

Sessions garanties disponibles

Formation Data Science Fondamentaux

Réf. SF33373

Apprenez les outils, la méthodologie, les bonnes pratiques et les principes clés de la Data Science grâce à notre formation Data Science: Fondamentaux de 5 jours !

5 j Durée	3 750 € HT Inter / participant	8 750 € HT Intra (≤12 pers.)	25 Sessions à venir
--------------	-----------------------------------	---------------------------------	------------------------

Objectifs pédagogiques

- ✓ Découvrir la Data Science et le métier de Data Scientist
- ✓ Comprendre les concepts clés de la Data Science
- ✓ Savoir utiliser les outils et méthodologies du Data Scientist
- ✓ Pouvoir programmer en R ou en Python
- ✓ Avoir des connaissances en machine learning
- ✓ Gérer l'apprentissage supervisé, semi-supervisé et non-supervisé
- ✓ Pouvoir intégrer la Data Science dans votre organisation

Public visé & prérequis

Public visé Architectes techniques, Business analysts, Chefs de projet, Data scientists, Développeurs, Ingénieurs Big Data, Managers

Prérequis Connaissances en statistiques et en programmation

Programme détaillé

Présentation de la Data Science

Big Data: définition, architecture...
La Big Data, quel enjeu pour les entreprises?
Volume, Vitesse, Variété: les 3V
Cas d'usage
Traiter les données: deep learning, analyse statistique...
Le Data Mining contre la Business Intelligence
Nouveaux challenges pour les organisations
Gouvernance des données

Principes, concepts clés de la Data Science

Data Science: présentation, définitions, vocabulaire et terminologie
Le métier de Data Scientist, « métier le plus sexy du XXIème siècle » ?
Data scientist: Rôle, responsabilités, compétences, philosophie
Objectifs et besoins
Processus de Data Science et Data Mining

Outils pour la Data Science

Outils propriétaires et outils open-source
Python et R: les langages pour la Data Science et leur environnement de développement
Notebooks Jupyter
Pandas, Numpy, SciKit-Learn... : les bibliothèques majeures
SQL, NoSQL, MongoDB : les principales bases de données pour la Data Science
Utiliser Excel, Tableau, D3.js ou encore Matplotlib pour visualiser les données
Installation d'outils pour les cas pratiques

Rappels mathématiques

Programmer avec Python et R

Python, R: les langages de la Data Science
Caractéristiques, structure
Opérations basiques, données, assignation de variables
Fonctions, listes, tableaux et packages

Obtenir et explorer les données

Trouver et importer des données
Données publiques, données privées
Entreposer les données (datawarehouse et datalake...)
Installation de bibliothèques et de packages
Pertinence des données: identification des caractéristiques des données
Contrôle de qualité: les bonnes pratiques

Prétraitement des données

De l'importance de nettoyer les données
Exemple d'un ensemble de données non-structurées
Nettoyage et préparation d'ensembles de données
Identification et résolution des valeurs manquantes, des valeurs « aberrantes »
Apache Spark et Apache Hadoop, MapReduce pour le Big Data
L'analyse en composantes principales (ACP, ou PCA pour Principal Component Analysis)
Extraire et sélectionner des features (Feature engineering)

Machine Learning pour analyser et modéliser

Entrées et sorties pour la modélisation d'un problème de Data Science
Machine Learning: capacités
Algorithmes supervisés, non-supervisés, semi-supervisés
Classification, régression d'algorithmes
Machine Learning et intuition
SciKit-Learn, gradDescent, TensorFlow... : les bibliothèques Machine Learning
Text Mining
Gestion des volumes du Big Data

Apprentissage supervisé: les méthodes

Construction d'un modèle de régression linéaire pour estimer les valeurs
Régression non-linéaire, logistique
Interpréter les coefficients de régression
Descente de gradient
Labels de jeux de données: automatisation
Méthodes ensemblistes
Forêts aléatoires, arbres de décision
SVM (machines à vecteurs de support)
Réseaux Bayésiens
Classification naïve bayésienne

Apprentissage semi-supervisé, non-supervisé, clustering

Algorithmes majeurs
Partitionnement en k-moyennes
Regroupement hiérarchique
Clustering basé sur la densité
Réseaux de neurones (Deep Learning)

Modèles d'apprentissage: évaluation et tests

Sur-apprentissage et cross-validation pour l'amélioration des modèles
Maintenance des modèles
Détérioration de la performance des modèles d'apprentissage

Visualisation et restitution des données

Prendre des décisions à partir des données

Visualiser les données: les grands principes

Utiliser Tableau Software ou encore QlikSense pour la datavisualization

Raconter une histoire avec les données avec le Data storytelling

Intégrer la Data Science dans votre organisation

Implémentations dans les processus actuels

Considérer les objectifs et le contexte professionnel pour le choix des outils

Enjeux organisationnels, éthiques et juridiques

Prochaines sessions

DATE	LIEU	STATUT
26/10/2026	à distance	Session garantie
26/10/2026	Lille	Sur confirmation
26/10/2026	Nantes	Sur confirmation
26/10/2026	Paris	Sur confirmation
30/11/2026	Lyon	Sur confirmation
30/11/2026	à distance	Sur confirmation

Pourquoi choisir Sparks Formation ?

20 ans d'expertise IT, des formateurs experts et des sessions garanties contre l'annulation.

20 ans
d'expertise

1 068+
formations

11 178
apprenants formés/an

+67
Net Promoter Score

Avis des participants

4,9/5 — 21 avis

Anthony G. — Formation très bien organisée, très claire Alison explique très clairement et est très à l'écoute, avec toujours une approche pratique essentielle. je n'avais jamais fait de Python et...

François H. — François-Marie est un excellent formateur ! Un grand merci à lui d'avoir été aussi pédagogue et clair dans ses explications du début à la fin de la formation,...

Votre formateur

Yaroub E. — Formateur Data Science

Yaroub intervient sur des formations portant sur la Data Science et le Machine Learning. Parmi les formations qui lui sont confiées figurent notamment « Machine Learning » et « Data...

Ils nous font confiance

Cette année notre filière Data Science & Machine Learning a accompagné plus de 70 entreprises dont :

SOPRA STERIA RTE AMADEUS SAS CEGEDIM LBC FRANCE SNCF - E.SNCF SOLUTIONS AMADEUS DATA PROCESSING GMBH CSTB GERS SA
ADENTIS +60 autres

Financement & certifications

Cette formation est éligible aux financements OPCO et FNE-Formation. Nos équipes vous accompagnent dans le montage de votre dossier.



Réservez votre place dès maintenant

20 ans d'expertise · 1 000+ formations · NPS +67 · réponse sous 24h ouvrées

0 805 950 800

demande@sparks-formation.com