

• **DURÉE**

2 jours / 14h

APPRÉCIATION

Évaluation qualitative de fin de stage

**MODALITÉS
ET MOYENS**

PÉDAGOGIQUES

Démonstrations, cas pratiques, synthèse et évaluation des acquis

► **Objectifs
pédagogiques**

- Apprendre à modéliser un problème Data Science
- Apprendre à extraire de la valeur de grands ensembles de données en utilisant une variété de méthodes de Machine Learning
- Apprendre les meilleures pratiques en matière de nettoyage et de préparation de données.

► **Public concerné**

Développeurs, Data Scientists et toute personne souhaitant acquérir des connaissances dans le domaine de la Data Science.

► **Prérequis**

Avoir des connaissances de base en mathématiques, statistiques et programmation.

Data Science

DATA SCIENCE - FONDAMENTAUX

1 380 € HT

Inter-Entreprises
Intra-entreprises sur devis

La Data Science est aujourd'hui centrale dans les entreprises digitalisées. Le but de cette formation est de connaître les bonnes pratiques ainsi que les outils à utiliser en Data Science.

Introduction à la Data Science

- Qu'est-ce que la Data Science ?
- Définition
- Domaines d'application de la Data Science
- Outils et algorithmes pour la Data Science
- De l'analyse statistique au Machine Learning
- Jeux de la Data Science

Étapes de réalisation d'un projet Data Science

- Collecte des données
- Préparation et nettoyage des données
- Construction du modèle
- Évaluation du modèle
- Visualisation et analyse des résultats
- Les bonnes pratiques à adopter

Outils de la Data Science

- Bibliothèques : Scikit-Learn, Pandas, MLlib...
- Outils de développement: Jupyter Notebook, Tableau, etc.

Prétraitement des données

- Nettoyage des données
- Intégration des données
- Transformation des données
- Échantillonnage

Feature Engineering

- Types de variables : variables catégorielles / discrètes, variables continues
- Méthodes principales : sélection des caractéristiques, réduction de dimensions

Les différentes familles d'algorithmes de Machine Learning

- L'apprentissage supervisé
- L'apprentissage non-supervisé
- L'apprentissage renforcé

Techniques d'évaluation de la performance d'un modèle

- Choix de la métrique de performance
- La validation croisée
- Sur-apprentissage et sous-apprentissage
- Régularisation

La visualisation des données

- Principe de la visualisation des données
- Importance de la visualisation des données
- Principaux outils de visualisation des données

Exemples de cas d'utilisation

