

Voici notre projet

Rétro-ingénierie, maintenance et évolution des logiciels

Groupe E - Sujet 1

26/11/2023

Thomas FARINEAU

Léo KITABADJIAN

Ludovic BAILET

Mohamed MAHJOUB

La question générale & pourquoi vous la trouvez intéressante

1. Signification et Impact de MLflow dans le Machine Learning

MLflow transcende la notion d'un simple outil pour devenir un mouvement vers une gestion plus structurée et systématique des processus en machine learning (ML). Cette évolution est essentielle dans un domaine où le ML est fondamentalement expérimental et itératif. La nécessité d'une traçabilité rigoureuse des expériences est cruciale pour comprendre les variations de performance et les ajustements des modèles. MLflow répond à ce besoin en offrant un cadre pour le suivi des expériences, la gestion des modèles, et l'orchestration des workflows, ce qui est indispensable pour gérer les complexités croissantes des projets d'apprentissage automatique.

2. Rationalisation du Développement des Modèles

L'intégration de MLflow dans les projets de ML permet aux organisations de rationaliser le développement de leurs modèles. En facilitant une meilleure reproductibilité, efficacité et collaboration entre les équipes, MLflow transforme les pratiques de ML, les rendant plus systématiques et alignées avec les principes de MLOps. Cela permet de gérer plus efficacement le cycle de vie complet des modèles de ML, depuis leur conception jusqu'à leur déploiement et maintenance, assurant ainsi une meilleure qualité et fiabilité des solutions de ML déployées.

3. Gestion des Expériences avec MLflow

L'étude de la gestion des expériences par MLflow est particulièrement captivante, car elle aborde un défi clé du ML : la gestion systématique et efficace des cycles d'expérimentation. MLflow offre des solutions pour tracer, documenter et comparer les différentes étapes du développement de modèles ML, y compris le suivi des paramètres, des métriques, des résultats et des versions de modèles. Comprendre comment MLflow facilite ces processus permet d'améliorer la reproductibilité et la traçabilité des projets ML et d'optimiser la collaboration entre les équipes de développement et de données.

Une première décomposition en sous-questions et les métriques/KPI/outils envisagés pour y répondre

Pour approfondir la gestion des expériences dans MLflow, nous pouvons décomposer la question principale en sous-questions spécifiques, tout en se concentrant sur des KPI (Key Performance Indicators) et outils spécifiques :

- 1. Stockage des Informations** : Quelles, comment et où les données d'expérience sont-elles stockées dans MLflow ?
KPI/Outils : Capacité de stockage (volume de données gérées), efficacité du stockage (temps d'accès aux données).
- 2. Suivi des métriques et paramètres** : Comment MLflow suit-il les métriques et paramètres des expériences ?
KPI/Outils : Précision du suivi (exactitude des données enregistrées), intégrité des données (cohérence et complétude des données).
- 3. Traçabilité des expériences** : MLflow garantit-il une traçabilité des expériences ?
KPI/Outils : Traçabilité des modifications (capacité à suivre les changements dans les expériences), transparence des processus (clarté dans la documentation des expériences), historique des versions (gestion et accessibilité de l'historique des expériences et des modèles).
- 4. Reproductibilité des expériences** : La reproductibilité des expériences est-elle assurée par MLflow ?
KPI/Outils : Taux de réussite de reproductibilité (pourcentage d'expériences reproduites avec succès), facilité de reconfiguration (temps et effort nécessaires pour répéter une expérience).

Ces sous-questions et KPI visent à évaluer en détail l'efficacité de MLflow dans la gestion des expériences de ML, en se concentrant sur la reproductibilité et la traçabilité, des éléments clés pour les pratiques de MLOps.

La démarche prévue

1. **Exploration initiale de la documentation et des existants** : Nous débuterons par une immersion dans la documentation officielle de MLflow et des articles académiques pour construire une base théorique solide. En parallèle, nous explorerons les contributions de la communauté pour saisir les utilisations réelles et les défis rencontrés par les utilisateurs de MLflow (notamment grâce au topic GitHub disponible).
2. **Développement de cas d'utilisation** : Nous développerons des scénarios d'utilisation reflétant divers aspects de la gestion des expériences dans MLflow, en mettant l'accent sur le stockage des données, le suivi des métriques, et la reproductibilité.
3. **Tests pratiques et évaluation des outils** : Nous confronterons nos résultats aux standards de MLOps pour évaluer l'adéquation de MLflow avec les meilleures pratiques du secteur.
4. **Comparaison avec les standards de MLOps** : Confronter les résultats obtenus avec les standards de MLOps pour évaluer l'alignement de MLflow avec les meilleures pratiques de l'industrie.
5. **Synthèse et recommandations** : Enfin, nous rassemblerons nos découvertes pour obtenir une vue d'ensemble de la gestion des expériences dans MLflow et proposerons des recommandations pour son utilisation efficace dans divers contextes MLOps.

Les sources dont au moins un article sur lesquels vous pensez vous baser pour répondre à la question.

- **Documentation officielle de MLflow** : La documentation officielle est une ressource intéressante pour comprendre en profondeur les fonctionnalités et les mécanismes internes de MLflow. Elle offre des explications détaillées, des guides d'utilisation, et des exemples pratiques.
 - Docs
 - GitHub
- **Articles académiques et de recherche** : Des articles provenant de Moodle et d'autres bases de données académiques fournissent une analyse théorique et des études de cas sur l'utilisation de MLflow, ses meilleures pratiques, et ses spécificités techniques.
 - Moodle:
 - * [ieee_mlflow.pdf](#) (disponible ici)
 - * [deem_mlflow.pdf](#) (disponible ici)
- **Article de blog** : Des articles de blog et publications en ligne offrent des perspectives pratiques, des retours d'expérience, et des analyses sur l'utilisation de MLflow dans différents contextes.
 - Prise en main de MLflow, un outil pour tracer les résultats de vos expériences
 - Gardez une trace de chacune de vos expériences avec MLFlow Tracking
 - MLOps: How MLflow effortlessly tracks your experiments and helps you compare them?)
- **Discussion instantanée** : Interagir avec la communauté MLflow peut fournir des insights précieux, des conseils pratiques, et des retours d'expérience directement des utilisateurs et développeurs de MLflow.
 - Slack

Les sources (codes, données, ...) que vous pensez exploiter.

- Topic MLFlow
- GitHub MLFlow