

MLOps and Model Deployment

Duration: 5 Days

Language: en

Course Code: PI2-110

Objective

By the end of this course, participants will be able to:

- Understand the key principles and goals of MLOps.
- Design scalable ML workflows from training to deployment.
- Implement CI/CD pipelines for machine learning models.
- Choose appropriate deployment strategies based on business needs.
- Monitor model performance and detect concept/data drift.
- Automate retraining and version control in production environments.
- Collaborate across cross-functional teams for continuous delivery of ML solutions.

Audience

This course is ideal for:

- Machine learning engineers and data scientists.
- DevOps and software engineers working with AI solutions.
- Technical project leads and MLOps professionals.
- IT operations and infrastructure teams supporting ML initiatives.
- Product managers and technical consultants in AI-driven projects.
- Developers transitioning into AI operations roles.

Training Methodology

This course combines technical instruction with hands-on labs, code walkthroughs, and real-world case studies. Participants will build sample CI/CD pipelines, deploy models using open-source tools, and practice monitoring and updating production models. Emphasis is placed on team collaboration, infrastructure scalability, and reproducibility of results.

Summary

As machine learning (ML) continues to evolve from experimental research into large-scale production, the need for robust deployment, monitoring, and lifecycle management has never been greater. This course introduces participants to MLOps—a set of practices that bridge the gap between data science and operations to ensure models are not only accurate but also scalable, reliable, and maintainable in real-world environments.

Participants will explore the end-to-end ML workflow with a focus on deployment strategies, CI/CD pipelines, infrastructure automation, and model governance. The course emphasizes collaboration between data scientists, engineers, and DevOps teams to build production-ready ML systems that are efficient, traceable, and adaptable.

Course Content & Outline

Section 1: Introduction to MLOps and Production ML

- What is MLOps? Definitions, components, and business value.
- Comparison of traditional DevOps vs MLOps.
- The ML lifecycle: from data ingestion to retraining.
- Challenges in deploying ML models at scale.
- Benefits of automation and monitoring in ML workflows.
- Key tools and platforms in the MLOps ecosystem (MLflow, Kubeflow, DVC, etc.).
- Case study: MLOps in a real-world product environment.

Section 2: Building and Automating ML Pipelines

- Designing reproducible ML workflows using pipeline tools.
- Data validation, feature engineering, and model training pipelines.
- Experiment tracking and parameter logging.
- Version control for datasets, code, and models.
- Using MLflow or DVC for tracking and reproducibility.
- Containerization using Docker for ML applications.
- Best practices for modular pipeline design.

Section 3: Model Deployment Strategies and Infrastructure

- Deployment types: batch, real-time (online), edge, and hybrid.
- Model serving tools: TensorFlow Serving, TorchServe, FastAPI, BentoML.
- API integration and microservice architecture.
- Model packaging and container orchestration with Kubernetes.
- Choosing the right cloud or on-prem environment.
- Hands-on example: deploying a model via REST API.
- Rollout strategies: blue-green, canary, and shadow deployments.

Section 4: Monitoring, Testing, and Model Governance

- Importance of monitoring models in production.
- Key metrics: latency, prediction accuracy, drift detection.
- Automating testing: unit tests, integration tests for ML models.
- Alerting and rollback strategies for failed deployments.
- Model governance: auditing, compliance, and reproducibility.
- Toolkits for drift detection and quality assurance.
- Logging and observability frameworks for ML services.

Section 5: CI/CD for Machine Learning

- Continuous integration and delivery pipelines for ML.
- Integrating Git, Jenkins, GitHub Actions, or GitLab CI.
- Automating retraining and redeployment on new data.
- Scheduling and orchestration using Airflow or Prefect.
- Building end-to-end automated workflows.
- Cross-functional collaboration and workflow ownership.
- Scaling MLOps for multiple models and business use cases.

Certificate Description

Upon successful completion of this training course, delegates will be awarded a Holistique Training Certificate of Completion. For those who attend and complete the online training course, a Holistique Training e-Certificate will be provided.

Holistique Training Certificates are accredited by The CPD Certification Service (CPD), and are certified under ISO 9001 and ISO 29993 standards.

CPD credits for this course are granted by our Certificates and will be reflected on the Holistique Training Certificate of Completion. In accordance with the standards of The CPD Certification Service, one CPD credit is awarded per hour of course attendance. A maximum of 50 CPD credits can be claimed for any single course we currently offer.

Categories

AI, Data and Visualisation, IT & Computer Application, Technology

Tags

MLOps, Model Deployment

Related Articles



Top 15 Skills Every Data Scientist Needs in 2025

Unlock the top 15 data scientist skills for 2025—from tech to soft skills—plus expert tips to help you grow, adapt, and thrive in a data-driven world.