



Data Science

Big Data - Développement d'applications de Machine Learning et d'Intelligence Artificielle (IA)

5 jours (35h00) | ★★★★★ 4,6/5 | BIG-DEVD | Évaluation qualitative de fin de stage | Formation délivrée en présentiel ou distanciel

Formations Informatique › Big Data › Data Science

Contenu mis à jour le 13/10/2023. Document téléchargé le 23/05/2024.

Objectifs de formation

A l'issue de cette formation, vous serez capable de :

- Développer des applications de Machine Learning et d'IA (Intelligence Artificielle) avec Spark et Python
- Utiliser la programmation parallèle sur un cluster
- Développer et optimiser des algorithmes standards de Machine Learning et d'IA
- Utiliser les bibliothèques Python pour le Machine Learning et d'IA
- Décrire le cycle de vie d'un projet Data Science.

Modalités, méthodes et moyens pédagogiques

Formation délivrée en présentiel ou distanciel* (blended-learning, e-learning, classe virtuelle, présentiel à distance).

Le formateur alterne entre méthode** démonstrative, interrogative et active (via des travaux pratiques et/ou des mises en situation).

Variables suivant les formations, les moyens pédagogiques mis en oeuvre sont :

- Ordinateurs Mac ou PC (sauf pour certains cours de l'offre Management), connexion internet fibre, tableau blanc ou paperboard, vidéoprojecteur ou écran tactile interactif (pour le distanciel)
- Environnements de formation installés sur les postes de travail ou en ligne
- Supports de cours et exercices

En cas de formation intra sur site externe à M2i, le client s'assure et s'engage également à avoir toutes les ressources matérielles pédagogiques nécessaires (équipements informatiques...) au bon déroulement de l'action de formation visée conformément aux prérequis indiqués dans le programme de formation communiqué.

* nous consulter pour la faisabilité en distanciel

** ratio variable selon le cours suivi

Prérequis

Avoir des connaissances en mathématiques algébriques pour pouvoir interpréter les algorithmes à mettre en oeuvre, et connaître le langage Python (des rappels seront faits pendant la formation).

Public concerné

Toute personne souhaitant acquérir des compétences dans le développement d'applications de Machine Learning et d'IA (Intelligence Artificielle).

Cette formation :

- Est animée par un consultant-formateur dont les compétences techniques, professionnelles et pédagogiques ont été validées par des diplômes et/ou testées et approuvées par l'éditeur et/ou par M2i Formation
- Bénéficie d'un suivi de son exécution par une feuille de présence élargée par demi-journée par les stagiaires et le formateur.

Programme

Jour 1

Vue d'ensemble du Big Data, du Machine Learning et de l'IA

- Introduction au Big Data et à l'IA (Intelligence Artificielle) : de quoi s'agit-il ?
- Perspectives offertes par le Big Data et l'IA
- Les acteurs du Big Data et de l'IA
- Exemples pratiques
- Les technologies concernées
- Les différents métiers
- Aspects économiques (OPEX, CAPEX, TRI) du Cloud vs On-Premise
- Démonstration d'applications

Rappels sur la Data Science

- Les bases du langage Python
- Modélisation des problématiques grâce à des vecteurs et matrices
- Probabilités, statistiques, statistiques descriptives, statistiques explicatives
- Entropie, gain d'information
- Compromis biais-complexité
- La malédiction de la dimension
- Matrice de confusion
- Gérer les valeurs manquantes MCAR / MAR / MNAR
- Validation croisée
- Courbe d'apprentissage
- Fondamentaux du nettoyage des données
- Réduction de la dimension par sélection de variables et par transformation de variables
- Métriques de performance d'un modèle :
 - R^2
 - Accuracy
 - Précision et rappel (recall)
 - F1-score
 - AUC
 - P-valeur

Exemples de travaux pratiques (à titre indicatif)

- Chargement, exploration, analyse, nettoyage, et premières modélisations sur un jeu de données

Jour 2

Machine Learning et IA : algorithmes

- Les concepts du Machine Learning et de l'IA
- Les données supervisées ou non supervisées : quelles différences ?
- Les librairies :
 - Scikit-learn
 - Tensorflow
 - PyTorch
 - Keras
- Régression
- Modèles linéaires
- Classification
- Naive Bayes
- K-NN
- K-Means clustering

Exemples de travaux pratiques (à titre indicatif)

- Mise en oeuvre des algorithmes vus en fonction de la finalité recherchée (régression, classification, imputation de valeurs manquantes) et en utilisant les librairies Scikit-learn et Keras

Jour 3

Machine Learning et IA : algorithmes (suite)

- DBScan
- Arbres de décision et de régression
- Support Vector Machines (SVM)
- Réseaux de neurones et apprentissage profond (Deep Learning)
- Réseaux convolutifs, réseaux récurrents

Exemples de travaux pratiques (à titre indicatif)

- Modélisations avancées (DL, CNN, RNN) avec optimisation des architectures et hyper-paramètres

Jour 4

Développer avec Spark

- Introduction à Apache Spark
- Faire du Machine Learning avec Apache Spark MLlib
- Traiter les données en temps réel avec Apache Spark Streaming
- Faire des requêtes au format SQL avec Apache Spark SQL
- Modélisation de réseaux sociaux grâce à des graphes avec Apache Spark GraphFrames
- Démonstration du travail collaboratif avec MLFlow

Exemples de travaux pratiques (à titre indicatif)

- Exercices d'application avec Spark MLlib/Streaming/SQL/GraphFrames
- Mise en oeuvre de MLFlow pour comparer les modèles

Jour 5

Outil de visualisation des données (ou Dataviz)

- Définition de la Dataviz
- Les acteurs de la Dataviz
- Principes de la Dataviz
- Exemples d'analyse (sous forme de démo) :
 - Bibliothèques de Dataviz de Python
 - Tableau Desktop / Public
 - Microsoft Power BI

Introduction à MLOps et aux architectures serverless

- Les approches DevOps et GitOps
- Bases de la conteneurisation et de l'automatisation
- Les architectures serverless
- Développer et déployer des applications MLOps en production (démonstration)
- Monitoring et amélioration continue des applications MLOps (démonstration)

Exemples de travaux pratiques (à titre indicatif)

- *Prise en main de Tableau ou MS Power BI*
- *Utilisation de bibliothèques Python (Matplotlib, Seaborn)*
- *Création et exécution d'un pipeline MLOps avec serving et monitoring du modèle*

Examen M2i (en option)

- Prévoir l'achat de l'examen en supplément
- L'examen (en français) sera passé le dernier jour, à l'issue de la formation et s'effectuera en ligne
- Il s'agit d'un QCM dont la durée moyenne est d'1h30 et dont le score obtenu attestera d'un niveau de compétence
- L'examen n'est pas éligible au CPF, mais permettra néanmoins de valider vos acquis

Le contenu de ce programme peut faire l'objet d'adaptation selon les niveaux, prérequis et besoins des apprenants.

Modalités d'évaluation des acquis

- En cours de formation, par des études de cas ou des travaux pratiques
- Et, en fin de formation, par un questionnaire d'auto-évaluation et/ou un examen M2i

Les + de la formation

Un examen M2i permettant de valider vos acquis à l'issue de la formation est disponible sur demande (coût : 120€).

Accessibilité de la formation

Le groupe M2i s'engage pour faciliter l'accessibilité de ses formations. Les détails de l'accueil des personnes en situation de handicap sont consultables sur la page Accueil et Handicap.

Modalités et délais d'accès à la formation

Les formations M2i sont disponibles selon les modalités proposées sur la page programme. Les inscriptions sont possibles jusqu'à 48 heures ouvrées avant le début de la formation. Dans le cas d'une formation financée par le CPF, ce délai est porté à 11 jours ouvrés.