

Data Science - les fondamentaux

Référence : **DATASC**

Durée : **1 jour**

Certification : **Aucune**

CONNAISSANCES PREALABLES

- Connaître l'utilité du Data Mining et les problématiques du Big Data dans les approches statistiques.

PROFIL DES STAGIAIRES

- Ingénieurs, analystes, responsables marketing. • Toute personne intéressée par les techniques de Data Science.

OBJECTIFS

- Découvrir le monde de la Data Science et les grandes problématiques associées. • Comprendre comment modéliser un problème de Data Science. • Découvrir l'intérêt des algorithmes. • Disposer d'une première approche du Deep Learning.

CERTIFICATION PREPAREE

Aucune

METHODES PEDAGOGIQUES

- Mise à disposition d'un poste de travail par stagiaire
- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience
- Le suivi de cette formation donne lieu à la signature d'une feuille d'émargement

FORMATEUR

Consultant-Formateur expert Bigdata

METHODE D'EVALUATION DES ACQUIS

- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation de fin de stage adressée avec la facture

CONTENU DU COURS

Introduction

- Compétences
- Data Science
- Machine Learning
- Big data
- Deep learning
- L'apprentissage machine

Les fondamentaux

- Approche fonctionnelle de base
- Les variables prédictives
- Les variables à prédire
- Les fonctions hypothèses
- Les estimateurs (biais et variances)
- Le compromis biais - variance
- Les fonctions de perte
- La régularisation des paramètres
- Optimisation des paramètres

La classification

- Régression logistique
- Machines à vecteurs de support (SVM)
- Arbres de décisions
- K plus proches voisins (kNN)

Les pratiques

- Prétraitement
- Compressions des données
- Réglages des modèles

L'apprentissage d'ensembles

- Principes de l'apprentissage d'ensemble
- Les forêts aléatoires
- Le gradient boosting

La régression

- Principes de la régression
- Explorations des données régressives

- Performance des modèles de régression

Le clustering

- Les k-moyens (kMeans)
- Le clustering hiérarchique
- L'approche DBSCAN

Le Deep Learning

- Perceptron
- Réseaux neuronaux multicouches (MLP)
- Réseaux neuronaux convolutifs (CNN)