

MLOps Fundamentals

Durée: 3 Jours **Réf de cours: GK840034** **Méthodes d'apprentissage: Classe à distance**

Résumé:

La formation "Les Fondamentaux de MLOps" est un guide complet sur les principes, les composants et les outils utilisés dans les opérations de machine learning (MLOps). Elle offre une compréhension approfondie du cycle de vie du machine learning, du cycle de vie de MLOps, et des outils impliqués, tels que MLFlow et KubeFlow.

Vous apprendrez à mettre en place un projet de machine learning, y compris l'utilisation de Git et GitHub, la configuration d'environnements virtuels et les hooks de pré-commit. Le cours abordera ensuite les fondamentaux de la gestion des données, tels que la compréhension des cycles de vie des données, la versionnage des données, la gouvernance et les solutions de stockage.

Des démonstrations pratiques et concrètes seront fournies sur l'analyse exploratoire des données (EDA), l'ingénierie des caractéristiques et le nettoyage des données en utilisant Pandas et Matplotlib. Le cours explorera également le concept des magasins de caractéristiques, leurs types, leur fonctionnement, les meilleures pratiques et les défis de mise en œuvre.

Mise à jour : 09.12.2024

Public visé:

Les praticiens du ML voulant passer en production des jeux de données ; Data Scientists, Développeurs ou Software Engineers

Objectifs pédagogiques:

- À l'issue de la formation, les participants seront capables de :
- Expliquer les principes de base et les défis de MLOps dans le contexte du cycle de vie d'un projet ML
- Concevoir et structurer des projets ML en utilisant les meilleures pratiques et des outils standards de l'industrie
- Analyser et préparer des ensembles de données pour des applications de machine learning en utilisant des techniques avancées de gestion des données
- Développer et évaluer des modèles de machine learning en utilisant des métriques appropriées et des outils de suivi des expériences
- Appliquer des techniques de contrôle de version et de containerisation pour assurer la reproductibilité des projets ML
- Mettre en œuvre des stratégies de test complètes pour les composants et pipelines ML
- Déployer des modèles ML en utilisant des APIs RESTful et des technologies de containerisation
- Construire des pipelines CI/CD automatisés pour les projets ML,
- Concevoir et mettre en œuvre des systèmes de surveillance pour les modèles ML déployés
- Évaluer et traiter la dérive des modèles et la dégradation des performances dans les environnements de production
- Intégrer les pratiques d'ingénierie des données dans les workflows MLOps
- Créer un pipeline MLOps de bout en bout en incorporant tous les concepts appris

Pré-requis:

Pour assister à cette formation, il est recommandé pour les candidats de:

Avoir suivi la formation GK821537 "Construire des applications intelligentes avec l'IA/ML Niveau 2" ou posséder les connaissances équivalentes

- Maîtriser Python,
- Avoir une bonne compréhension du ML
- Connaître Git et le contrôle de version

■ Avoir une expérience avec les plateformes cloud
Afin de vous assurer que vous possédez toutes les connaissances
requises pour participer à la formation, notre équipe commerciale
vous proposera un QCM.

Contenu:

Jour 1

Introduction à MLOps

- Qu'est-ce que MLOps
- Aperçu du Machine Learning Life Cycle

Composants et Outils de MLOps

- Aperçu du cycle de vie de MLOps / Composants de MLOps et Avantages
- Aperçu des outils de MLOps (MLFlow, KubeFlow, etc.) et leur rôle dans l'automatisation des pipelines ML

Mise en place d'un projet ML

- Configuration de Git et GitHub
- Configuration des environnements virtuels
- Hooks de pré-commit

Fondamentaux de la gestion des données

- Comprendre les cycles de vie des données
- Versionnage des données
- Gouvernance des données
- Solutions de stockage des données

Démo : EDA, Ingénierie des caractéristiques et Nettoyage des données

- EDA pratique en utilisant Pandas pour résumer le jeu de données.
- Visualisation des distributions en utilisant Matplotlib (histogrammes, diagrammes de dispersion).
- Création de nouvelles caractéristiques et nettoyage des données en supprimant les valeurs manquantes et les valeurs aberrantes.

Feature Stores

- Introduction aux Feature Stores
- Types de Feature Stores
- Fonctionnement des Feature Stores
- Meilleures pratiques pour l'utilisation des Feature Stores
- Défis de la mise en œuvre des Feature Stores

Développement de modèles

- Aperçu du processus de développement de modèles
- Choisir le bon algorithme
- Entraînement et validation des modèles
- Éviter le surapprentissage
- Métriques d'évaluation des modèles

Mise en œuvre d'un pipeline ML de base

Stratégies de développement de modèles

- Aperçu des approches de développement de modèles
- Approches centrées sur les données vs. centrées sur les modèles
- Expérimentation dans le développement de modèles
- Développement collaboratif dans MLOps

Interprétabilité et explicabilité des modèles ML

- Introduction à l'interprétabilité et à l'explicabilité des modèles
- Techniques pour l'interprétabilité des modèles
- Explicabilité dans différents types de modèles
- Outils pour l'interprétabilité
- Défis de l'explicabilité

Mise en œuvre des algorithmes

- Sélection d'un algorithme
- Mise en œuvre de l'algorithme choisi
- Évaluation des performances de l'algorithme
- Comparaison de plusieurs algorithmes

Démo : Sélection, mise en œuvre et évaluation des algorithmes

- Sélectionner un ensemble de données, choisir deux algorithmes différents (par exemple, arbre de décision et SVM)
- Mettre en œuvre les algorithmes en utilisant scikit-learn
- Évaluer les performances de chaque algorithme
- Comparer les résultats en utilisant des métriques telles que la précision, la précision, etc.

Suivi des expériences et évaluation des modèles

- Introduction au suivi des expériences
- Mise en place du suivi des expériences
- Évaluation des performances des modèles
- Visualisation des performances des modèles

Mise en place de MLflow pour le suivi des expériences

- Introduction à MLflow
- Suivi des expériences avec MLflow
- Comparaison de plusieurs exécutions
- Stockage et récupération des modèles

Évaluation des modèles

Stratégies de service et de déploiement des modèles

- Introduction au service des modèles
- Stratégies de déploiement
- Conteneurisation des modèles ML
- Service des modèles avec Docker
- Cadres de service des modèles
- Déploiement des modèles sur des plateformes cloud

Problèmes juridiques et de conformité dans MLOps

- Introduction aux problèmes juridiques et de conformité dans MLOps
- Normes réglementaires clés
- Gouvernance et conformité des modèles
- Défis des problèmes juridiques et de conformité

Conteneurisation des modèles ML avec Docker

- Introduction à Docker
- Configuration de Docker
- Création d'une image Docker
- Déploiement des conteneurs Docker sur des plateformes cloud

Déploiement des modèles sur des plateformes cloud

- Introduction au déploiement sur le cloud
- Préparation du modèle pour le déploiement
- Configuration de l'infrastructure cloud
- Déploiement du modèle avec Ray Serve

Entraînement fédéré et déploiements en périphérie

- Introduction à l'apprentissage fédéré et à l'informatique en périphérie
- Architecture d'entraînement fédéré
- Déploiement de modèles en périphérie
- Outils et cadres
- Défis de l'apprentissage fédéré et de l'informatique en périphérie

CI/CD pour ML

- Introduction à CI/CD pour l'apprentissage automatique
- Mise en place de pipelines CI/CD pour ML
- Intégration de CI/CD avec le suivi des expériences
- Automatisation de la validation et des tests des modèles

Mise en place de pipelines CI/CD pour ML

- Introduction à GitHub Actions pour CI/CD

- Construction du pipeline
- Intégration du prétraitement et du développement du modèle
- Entraînement et évaluation du pipeline
- Introduction à l'automatisation du pipeline

Jour 2

- Préparation de l'environnement d'évaluation
- Évaluation des performances des modèles
- Comparaison des modèles en fonction de l'évaluation

Techniques de réglage des hyperparamètres

- Introduction au réglage des hyperparamètres
- Recherche en grille vs. recherche aléatoire
- Optimisation bayésienne
- Considérations pratiques

Réglage automatisé des hyperparamètres

- Introduction au réglage automatisé des hyperparamètres
- Exécution du réglage des hyperparamètres
- Analyse des résultats

Jour 3

- Automatisation de l'entraînement et du déploiement des modèles
- Intégration de MLflow avec CI/CD
- Test du pipeline CI/CD

Surveillance et maintenance des systèmes ML

- Introduction à la surveillance des systèmes ML
- Outils pour la surveillance des modèles ML
- Mise en place d'alertes pour la dérive des modèles
- Surveillance des performances des modèles en temps réel
- Boucles de rétroaction continues
- Mise à l'échelle de la surveillance pour les déploiements à grande échelle

Mise en œuvre des outils de surveillance

- Introduction aux outils de surveillance
- Instrumentation du modèle ML pour la surveillance
- Mise en œuvre du code - Exposition des métriques pour Prometheus
- Visualisation des métriques dans Grafana

Méthodes pédagogiques :

Les participants réalisent un test d'évaluation des connaissances en amont et en aval de la formation pour valider les connaissances acquises pendant la formation.

Un support de cours électronique sera remis aux participants.

Suivi de cette formation à distance depuis un site client équipé. Il suffit d'avoir une bonne connexion internet, un casque avec micro et d'être dans un endroit au calme pour en profiter pleinement

Une fiche explicative est adressée en amont aux participants pour leur permettre de vérifier leur installation technique et de se familiariser avec la solution technologique utilisée.

L'accès à l'environnement d'apprentissage, ainsi qu'aux feuilles d'émargement et d'évaluation est assuré.

En savoir plus : <https://www.globalknowledge.com/fr-fr/solutions/methodes-d'apprentissage/classe-a-distance>

Autres moyens pédagogiques et de suivi:

- Compétence du formateur : Les experts qui animent la formation sont des spécialistes des matières abordées et ont au minimum cinq ans d'expérience d'animation. Nos équipes ont validé à la fois leurs connaissances techniques (certifications le cas échéant) ainsi que leur compétence pédagogique.
- Suivi d'exécution : Une feuille d'émargement par demi-journée de présence est signée par tous les participants et le formateur.
- En fin de formation, le participant est invité à s'auto-évaluer sur l'atteinte des objectifs énoncés, et à répondre à un questionnaire de satisfaction qui sera ensuite étudié par nos équipes pédagogiques en vue de maintenir et d'améliorer la qualité de nos prestations.

Délais d'inscription :

- Vous pouvez vous inscrire sur l'une de nos sessions planifiées en inter-entreprises jusqu'à 5 jours ouvrés avant le début de la formation sous réserve de disponibilité de places et de labs le cas échéant.
- Votre place sera confirmée à la réception d'un devis ou ""booking form"" signé. Vous recevrez ensuite la convocation et les modalités d'accès en présentiel ou distanciel.
- Attention, si cette formation est éligible au Compte Personnel de Formation, vous devrez respecter un délai minimum et non négociable fixé à 11 jours ouvrés avant le début de la session pour vous inscrire via moncompteformation.gouv.fr.

Accueil des bénéficiaires :

- En cas de handicap : plus d'info sur globalknowledge.fr/handicap
- Le Règlement intérieur est disponible sur globalknowledge.fr/reglement