

Formation Apprentissage

Code : EAI-ML-100 Durée : 5 jours (35 heures)

Public visé

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

{{< formation-prerequis >}}

Objectifs pédagogiques

À l'issue de cette formation, les apprenants seront capables de : - Développer des applications de Machine Learning et d'IA 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 l'IA - Décrire le cycle de vie d'un projet Data Science

Programme

Jour 1

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

- Introduction au Big Data et à l'IA
- Perspectives offertes par le Big Data et l'IA
- Les acteurs du secteur
- Technologies concernées et métiers associés
- Aspects économiques (OPEX, CAPEX, TRI) du Cloud vs On-Premise

Rappels sur la Data Science

- Bases du langage Python
- Modélisation par vecteurs et matrices
- Probabilités et statistiques (descriptives/explicatives)
- Entropie et gain d'information
- Compromis biais-complexité et malédiction de la dimension
- Matrice de confusion
- Gestion des valeurs manquantes (MCAR/MAR/MNAR)
- Validation croisée et courbes d'apprentissage
- Nettoyage des données et réduction dimensionnelle

- Métriques : R^2 , Accuracy, Précision, Rappel, F1-score, AUC, P-valeur

Travaux pratiques

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

Jour 2

Machine Learning et IA : Algorithmes

- Concepts du Machine Learning et de l'IA
- Données supervisées vs non supervisées
- Bibliothèques : Scikit-learn, Tensorflow, PyTorch, Keras
- Régression et modèles linéaires
- Classification
- Naive Bayes, K-NN, K-Means clustering

Travaux pratiques

- Mise en oeuvre des algorithmes selon finalité (régression, classification, imputation) avec 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 et récurrents

Travaux pratiques

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

Jour 4

Développer avec Spark

- Introduction à Apache Spark
- Machine Learning avec Apache Spark MLlib
- Traitement temps réel avec Apache Spark Streaming
- Requêtes SQL avec Apache Spark SQL
- Modélisation de réseaux sociaux avec Apache Spark GraphFrames
- Travail collaboratif avec MLFlow

Travaux pratiques

- Exercices avec Spark MLlib/Streaming/SQL/GraphFrames et implémentation de MLFlow pour comparaison de modèles

Jour 5

Outil de visualisation des données (Dataviz)

- Définition et acteurs de la Dataviz
- Principes de la Dataviz

- Bibliothèques Python (Matplotlib, Seaborn)
- Tableau Desktop/Public et Microsoft Power BI

Introduction à MLOps et architectures serverless

- Approches DevOps et GitOps
- Conteneurisation et automatisation
- Architectures serverless
- Développement et déploiement d'applications MLOps
- Monitoring et amélioration continue

Travaux pratiques

- Prise en main de Tableau ou Power BI, utilisation de bibliothèques Python, création et exécution d'un pipeline MLOps avec serving et monitoring

Modalités d'évaluation des acquis

En cours de formation, par des travaux pratiques. En fin de formation, par un questionnaire d'auto-évaluation.

{{< formation-require >}}

{{< formation-seealso >}}