Data Science, les fondamentaux

Cours Synthèse de 2 jours - 14h
Réf : DTX - Prix 2024 : 1 950€ HT
Le prix pour les dates de sessions 2025 pourra être révisé

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 données.

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

Pendant ce cours de synthèse, le formateur effectue des démonstrations que chaque participant reproduit pour une mise en œuvre des principaux concepts.

TRAVAUX PRATIQUES

Mise en pratique guidée des fondamentaux au travers d'exercices. QCM et 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

Aucune connaissance particulière.

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 : 06/2023

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

- Les fondamentaux : big data, data lake, data mining, intelligence artificielle, machine et deep learning, text mining.
- Les nouveaux défis : l'émergence et la multiplication de nouvelles sources de données.
- Hétérogénéité 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'usage de la data science dans une chaîne de valeur métier (comportement client, offre produit, etc.).

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 base.

- 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.

Echanges : Analyses simples avec R ou Python pour illustrer les techniques de l'analyse supervisée (régression et classification) et 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 (Pandas, machine learning).
- Les outils de DataViz (Power BI, Qlik, tableau, etc.).
- Modélisation des données : représentation des processus, des flux, des contrôles et des conditions.
- Modélisation des données : 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.

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

4) Modélisation d'un problème de data science

- Récapitulatif de la démarche.
- Analyse de deux cas métier, à titre d'exemple la relation client et la détection des fraudes, mais peuvent être autres.
- 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

CLASSE À DISTANCE	PARIS
2024 : 19 nov.	2024 : 12 nov.
2025 : 03 avr., 10 juin, 25 sept., 16 déc.	2025 : 27 mars, 03 juin, 18 sept., 09 déc.