Formation Cursus Data Scientist

Description de la formation Cursus Data Scientist

Le métier de Data Scientist est apparu ces dernières années pour faire face à la multiplication des données, à la diversité de leurs formes et de leurs sources : le Big Data. Le rôle du Data Scientist : rendre les données exploitables, les traiter pour leur donner du sens et ainsi permettre à la direction générale d'adapter la stratégie de l'entreprise.

À la croisée du statisticien, du développeur et de l'expert métier, le Data Scientist doit être capable d'appréhender les évolutions majeures des nouvelles technologies d'analyse en y intégrant une nouvelle conception des dimensions. Ce n'est qu'ainsi que son objectif final d'aide à la décision sera atteint.

Pour l'entreprise, les enjeux sont multiples : devancer les besoins des clients, anticiper les risques financiers, modifier en temps réel une politique de prix, anticiper et éviter des maladies ou une panne...

Ce cursus Data Scientist vous forme directement à cette fonction en vous donnant toutes les clés pour appréhender, manipuler et restituer les données que vous aurez à analyser dans le cadre d'un projet Big Data.

Ce cursus regroupe plusieurs cours. Les dates affichées correspondent à celles du premier module de formation.

Objectifs

Objectifs opérationnels :

Pouvoir endosser la fonction de Data Scientist : rendre les données exploitables, et leur donner du sens.

Savoir présenter et commenter les données : permettre à l'entreprise d'adapter sa stratégie grâce aux analyses effectuées.

Objectifs pédagogiques :

À l'issue de ce cursus Data Scientist, vous serez à même de maîtriser tous les tenants et aboutissants du Big Data grâce à l'assimilation des connaissances et compétences suivantes :

- Comprendre le vocabulaire des statisticiens et savoir effectuer des calculs récurrents
- Savoir situer la frontière entre statistiques et probabilités
- Savoir choisir le bon outil pour représenter vos études statistiques, et bien communiquer dessus
- Connaître les acteurs du Big Data et leur niveau d'interdépendance
- Connaître les spécificités d'une infrastructure Big Data : stockage de données, analyse, visualisation...
- Manipuler des données, des objets et programmer avec R
- Maîtriser les fonctionnalités plus avancées de R : packages, structrures de données, Rmarkdown, purr...
- Comprendre les différences entre apprentissage automatique supervisé, non supervisé et meta-apprentissage
- Maîtriser l'utilisation d'algorithmes d'auto-apprentissage adaptés à une solution d'analyse, et appliquer ces techniques à des projets Big Data
- Gérer, collecter, analyser et visualiser vos données
- Mettre en récit vos analyses pour les promouvoir en interne ou en externe

À qui s'adresse cette formation ?

Public :

De manière générale, ce cursus Data Scientist s'adresse à toute personne amenée à évoluer vers une fonction de Data Scientist. Ce poste recoupe des profils variés : analystes, statisticiens, spécialistes BI...

Prérequis :

Pour suivre ce cursus Data Scientist, il est nécessaire de posséder des connaissances de base en statistiques (régression linéaire, échantillonnage) ainsi que des connaissances de base en programmation (variables, boucles, etc.).

Des connaissances de base en SQL et dans l'utilisation de Tableau Software sont également essentielles pour aborder sereinement le volet "Visualisation des données" de ce cursus.

Contenu du cours Cursus Data Scientist

Comprendre les statistiques pour le Big Data ou la Business Intelligence

- ✓ Le vocabulaire de base
- ✓ Calcul fondamental en statistique descriptive
- ✓ Probabilités
- ✓ Tests et intervalles de confiance
- ✓ Visualisation des données
- ✓ L'évolution des statistiques pour le Big Data

Big Data : Enjeux, concepts, architectures et outils

- ✓ Contexte et opportunités du Big Data
- ✓ Sécurité éthique et enjeux juridiques du Big Data
- ✓ Open data
- ✓ Les projets Big Data en entreprise
- ✓ Architecture et infrastructure Big Data
- ✓ L'analyse des données et la visualisation
- ✓ Le développement d'applications Big Data
- ✓ La visualisation des données (Dataviz)
- ✓ Démonstration d'un environnement distribué Hadoo
- ✓ Cas d'usage et success-stories

Logiciel R : Prise en main

- ✓ Présentation du logiciel R
- ✓ Première prise en main du logiciel R
- ✓ Les Objets
- ✓ Les Fonctions et programmation R
- ✓ Génération, gestion et visualisation des données
- ✓ Analyses statistiques
- ✓ Bilan

Logiciel R : Perfectionnement et bonnes pratiques

- ✓ Organiser son travail sous R
- ✓ Manipuler facilement ses données avec le package dplyr
- ✓ Exercices
- ✓ Manipulation des variables catégorielles avec le package forecats
- ✓ Exercices
- ✓ Manipuler les chaînes de caractères avec le package stringr
- ✓ Exercices
- ✓ Manipuler des données de date : utilisation du package lubridate
- ✓ Exercices

- ✓ Assemblage de tables
- ✓ Exercices
- ✓ Réaliser des représentations graphiques performantes avec le package ggplot2
- ✓ Générer dynamiquement son rapport d'analyse avec R Markdown
- ✓ Introduction à la programmation fonctionnelle avec le package purrr
- ✓ Exercices

Machine Learning : Introduction par la pratique (3 jours)

- ✓ Introduction au monde du Big Data et de la Data Science
- ✓ Un premier exemple de modélisation : la détection de Spams
- ✓ Les différents types d'application du Machine Learning
- ✓ Prise en main des outils
- ✓ Mise en pratique sur un problème de classification
- ✓ Mise en pratique sur un problème de régression
- ✓ La validation des modèles : 1ère partie
- ✓ Une approche non-supervisée : le clustering
- ✓ Nettoyage des données : 1ère partie
- ✓ Exploration et visualisation des donnéesLa validation des modèles : 2e partie
- ✓ Le processus de création d'un modèle
- ✓ Les méthodes ensemblistes
- ✓ Le nettoyage des données : 2e partie
- ✓ Le Feature Engineering
- ✓ Ouverture sur le Deep Learning

Big Data : Les techniques d'Analyse et de Visualisation (4 jours)

- ✓ Comprendre les spécificités du Big Data
- ✓ Les concepts fondamentaux et technologies associées du Big Data (stockage, recherche, visualisation)
- ✓ Gestion des données structurées ou non
- ✓ La collecte et exploration des données
- ✓ L'analyse des données
- ✓ La visualisation des données (Dataviz)

Data Storytelling : Racontez l'histoire de vos données (1 jour)

- ✓ Concepts clés de la mise en récit des données
- ✓ Exercice
- ✓ Analyse d'une présentation, création d'indicateurs de mesure de l'histoire
- ✓ Exercice pratique
- ✓ Rédaction d'un pitch et d'un schéma narratif
- ✓ Mise en pratique
- ✓ Exercice pratique
- ✓ Prise en main de l'outil de Data Storytelling de Tableau Software, en équipe

- ✓ Exercice pratique
- ✓ Exercice individuel de construction et de présentation d'une histoire entre les participants

Travaux Pratiques

- ✓ Ce cursus Data Scientist comporte de nombreux travaux pratiques favorisant l'assimilation des connaissances.
- ✓ Les calculs et études de cas servent de fil conducteur aux multiples démonstrations. Ces derniers sont réalisés sur Excel ou en Python pour ceux qui le souhaitent.
- ✓ Le logiciel R est également très souvent utilisé. Il est accompagné par Hive pour la gestion et l'exploration des données, par Pig ou Spark pour l'ETL et le traitement des données, et par Elastick Stack pour l'analyse et la visualisation des logs.