

Waterflow

Why did the ML engineer bring a parachute to work ? Because they were always prepared for a "cloud" deployment!

Machine Learning Operations



Dans le monde de l'intelligence artificielle et de l'apprentissage automatique, **la création de modèles performants n'est que la première étape d'un processus beaucoup plus vaste**. Une fois qu'un modèle a été développé, il doit être **déployé dans des environnements de production**, où il peut être utilisé pour **prendre des décisions** en temps réel, analyser des données en continu, ou fournir des recommandations personnalisées. C'est là que **le Machine Learning Operations** entre en jeu. Inarrêtable dans votre quête de connaissances, vous vous trouvez une nouvelle obsession, le **ML Ops** ! Vous



vous lancez donc dans un nouveau projet qui se focalise sur **cet ensemble de pratiques** dont vous ne connaissez pas grand-chose.

Drink safe

L'accès à l'eau potable est essentiel à la santé, un droit humain fondamental. Il s'agit d'une **question de santé** et de développement importante aux niveaux national, régional et local. Dans certaines régions, il a été démontré que les **investissements dans l'approvisionnement en eau et l'assainissement peuvent produire un bénéfice économique net**, puisque les réductions des effets néfastes sur la santé et des coûts des soins de santé l'emportent sur les coûts de mise en œuvre des interventions.





Votre sensibilité et votre sens du devoir l'emportent sur vous, **vous décidez de développer un outil pour l'identification d'eau potable à la consommation humaine**. Vous suivez consciencieusement les étapes suivantes :

1. Vous réalisez **une veille sur le Machine Learning Operations** et vous le définissez de façon claire et concise.
2. Vous utilisez un ensemble de données contenant **9 mesures de la qualité de l'eau pour 3276 étendues d'eau différentes**. Récupérez les données d'intérêt [ici](#) ainsi que leur description [ici](#).
3. Comme à votre habitude, **vous pré-traitez** vos données et les préparez à l'emploi. Vous réalisez **une analyse exploratoire des données d'intérêt**.
4. **MLflow** est un outil puissant de gestion du cycle de vie de l'apprentissage automatique. Il permet le suivi, le packaging et le versioning des modèles de Machine Learning, ce qui facilite la reproduction et la comparaison des résultats. **Vous réalisez une veille cet l'outil et l'étendue de son utilisation et vous l'installez**.
5. Vous séparez vos données en ensembles d'entraînement et de validation puis, **vous entraînez un modèle de classification binaire** afin de prédire la qualité de l'eau. Pour la modélisation des données, **deux options s'offrent à vous** :
 - a. vous restez **sur les bonnes vieilles habitudes** en entraînant un algorithme tel que le Random Forest, le XGboost, ou n'importe quel modèle que vous jugez pertinent,



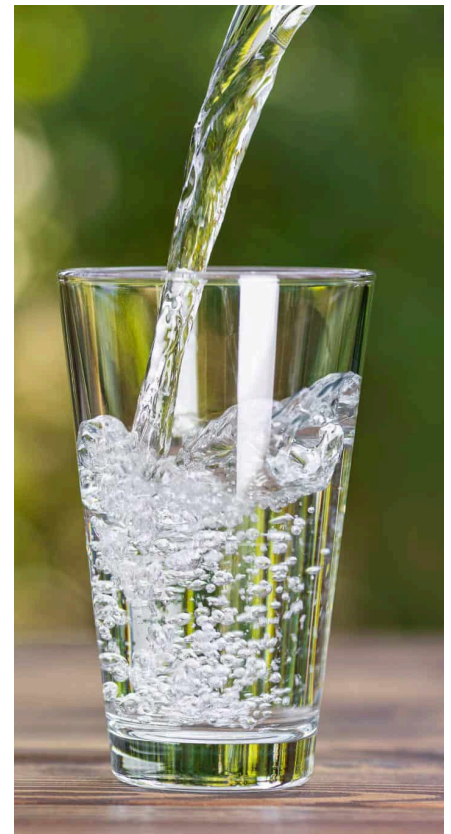
- b. OU vous **vous challengez en utilisant un perceptron multicouches**, un modèle d'apprentissage profond.
6. Vous réalisez des **prédictions sur l'ensemble de validation** et vous calculez une ou plusieurs **métriques d'évaluation**.

Cette partie couvre le suivi des expériences avec MLflow

7. Vous configurez **un serveur MLflow** sur le **port 5000**.
8. Si elle n'existe pas, vous **initialisez une expérience MLflow** avec le nom **"experiment_water_quality"** puis vous récupérez cette expérience.
9. Vous **démarrez l'expérience MLflow** et vous enregistrez le modèle avec ses métadonnées (les paramètres, les métriques d'évaluations, etc).

Cette partie couvre le packaging et déploiement du modèle avec MLflow

10. **Vous ouvrez MLflow UI** depuis le serveur définie précédemment et vous explorez l'interface.
11. **Vous explorez les différentes versions du modèle**. Si vous êtes satisfait de votre modèle, vous pouvez le transitionner depuis **les expériences au Model Registry de MLflow**.
12. Vous **déployez le modèle enregistré** avec MLflow dans une application Flask et créez une API **pour effectuer des prédictions en temps réel**.

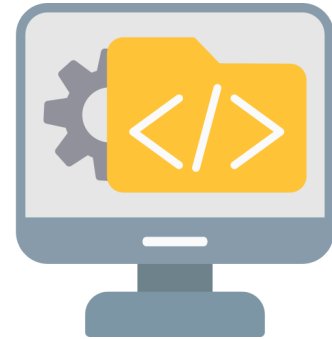




Prêts ? Testez !

Afin d'être sûr que votre **application fonctionne comme prévu**, vous vous penchez sur les tests logiciels. Vous renseignez sur les tests suivants et **vous les appliquez à votre problématique** :

- a. Les tests **unitaires**,
- b. Les tests **fonctionnels**,
- c. Les tests de **non-régression**.



Compétences visées

→ MLOps

Rendu

L'évaluation de ce projet se fera sur deux aspects :

1. Une **présentation explicative de votre travail** sous forme de diapositives. Celle-ci doit raconter la réflexion derrière votre travail et inclure votre **Trello d'organisation**.
2. Un repository github public nommé **waterflow**, contenant les éléments suivants :
 - a. Un notebook Jupyter **propre et commenté** (introduction, conclusion, etc) contenant le procédé de développement de votre outil. **Pensez à répondre à la problématique**.
 - b. Un fichier **experiment.py** contenant les différentes étapes d'utilisation de MLflow.
 - c. Un fichier **app.py** de votre application Flask.



- d. **Votre environnement de tests** : un dossier contenant vos différents scripts de tests à l'aide de la librairie pytest.
- a. Un fichier **README.md** présentant le contexte du projet, les données et leur analyse, l'algorithme utilisé et une conclusion sur votre travail. Pensez à inclure la veille réalisée.

Base de connaissances

- [Machine Learning Operations](#)
- [Getting Started with MLflow](#)
- [MLflow Model Registry](#)
- [MLFlow, une plateforme opensource de Machine Learning pour des projets en Data Science optimisés](#)
- [The Essential Main Ideas of Neural Networks](#)
- [Neural Networks Pt. 2: Backpropagation Main Ideas](#)
- [Qu'est-ce que le test logiciel ?](#)