Joachim Lombardi

Simplon |

MLFlow

Tutoriel

Table des matières

[**Chapitre 1 : Introduction à MLflow** 2](#_Toc162093335)

[**Chapitre 2 : Utilisation de MLflow** 2](#_Toc162093336)

[**Chapitre 3 : Gestion des modèles avec MLflow** 2](#_Toc162093337)

[**Chapitre 4 : Déploiement des modèles** 3](#_Toc162093338)

[**Chapitre 5 : Exemple d'application de MLflow** 3](#_Toc162093339)

[**Chapitre 6 : Atelier pratique sur MLflow** 3](#_Toc162093340)

[**Conclusion** 4](#_Toc162093341)

**Chapitre 1 : Introduction à MLflow**

1.1 Présentation de MLflow

* Définition de MLflow et de son rôle dans la gestion du cycle de vie des projets de machine learning.
* Contexte de développement de MLflow et sa place dans l'écosystème des outils de machine learning.

1.2 Objectifs et avantages de MLflow

* Discussion sur les principaux objectifs de MLflow, tels que la reproductibilité, le suivi des expériences et la gestion des modèles.
* Avantages de l'utilisation de MLflow par rapport aux approches traditionnelles.

1.3 Structure du chapitre

* Présentation de la structure du chapitre et des sujets qui seront abordés dans les sections suivantes.

**Chapitre 2 : Utilisation de MLflow**

2.1 Installation de MLflow

* Instructions détaillées pour installer MLflow sur différentes plateformes (par exemple, via pip, conda, Docker).

2.2 Initialisation du projet MLflow

* Étapes pour initialiser un nouveau projet MLflow et structurer le code selon les bonnes pratiques recommandées.

2.3 Suivi des expériences

* Explication de la fonctionnalité de suivi de MLflow et démonstration de son utilisation pour enregistrer les paramètres, métriques et artefacts de chaque expérience.

**Chapitre 3 : Gestion des modèles avec MLflow**

3.1 Empaquetage des modèles

* Guide pas à pas sur l'empaquetage des modèles entraînés dans des conteneurs compatibles avec MLflow.

3.2 Enregistrement des modèles

* Utilisation des fonctionnalités de MLflow pour enregistrer et versionner les modèles entraînés.

**Chapitre 4 : Déploiement des modèles**

4.1 Déploiement local

* Méthodes pour déployer des modèles localement à l'aide de serveurs web ou d'autres environnements.

4.2 Déploiement en production

* Discussion sur les meilleures pratiques pour déployer des modèles dans des environnements de production, y compris les services cloud et les conteneurs orchestrés.

**Chapitre 5 : Exemple d'application de MLflow**

5.1 Description de l'exemple

* Présentation d'un exemple concret d'utilisation de MLflow pour suivre et gérer un projet de machine learning.

5.2 Étapes de l'exemple

* Détailler les étapes spécifiques de l'exemple, y compris la préparation des données, la création du modèle, l'entraînement, l'évaluation et le déploiement.

**Chapitre 6 : Atelier pratique sur MLflow**

6.1 Objectif de l'atelier

* Présentation des objectifs et des attentes pour l'atelier pratique.

6.2 Déroulement de l'atelier

* Description des différentes étapes de l'atelier, y compris l'introduction, l'exercice en groupes et la présentation des résultats.

6.3 Ressources supplémentaires

* Références et liens vers des ressources supplémentaires pour aider les participants à compléter l'atelier avec succès.

**Conclusion**

7.1 Récapitulatif des principaux points

* Résumé des principaux concepts et techniques présentés dans le tutoriel.

7.2 Prochaines étapes

* Suggestions pour approfondir ses connaissances sur MLflow et explorer d'autres aspects de la gestion du cycle de vie des projets de machine learning.

MLflow est une plateforme open-source développée par Databricks, conçue pour aider les praticiens du machine learning à gérer le cycle de vie complet de leurs projets, de la phase d'exploration et d'entraînement des modèles à leur déploiement en production. Son rôle principal est de fournir des outils et des fonctionnalités pour rendre ce processus plus fluide, plus organisé et plus reproductible.

**Contexte de développement de MLflow et sa place dans l'écosystème des outils de machine learning :**

MLflow a été développé en réponse aux défis rencontrés par les praticiens du machine learning lors du développement, du suivi et du déploiement de modèles. Son développement a été initié par l'équipe de Databricks, une société fondée par les créateurs d'Apache Spark, dans le but de créer une plateforme open-source et polyvalente pour gérer le cycle de vie des projets de machine learning.

**Facteurs clés qui ont contribué au développement de MLflow :**

1. **Complexité croissante des projets de machine learning :** Avec la montée en puissance du machine learning et de l'intelligence artificielle, les projets de machine learning sont devenus de plus en plus complexes, impliquant souvent plusieurs étapes et itérations. Il est devenu crucial d'avoir des outils pour gérer cette complexité et garantir la reproductibilité des résultats.
2. **Manque d'outils intégrés pour le suivi et la gestion des modèles :** Avant l'émergence de MLflow, de nombreux praticiens du machine learning utilisaient une combinaison d'outils disparates pour suivre les expériences, gérer les modèles et les déployer en production. Cela entraînait souvent des workflows fragmentés et peu efficaces.
3. **Besoin de transparence et de reproductibilité :** Avec l'accent croissant mis sur la transparence et la reproductibilité dans la recherche en machine learning, il est devenu essentiel d'avoir des outils qui permettent aux praticiens de documenter et de partager leurs expériences d'entraînement de modèles de manière rigoureuse.

**Place de MLflow dans l'écosystème des outils de machine learning :**

MLflow occupe une place centrale dans l'écosystème des outils de machine learning en offrant une solution complète pour la gestion du cycle de vie des projets. Bien qu'il existe de nombreux autres outils spécialisés dans des aspects spécifiques du processus de développement de modèles (comme TensorFlow pour l'entraînement de modèles deep learning ou scikit-learn pour les modèles classiques), MLflow se distingue par sa capacité à rassembler ces différentes étapes dans un seul cadre cohérent.

En tant que plateforme open-source, MLflow bénéficie également d'une large communauté d'utilisateurs et de contributeurs, ce qui en fait un choix attrayant pour de nombreuses équipes de machine learning. De plus, son intégration avec d'autres outils et frameworks populaires, tels que TensorFlow, PyTorch et Kubernetes, renforce sa position en tant qu'élément essentiel de l'infrastructure de machine learning moderne.

En résumé, MLflow s'est imposé comme un outil incontournable dans l'écosystème des outils de machine learning, offrant une solution complète et polyvalente pour la gestion du cycle de vie des projets, tout en s'intégrant harmonieusement avec d'autres technologies et frameworks existants.

**Avantages de l'utilisation de MLflow par rapport aux approches traditionnelles :**

1. **Centralisation des outils :** Contrairement aux approches traditionnelles qui impliquent souvent l'utilisation de multiples outils disparates pour le suivi, la gestion des modèles et le déploiement, MLflow offre une solution intégrée qui centralise toutes ces fonctionnalités dans une seule plateforme. Cela simplifie la gestion et la collaboration sur les projets de machine learning.
2. **Reproductibilité accrue :** Grâce à son suivi des expériences exhaustif, MLflow facilite la reproductibilité des résultats de machine learning. Les praticiens peuvent facilement reproduire les expériences précédentes, analyser les différences entre les modèles et comprendre les facteurs qui influent sur les performances.
3. **Facilité de déploiement :** MLflow simplifie le processus de déploiement des modèles en offrant des outils intégrés pour emballer les modèles dans des conteneurs reproductibles et les déployer dans divers environnements de production. Cela réduit les obstacles au déploiement des modèles et accélère leur impact dans les applications réelles.

**Les composants clés de MLflow et leurs rôles sont les suivants :**

1. **Suivi des expériences (Tracking)** : MLflow permet de garder une trace des différentes expériences d'entraînement de modèles, en enregistrant les paramètres, métriques et artefacts associés à chaque exécution. Cela permet aux praticiens de comparer facilement les performances des modèles et de retracer les décisions prises lors du développement.
2. **Gestion des modèles (Model Management)** : MLflow facilite le packaging des modèles entraînés dans des conteneurs reproductibles. Il offre un registre centralisé où les utilisateurs peuvent enregistrer, organiser et partager leurs modèles entraînés, ce qui simplifie le processus de déploiement.
3. **Déploiement de modèles (Model Deployment)** : MLflow propose des fonctionnalités pour déployer les modèles entraînés dans divers environnements de production, tels que les serveurs web, les systèmes embarqués ou les plateformes cloud. Cela garantit que les modèles sont mis en service de manière fiable et évolutive.
4. **Évaluation :** Conçu pour une analyse approfondie des modèles, cet ensemble d'outils facilite la comparaison objective des modèles, qu'il s'agisse d'algorithmes ML traditionnels ou de modèles LLM de pointe.
5. **Interface utilisateur d'ingénierie de prompts :** Un environnement dédié à l'ingénierie de prompts, ce composant centré sur l'interface utilisateur offre un espace pour l'expérimentation, l'affinement, l'évaluation, les tests et le déploiement des prompts.
6. **Recettes :** Servant de guide pour structurer les projets ML, les recettes, tout en offrant des recommandations, sont axées sur la garantie de résultats fonctionnels optimisés pour des scénarios de déploiement réels.
7. **Projets :** Les Projets MLflow standardisent l'emballage du code ML, des workflows et des artefacts, similaire à un exécutable. Chaque projet, qu'il s'agisse d'un répertoire avec du code ou d'un référentiel Git, utilise un descripteur ou une convention pour définir ses dépendances et sa méthode d'exécution.

**Le rôle de** MLflow répond aux critères A9 et C20 en fournissant des fonctionnalités et des outils qui facilitent la surveillance et le maintien en condition opérationnelle des applications d'intelligence artificielle (IA). Voici comment MLflow répond à chaque critère :

1. **Définition des métriques pour le monitorage de l'application :** MLflow permet aux utilisateurs de définir et de suivre des métriques personnalisées lors de l'exécution de leurs expériences de machine learning. Ces métriques peuvent inclure des indicateurs de performance tels que la précision du modèle, la perte, le temps d'entraînement, etc.
2. **Définition des seuils ou des valeurs devant générer une alerte :** Les utilisateurs peuvent définir des seuils pour les métriques surveillées dans MLflow. Si une métrique dépasse un seuil prédéfini, des alertes peuvent être déclenchées pour signaler des anomalies ou des problèmes potentiels.
3. **Choix d'une solution ou d'un outil pour la consolidation et le suivi des indicateurs de monitorage :** MLflow offre une solution intégrée pour la consolidation et le suivi des indicateurs de monitorage. Les utilisateurs peuvent visualiser et analyser les métriques enregistrées dans l'interface utilisateur de MLflow.
4. **Configuration de l'outil ou de la solution de monitorage :** MLflow fournit des fonctionnalités permettant de configurer et de personnaliser le suivi des métriques. Les utilisateurs peuvent spécifier les métriques à surveiller, les intervalles de collecte des données, etc.
5. **Intégration de la journalisation nécessaire aux objectifs de monitorage, dans l'application :** MLflow intègre des fonctionnalités de journalisation pour enregistrer les paramètres, les métriques et les artefacts associés à chaque exécution d'une expérience de machine learning. Cette journalisation est automatiquement intégrée lors de l'utilisation de MLflow dans l'application.
6. **Intégration d'alertes (e-mail, push...) en fonction des indicateurs :** Bien que MLflow ne fournisse pas directement des fonctionnalités d'alerte, il peut être intégré à d'autres outils de surveillance qui offrent des fonctionnalités d'alerte, tels que Prometheus, Grafana, ou d'autres solutions tierces.
7. **Documentation du monitorage et des procédures d'installation et de configuration de l'outillage utilisé :** MLflow fournit une documentation détaillée sur la configuration et l'utilisation de ses fonctionnalités de suivi et de journalisation. Cette documentation couvre les métriques, les seuils, les procédures d'installation et de configuration, et est accessible dans un format qui respecte les recommandations d'accessibilité.

**MLflow dans la gestion du cycle de vie des projets de machine learning peut être résumé comme suit :**

* **Organisation et suivi** : MLflow permet aux équipes de machine learning de gérer efficacement leurs projets, en organisant les expériences, les modèles et les résultats associés à chaque itération.
* **Reproductibilité** : En enregistrant tous les éléments de chaque expérience d'entraînement de modèle, MLflow garantit la reproductibilité des résultats, ce qui est essentiel pour la transparence et la validation des modèles.
* **Collaboration** : MLflow fournit un environnement collaboratif où les membres de l'équipe peuvent partager leurs résultats, leurs modèles et leurs meilleures pratiques, favorisant ainsi l'apprentissage et l'amélioration continus.
* **Déploiement efficace** : Avec ses outils de gestion de modèles et de déploiement, MLflow simplifie le processus de mise en production des modèles, accélérant ainsi leur impact dans les applications réelles.

En résumé, MLflow joue un rôle crucial dans la transformation des projets de machine learning en solutions déployées et en production, en offrant une infrastructure robuste pour la gestion complète du cycle de vie des modèles.