif du stage est de réaliser un projet complet de data science allant de la conception d'un modèle de machine learning jusqu'à la phase d'industrialisation.

Travaux à réaliser

Intégré pour la durée de votre stage à la Taskforce Data, vous serez amené, sous la supervision d'un consultant, à réaliser les travaux suivants :

- Rédiger une note de cadrage du sujet de stage, comprenant :
 - Votre compréhension du stage et ses objectifs
 - La définition d'un cas d'usage
 - Un planning prévisionnel et ses jalons clés

Le stage se déroulera selon deux grandes phases principales et une phase optionnelle en fonction du temps restant :

- La réalisation d'un modèle de Machine Learning à partir d'un notebook Python / R qui comprendra les phases suivantes d'un projet de data science :
 - Exploration des données
 - Préparation des données
 - Feature engineering
 - Modélisation
 - Evaluation
- La mise en place d'une chaîne d'intégration continue basée sur le travail réalisé dans la première phase :
 - Déploiement du modèle sur une des grandes solutions cloud
 - Automatisation du déploiement et gestion de versions
 - Entraînement automatique avec l'arrivée de nouvelles données et A/B testing pour la vérification de non-régression entre le déploiement de deux modèles
 - Création d'alertes basées sur des métriques
- Test de charge (Phase optionnelle)
 - Création d'injecteurs de données
 - Test sur la prédiction en temps réel de plusieurs centaines/milliers d'évènements
 - Intégration de la prédiction en temps réel dans des outils de data streaming

En parallèle, vous participerez à des missions de conseil, en intégrant une équipe de consultants Nexworld placée sous la responsabilité d'un directeur de mission.

Outils et technologies (non exhaustif)

- Jupyter, RStudio
- Python, R
- Sagemaker / Azure ML / AI platform

Profil

Etudiant spécialisé en Machine Learning avec une forte appétence pour le développement et la volonté de réaliser un projet de bout-en-bout.