

Maîtriser la Data Science : méthodes, outils et apprentissage par renforcement

Date et durée
Code formation : DATA004FR Durée : 5 jours Nombre d'heures : 35 heures
Description
La Data Science est devenue un pilier fondamental de l'innovation et de la prise de décision stratégique dans tous les secteurs. Elle permet aux entreprises d'exploiter la valeur cachée de leurs données, de prédire des tendances, d'optimiser des processus et de créer de nouvelles opportunités. Grâce à des méthodes d'analyse avancées et des outils puissants, la science des données ouvre la voie à une meilleure compréhension des phénomènes complexes et à l'automatisation intelligente. Cette formation intensive de 5 jours est conçue pour vous permettre d'acquérir une maîtrise solide des fondations de la data science, en abordant les principales méthodes d'analyse des données et les outils de traitement et modélisation. Vous aborderez le prétraitement et l'exploration des données, en utilisant des algorithmes de machine learning pour la classification, la régression et la réduction de dimensions. Le tout, en appliquant l'apprentissage par renforcement pour une prise de décision autonome. À l'issue de ce programme complet, vous maîtriserez les méthodes, les outils clés (Python, Pandas, NumPy, etc.) et les techniques avancées, vous permettant de concevoir, d'implémenter et de déployer des projets concrets, y compris ceux intégrant l'apprentissage par renforcement. Vous serez ainsi prêt à transformer les données en décisions stratégiques et à piloter des initiatives innovantes.
Objectifs
À l'issue de cette formation avancée en Data Science, vous atteindrez les objectifs de compétences suivants : • comprendre les fondements de la Data Science et ses étapes clés ; • maîtriser les outils essentiels de la Data Science, tels que Python, Pandas et NumPy pour manipuler et analyser des données ; • préparer et explorer les données, incluant le nettoyage, la transformation et la visualisation préliminaire ; • utiliser les algorithmes du machine learning pour des tâches de classification, de régression et de réduction de dimension ; • appliquer les concepts et les algorithmes de l'apprentissage par renforcement pour la prise de décision autonome ; • implémenter des projets complets de Data Science, y compris le développement de systèmes basés sur l'apprentissage par renforcement ; • déployer des modèles de Data Science en les sauvegardant et en les exposant via des API REST.
Points forts
• Formateurs experts et praticiens : bénéficiez de l'expertise de formateurs reconnus, dotés d'une connaissance approfondie des méthodes d'analyse de données, des outils essentiels de la Data Science, et de l'Apprentissage par renforcement.

- **Mise en pratique interactive** : maîtrisez les outils et techniques grâce à des démonstrations concrètes et des ateliers pratiques. Vous serez ainsi préparé à la préparation des données, à l'utilisation du Machine Learning, à l'application de l'apprentissage par renforcement, et au déploiement de modèles de Data Science.
- **Acquisition de compétences opérationnelles** : développez une expertise concrète pour concevoir et mettre en œuvre des projets de Data Science, du prétraitement des données à l'application d'algorithmes complexes, incluant l'apprentissage par renforcement.

Modalités d'évaluation

Quiz / QCM

Travaux Pratiques

Pré-requis

Suivre cette formation nécessite les prérequis suivants :

- une familiarité avec l'environnement informatique et les concepts de base en programmation ;
- des connaissances en mathématiques et statistiques de niveau intermédiaire sont un plus, mais non obligatoires, les concepts nécessaires seront abordés durant la formation.

Public

Cette formation s'adresse aux publics suivants :

- **les débutants ou professionnels IT en reconversion** qui souhaitent acquérir des bases essentielles et une compréhension pratique de la Data Science ;
- **les analystes de données, ingénieurs et chercheurs** qui désirent approfondir leurs compétences en traitement, en modélisation des données et en apprentissage par renforcement

Programme

Module 1 : préparer et explorer les données

- Les principes fondamentaux de la data science (révision).
- La préparation et le nettoyage des données, incluant la gestion des valeurs manquantes, des doublons et des incohérences.
- La transformation des données (normalisation, standardisation, et catégorisation).
- La visualisation préliminaire des données avec Matplotlib et Seaborn.
- Les statistiques descriptives et la visualisation des distributions.
- La détection et l'identification des corrélations.
- Les méthodes de réduction de dimensions (PCA).

Module 2 : appliquer le Machine Learning (ML) et ses concepts

- Les principes fondamentaux du Machine Learning (révision).
- Les algorithmes de classification (régression logistique, K-Nearest Neighbors et arbre de décision).
- Les algorithmes de régression (régression linéaire et régression polynomiale).
- L'évaluation des modèles (matrice de confusion, précision, rappel et courbe ROC).
- Les concepts de base de l'apprentissage par renforcement (reinforcement learning).
- Les différences entre apprentissage supervisé, non supervisé et par renforcement.
- Les applications de l'apprentissage par renforcement (jeux, robotique, systèmes de recommandation).
- L'implémentation pratique (développement d'un modèle de ML).

Module 3 : maîtriser la modélisation avancée et l'apprentissage par renforcement

- Les algorithmes de clustering (K-means, DBSCAN et clustering hiérarchique).
- Les applications des algorithmes de clustering dans l'analyse des segments de données.
- L'évaluation des modèles de clustering (silhouette score et coefficient de Rand).
- Les concepts de base du Deep Learning et des réseaux de neurones.
- L'apprentissage par renforcement avancé (Q-learning et Deep Q Networks).
- La mise en place d'un problème simple d'apprentissage par renforcement (jeu simple ou robot).
- L'application d'un algorithme d'apprentissage par renforcement pour un cas concret.

Module 4 : déployer les modèles de Data Science

- La sauvegarde d'un modèle avec Joblib ou Pickle.
- Le déploiement d'un modèle sur un environnement cloud ou un serveur.
- La création d'API REST pour exposer le modèle et réaliser des prédictions à distance.