Projet MLOps — Industrialisation d’un modèle de Machine Learning (M2)

Ce projet vise à vous placer dans une situation réaliste de MLOps. L’objectif est d’industrialiser le cycle de vie d’un modèle de Machine Learning, et non d’optimiser ses performances statistiques.

1. Contexte du projet

Vous rejoignez une équipe Data Platform. Un data scientist a développé un modèle fonctionnel sur son poste local. Votre mission est de rendre ce modèle reproductible, déployable, observable et maintenable dans le temps.

2. Problème métier

- Classification binaire
- Régression
- Détection d’anomalies

3. Artefacts à produire

- Scripts d’entraînement et d’inférence séparés
- Sauvegarde et chargement explicite du modèle
- Reproductibilité via seed fixée

4. Conteneurisation

- Dockerfile pour l’entraînement
- Dockerfile pour l’inférence
- Images légères, sans tag latest
- Exécution en utilisateur non-root

5. Pipeline MLOps

- Validation des données
- Entraînement et évaluation du modèle
- Versioning des modèles et métriques
- Packaging et publication

6. Déploiement

- API HTTP avec endpoint /predict
- Endpoint /health

- Chargement du modèle au démarrage

7. Observabilité

- Logs structurés
- Métriques simples : latence, volume
- Bonus : détection de dérive

8. Sécurité et fiabilité

- Validation des entrées
- Gestion propre des variables d’environnement
- Pas de données sensibles dans les images

9. Scénario incident

Un incident est simulé (dérive des données, modèle obsolète ou performance dégradée). Vous devez analyser la situation, proposer une remédiation et éviter la régression.

10. Documentation attendue

Un README décrivant l’architecture, le pipeline MLOps, la procédure d’entraînement, le déploiement, la gestion des versions et les choix techniques.

L’évaluation porte sur la qualité de l’industrialisation MLOps, la reproductibilité, la fiabilité du système et la clarté de la documentation.