

Data Science, les fondamentaux

Cours Synthèse de 2 jours

Réf : DTX - Prix 2022 : 1 830€ HT

Enjeu majeur de stratégie pour les organisations, la science des données permet, à partir d'outils mathématiques, de faire apparaître le comportement des données et d'analyser les événements qu'elles décrivent. Ce cours aborde les fondamentaux de la data science et permet d'acquérir une démarche d'analyse de donnée.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Connaître les principes de base de la data science et l'organisation de la démarche

Appréhender l'application de la data science afin de résoudre des questions et ses limites

Développer sa capacité d'analyse et d'interprétation des chiffres par la représentation graphique

Comprendre comment utiliser les outils de la data science et développer les modèles à des fins professionnelles

Ouverture sur l'enjeu de l'exploitation de la donnée dans un contexte concurrentiel et d'amélioration continue

Appréhender l'organisation et l'infrastructure pour les services et pour les projets de data science

MÉTHODES PÉDAGOGIQUES

Le support est organisé sous forme d'un arbre de décisions. Ainsi chacun peut sélectionner les méthodes, outils... qui le concernent directement. En fin de formation, le support devient un roadbook.

TRAVAUX PRATIQUES

Mise en pratique des fondamentaux au travers d'exercices et de mise en situation. QCM, tableaux de synthèse permettent à chacun de se positionner.

PARTICIPANTS

Directeurs/responsables des SI, responsables de projets en lien avec l'analyse de données, responsables d'études statistiques.

PRÉREQUIS

Avoir des bases en statistiques serait un plus.

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

LE PROGRAMME

dernière mise à jour : 06/2021

1) Qu'est-ce que la data science ?

- Les fondamentaux : big data, data lake, data mining, intelligence artificielle, machine et deep learning, texte mining.
- Les nouveaux défis : l'émergence et la multiplication de nouvelles sources de données.
- Hétérogénéités des données, flux temps réel et explosion des volumes de données, à prendre en compte.
- L'écosystème technologique du big data.
- Démystifier le monde de la data science : analyse descriptive, prédictive et prescriptive.
- Le métier, les outils et les méthodes du data scientist.
- Introduction au machine learning, à l'analyse supervisée et à l'analyse non supervisée.
- Notions de sur et sous-apprentissage.

Démonstration : Cas d'usages de la data science dans une chaîne de valeur métier (comportement client, offre produit...).

2) Les méthodes et les modèles de la data science

- Collecte, préparation et exploration des données.
- L'importance de la démarche de la qualité des données (nettoyer, transformer, enrichir).
- Définition des métriques.
- Les méthodes statistiques de bases.

- Les principales classes d'algorithmes supervisés : arbres de décision, K plus proches voisins, régression, Naive Bayes.
- Les principales classes d'algorithmes non-supervisés : clustering, ACP, CAH, réseaux de neurones.
- Le text mining et les autres familles d'algorithmes.

Travaux pratiques : Analyses simples avec R pour illustrer les techniques de l'analyse supervisée (régression et classification) et de l'analyse non-supervisée (clustering, segmentation et détection d'anomalies).

3) Représentation graphique et restitution des données

- Les langages de l'analyse statistique R et Python,.
- Leurs environnements de développement (R-Studio, Anaconda, PyCharm) et leurs librairies (Panda, machine learning).
- Les outils de DataViz (Power BI, Qlik, tableau...).
- Modélisation des algorithmes : représentation des processus, des flux, des contrôles et des conditions.
- Modélisation des algorithmes : les outils (Orange, Power BI).
- Communiquer les résultats par le data storytelling : organiser le visuel (diagrammes, classements, cartographies).
- Communiquer les résultats par le data storytelling : restituer la signification des résultats.

Travaux pratiques : Exercices d'exploration graphique des données, analyse de la position et de l'étendue des données (nuages, histogrammes...).

4) Modélisation d'un problème de data science

- Récapitulatif de la démarche.
- Cas métier 1 : la relation client dans l'assurance :
 - Cibler les campagnes marketing. Comprendre les causes d'attrition client. Quels produits pour quels clients ?
- Cas métier 2 : la détection des fraudes :
 - Comparer la recherche par statistiques classiques et data mining.
 - Détection par méthode supervisée. Détection par méthode non supervisée.

Etude de cas : Mise en application pratique de la méthode au storytelling sur des cas métier.

LES DATES

AIX-EN-PROVENCE

2022 : 16 juin, 25 août, 15 sept., 24 nov.

BORDEAUX

2022 : 18 juil., 13 oct., 05 déc.

LILLE

2022 : 25 juil., 15 sept., 03 nov.

LYON

2022 : 28 juil., 13 oct., 15 déc.

NANTES

2022 : 28 juil., 13 oct., 08 déc.

PARIS LA DÉFENSE

2022 : 30 mai, 25 juil., 15 sept., 13 oct., 03 nov.

SOPHIA-ANTIPOLIS

2022 : 25 août, 15 sept., 24 nov.

STRASBOURG

2022 : 28 juil., 13 oct., 08 déc.

TOULOUSE

2022 : 18 juil., 13 oct., 05 déc.

CLASSE A DISTANCE

2022 : 30 mai, 25 juil., 15 sept., 13 oct., 03 nov.