

Machine Learning Operations (ML OPS)

Overview

Machine Learning Operations (ML OPS) is a critical discipline within our organisation that focuses on streamlining the deployment, monitoring, and management of machine learning models. By implementing ML OPS practices, we ensure that our models are reliable, scalable, and maintainable across their lifecycle. This page outlines our ML OPS process using Databricks workspaces, supported by Databricks Unity Catalog and Azure DevOps.

Our deployment of ML OPS has been reviewed by Databricks for best practice  [MLOps Architecture Review - Three UK.pdf](#)

- Repo: [Unit test demo.py](#) - [Repos \(azure.com\)](#)
- Documentation: [README.md](#) - [Repos \(azure.com\)](#)

Environment Setup

Workspaces

We utilise three distinct Databricks workspaces to manage the lifecycle of our machine learning models:

1. **PPD (Development)**: This is the development environment where data scientists perform data understanding, model development, and initial evaluation.
2. **PPS (Testing)**: This environment is used for testing purposes, including unit tests, integration tests, and validation.
3. **PRD (Production)**: This is the production environment where validated models are deployed and monitored.

Unity Catalog

Databricks Unity Catalog supports each workspace, providing a unified governance solution for data and AI. It ensures consistent data policies and governance across all stages of the ML lifecycle.

Azure DevOps

We use Azure DevOps to manage our projects, with each project organised as a mono repo. This setup allows for streamlined version control, collaboration, and automation throughout the ML OPS process.

ML OPS Process

1. Development in PPD (Dev)

- **Data Understanding:** Data scientists begin by exploring and understanding the data available in the development environment.
- **Modeling:** Once familiar with the data, they develop machine learning models using various techniques and tools available in Databricks.
- **Evaluation:** Initial evaluation of the models is conducted to ensure they meet the basic performance criteria.
- **Version Control:** All code, including notebooks, scripts, and configurations, are managed in Azure DevOps repositories.

2. Testing in PPS (Test)

- **Notebook Promotion:** After a model is developed and evaluated in the development environment, the corresponding notebooks are promoted to the testing environment via Azure DevOps pipelines.
- **Unit Tests:** Unit tests are executed to validate individual components of the model.
- **Integration Tests:** Integration tests are performed to ensure that the model works well with other systems and components.
- **Validation:** Comprehensive validation is done to ensure that the model performs as expected under different scenarios.
- **Continuous Integration (CI):** Automated CI pipelines in Azure DevOps ensure that all tests are run on every commit to the repository.

3. Deployment in PRD (Prod)

- **Model Promotion:** Once a model passes all tests in the testing environment, it is promoted to the production environment using Azure DevOps release pipelines.
- **Deployment:** The model is deployed to production where it becomes available for use in real-time applications.
- **Monitoring:** Continuous monitoring of the deployed model is conducted to track its performance and detect any anomalies or drifts.
- **Continuous Deployment (CD):** Automated CD pipelines manage the deployment process, ensuring consistent and reliable delivery of models to production.

Benefits of ML OPS

Implementing ML OPS offers several benefits:

- **Scalability:** Ensures models can handle increasing volumes of data and requests.
- **Reliability:** Reduces the risk of model failures and ensures consistent performance.
- **Maintainability:** Simplifies the process of updating models and addressing issues.

- **Efficiency:** Streamlines the workflow from development to deployment, reducing time to market.
- **Governance:** Provides a framework for managing and auditing models, ensuring compliance with policies and regulations.
- **Collaboration:** Azure DevOps enhances collaboration by providing tools for version control, project management, and automated workflows.

Training resources

ML Ops overview:  Data Science - ML Ops Implementation.pptx

Version Control Introduction:  Data Science - VCS Training.pptx

Unit Testing Introduction:  Data Science - Unit testing training.pptx