



Big Data - MLOps et DataOps

# Python et MLflow - Concevoir, déployer et surveiller un pipeline MLOps

5 jours (35h00) | 9 4,6/5 | MLOPS-MLF | Évaluation qualitative de fin de stage | Formation délivrée en présentiel ou distanciel

Formations Informatique > Data > Big Data - MLOps et DataOps

Contenu mis à jour le 13/10/2023. Document téléchargé le 27/07/2024.

## Objectifs de formation

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

- Concevoir et tester un modèle de ML (Machine Learning) avec MLflow
- Concevoir et tester un pipeline MLOps robuste avec MLflow
- Déployer votre modèle en production
- Surveiller votre modèle de production et corriger les écarts.

## 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 bases de données et/ou analytiques.

## Public concerné

Data Scientists, Data Analysts, ingénieurs Data, ingénieurs DevOps, développeurs, architectes Data, chefs de projets, managers.

## 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

### Fondamentaux

- Qu'est-ce qu'une donnée ?
- Cycle de vie d'une donnée
- Cycle de vie de la Data Science
- Evolution des analytiques
- Apports des analytiques et du Cloud
- Qu'est-ce qu'une organisation Data Driven ?
- Business Intelligence vs Business Analytics
- Du Data Warehouse au Data Lake et au Lakehouse
- Gouvernance et qualité des données
- Lean, Agilité et DevOps, MLOps appliqués aux données
- Conteneurisation et architectures micro-services
- Apports de MLOps
- Workflow de MLOps
- DevOps vs DataOps vs MLOps
- Machine Learning et opérations : briser le mur de la confusion
- Constituer et former une équipe MLOps
- Plateformes et outils de MLOps
- Qu'est-ce que MLflow ?

### **Exemple de travaux pratiques (à titre indicatif)**

- *Mise en pratique : développer un premier modèle avec MLflow*

## Jour 2

### Démarrer avec MLflow

- MLflow Projects
- MLflow Tracking
- MLflow Models
- MLflow Model Registry

### **Exemple de travaux pratiques (à titre indicatif)**

- *Création d'un premier pipeline avec MLflow*

### Développement du modèle et expérimentations

- Définir l'expérimentation : explorer le jeu de données
- Ajouter des expérimentations : étapes pour concevoir un classificateur logistique
- Comparer différents modèles
- Affiner le modèle en optimisant les hyperparamètres
- Gérer les signatures et schémas des modèles
- Ajouter le meilleur modèle au Model Registry
- Gérer le cycle de vie de développement du modèle

## Jour 3

### Le Machine Learning en production

- Comprendre les challenges des systèmes et projets de ML
- Plateformes de ML : Michelangelo, Kubeflow
- Spécifier l'architecture d'une plateforme de ML
- Architecture des systèmes de haut niveau
- MLflow et les autres outils de l'écosystème

## Jour 4

### Préparer les données et entraîner le modèle

- Structurer le pipeline de données du projet
- Sourcer les données
- Vérifier la qualité des données
- Générer l'ensemble des attributs et des données d'apprentissage
- Exécuter le pipeline de bout-à-bout
- Utiliser les feature stores
- Créer le projet d'apprentissage avec MLflow
- Implémenter le job d'apprentissage
- Evaluer le modèle
- Déployer le modèle dans le Model Registry
- Créer une image Docker pour le job d'apprentissage

## Jour 5

### Déploiement et inférence avec MLflow

- Créer un Model Registry local
- Créer un job batch d'inférence
- Créer un process d'API pour l'inférence
- Déployer les modèles pour du scoring batch dans Kubernetes
- Réaliser un déploiement sur le Cloud avec AWS SageMaker

### Mettre à niveau le workflow de ML

- Intégrer MLflow avec Spark
- Intégrer MLflow avec NVIDIA RAPIDS (GPU)
- Suivre la performance des modèles en production
- Monitoring du Data Drift, du Model Drift, et du Target Drift
- Monitoring et alertes de l'infrastructure

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

## **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.