



## DATA SCIENCE – EXPLORER LES PLATEFORMES DE DATA SCIENCE, OUTILS DES CITIZEN DATA SCIENTIST

Une plateforme data consiste à stocker des données en très grande quantité, puis de les préparer dans le but de les analyser pour répondre à une problématique métier (optimisation, meilleure connaissance client, etc.). Cette étape d'analyse est généralement réalisée par les data scientists ou les data analysts et permet d'extraire de la connaissance à partir des données. Avec seulement une base de connaissances des algorithmes et des méthodes de machine learning, le métier peut aujourd'hui lui-même approcher l'univers de la data science en devenant citizen data scientist. Cette compétence facilite les échanges avec les data scientists et participe à réduire les incompréhensions entre les métiers et la technique. Cette possibilité est donnée par des outils facilitant l'accès à l'analyse de données, qui sont appelées plateformes de data science.

### Objectif du stage

L'objectif du stage est de comparer différentes plateformes de data science pour avoir une vue d'ensemble des possibilités que chacune propose. Le stage se déroulera selon deux grands axes :

- 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 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 :

### 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
- Définir les critères d'évaluation des solutions
- Créer et déployer un modèle de machine learning
- Tester et évaluer l'utilisation des plateformes de data science et leur capacité à créer des modèles pertinents
- Etudier et évaluer les offres, en identifiant les forces et faiblesses des outils du marché
- Réaliser la matrice des usages des techniques et des solutions
- Créer un white paper / séminaire sur l'utilisation des plateformes de data science par les Citizen Data Scientists

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
- RapidMiner, DataRobot, Dataiku, H2O.ai, Microsoft ML Studio

### Profil

Etudiant avec de bonnes compétences en Machine Learning, capable de créer un modèle pertinent et d'évaluer en profondeur les fonctionnalités et les modèles proposés par les plateformes de data science.