

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

5 jours (35 heures)

Perfectionnement

Présentiel ou classe à distance

à partir de

4 375 € HT par personne (TVA en vigueur) **Repas inclus en présentiel**

Contenu mis à jour le 25/09/2026. Document téléchargé le 07/10/2026.

Objectifs

À 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.

Les compétences visées par cette formation font l'objet d'une évaluation.

À qui s'adresse cette formation ?

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

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).

Programme

1 Pendant votre formation

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

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 (démon)
- Monitoring et amélioration continue des applications MLOps (démon)

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

Financement

Plusieurs dispositifs de financement sont accessibles via les OPCO (Opérateurs de Compétences), organismes agréés par le ministère du Travail dont le rôle est d'accompagner, collecter et gérer les contributions des entreprises au titre du financement de la formation professionnelle.

Pour plus d'information, une équipe de gestionnaires SKOLAE Formation spécialisée vous accompagne dans le choix de vos formations et la gestion administrative.

Nous contacter pour en savoir plus →

Modalités 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

Accessibilité aux personnes en situation de handicap

Détails de la session

Du 10 mai 2027 au 14 mai 2027 (5 jours / 35 heures)

Paris La Défense

Du 10 mai 2027 au 14 mai 2027

Paris La Défense

- 10 mai 2027 9h00 – 15h00
- 11 mai 2027 9h00 – 15h00
- 12 mai 2027 9h00 – 15h00
- 13 mai 2027 9h00 – 15h00
- 14 mai 2027 9h00 – 15h00