

MLOps Fundamentals

Duration: 3 Days **Course Code: GK840034** **Delivery Method: Virtual Learning**

Overview:

MLOps Fundamentals is a comprehensive guide to the principles, components, and tools used in Machine Learning Operations (MLOps). It provides a thorough understanding of the machine learning lifecycle, MLOps lifecycle, and the benefits and tools involved, such as MLFlow and KubeFlow.

We will take a look at setting up an ML project, including using Git and GitHub, setting up virtual environments, and pre-commit hooks. The course will then delve into the fundamentals of data management, such as understanding data lifecycles, data versioning, governance, and storage solutions.

Practical, hands-on demonstrations will be provided on Exploratory Data Analysis (EDA), feature engineering, and data cleaning using pandas and matplotlib. The course will further explore the concept of feature stores, their types, working, best practices, and implementation challenges.

Virtual Learning

This interactive training can be taken from any location, your office or home and is delivered by a trainer. This training does not have any delegates in the class with the instructor, since all delegates are virtually connected. Virtual delegates do not travel to this course, Global Knowledge will send you all the information needed before the start of the course and you can test the logins.

Target Audience:

ML practitioners looking to make the leap from toy ML demos to productionized ML applications

- Data Scientists
- Developers
- Software Engineers

Objectives:

- Explain the core principles and challenges of MLOps in the context of the ML project lifecycle
- Design and structure ML projects using best practices and industry-standard tools
- Analyze and prepare datasets for machine learning applications using advanced data management techniques
- Develop and evaluate machine learning models using appropriate metrics and experiment tracking tools
- Apply version control and containerization techniques to ensure reproducibility in ML projects
- Implement comprehensive testing strategies for ML components and pipelines
- Deploy ML models using RESTful APIs and containerization technologies
- Construct automated CI/CD pipelines for ML projects
- Design and implement monitoring systems for deployed ML models
- Evaluate and address model drift and performance degradation in production environments
- Integrate data engineering practices into MLOps workflows
- Create an end-to-end MLOps pipeline incorporating all learned concepts

Prerequisites:

Proficiency in Python; strong beginner/intermediate grasp of ML, familiarity with Git and version control, experience with cloud platforms

Prerequisite Course:

Content:

1- Introduction to MLOps

- What is MLOps
- Machine Learning Life Cycle Overview

2- MLOps Components and Tools

- Brief overview of MLOps Life Cycle / Components of MLOps and Benefits
- Brief Overview of MLOps tools (MLFlow, KubeFlow, etc) and their role in automating ML Pipelines

4- Setting up an ML Project

- Git and GitHub Setup
- Setting Up Virtual Environments
- Pre-commit Hooks

5- Data Management Fundamentals

- Understanding Data Lifecycles
- Data Versioning
- Data Governance
- Data Storage Solutions

6- Demo: EDA, Feature Engineering, and Data Cleaning

- Hands-on EDA using pandas to summarize the dataset.
- Visualizing distributions using matplotlib (histograms, scatter plots).
- Creating new features and cleaning data by removing missing values and outliers.

7- Feature Stores

- Introduction to Feature Stores
- Types of Feature Stores
- How Feature Stores Work
- Best Practices for Using Feature Stores
- Challenges in Implementing Feature Stores

8- Model Development

- Overview of Model Development Process
- Choosing the Right Algorithm
- Model Training and Validation
- Avoiding Overfitting
- Model Evaluation Metrics

9- Implementing a Basic ML Pipeline

- Building the Pipeline
- Integrating Preprocessing and Model Development
- Training and Evaluating the Pipeline
- Introduction to Pipeline Automation

10- Model Development Strategies

- Overview of Model Development

11- ML Model Interpretability and Explainability

- Introduction to Model Interpretability and Explainability
- Techniques for Model Interpretability
- Explainability in Different Model Types
- Tools for Interpretability
- Challenges in Explainability

12- Implementing Algorithms

- Selecting an Algorithm
- Implementing the Chosen Algorithm
- Evaluating Algorithm Performance
- Comparing Multiple Algorithms

13- Demo: Selecting, Implementing, and Evaluating Algorithms

- Select a dataset, choose two different algorithms (e.g., Decision Tree and SVM)
- Implement the algorithms using scikit-learn
- Evaluate the performance of each algorithm
- Compare the results using metrics like accuracy, precision, etc

14- Experiment Tracking and Model Evaluation

- Introduction to Experiment Tracking
- Setting Up Experiment Tracking
- Evaluating Model Performance
- Visualizing Model Performance

15- Setting Up MLflow for Experiment Tracking

- Introduction to Mlflow
- Tracking Experiments with Mlflow
- Comparing Multiple Runs
- Storing and Retrieving Models

16- Evaluating Models

- Preparing the Evaluation Environment
- Evaluating Model Performance
- Comparing Models Based on Evaluation"

17- Hyperparameter Tuning Techniques

- Introduction to Hyperparameter Tuning
- Grid Search vs. Random Search
- Bayesian Optimization
- Practical Considerations"

18- Automated Hyperparameter Tuning

- Introduction to Automated Hyperparameter Tuning
- Running Hyperparameter Tuning

20- Legal and Compliance issues in MLOps

- Introduction to Legal and Compliance in MLOps
- Key Regulatory Standards
- Model Governance and Compliance
- Challenges in Legal and Compliance Issues

21- Containerizing ML Models with Docker

- Introduction to Docker
- Setting Up Docker
- Building a Docker Image
- Deploying Docker Containers on Cloud Platforms

22- Deploying Models to Cloud Platforms

- Introduction to Cloud Deployment
- Preparing the Model for Deployment
- Setting Up Cloud Infrastructure
- Deploying the Model with Ray Serve

23- Federated Training and Edge Deployments

- Introduction to Federated Learning and Edge Computing
- Federated Training Architecture
- Edge Model Deployment
- Tools and Frameworks
- Challenges in Federated Learning and Edge Computing

24- CI/CD for ML

- Introduction to CI/CD for Machine Learning
- Setting Up CI/CD Pipelines for ML
- Integrating CI/CD with Experiment Tracking
- Automating Model Validation and Testing

25- Setting up CI/CD Pipelines for ML

- Introduction to GitHub Actions for CI/CD
- Automating Model Training and Deployment
- Integrating MLflow with CI/CD
- Testing the CI/CD Pipeline

26- Monitoring and Maintaining ML Systems

- Introduction to Monitoring ML Systems
- Tools for Monitoring ML Models
- Setting Up Alerts for Model Drift
- Monitoring Model Performance in Real-Time
- Continuous Feedback Loops
- Scaling Monitoring for Large-Scale Deployments

27- Implementing Monitoring Tools

Approaches	■ Analyzing the Results"	
■ Data-Centric vs. Model-Centric Approaches	19- Model Serving and Deployment Strategies	■ Introduction to Monitoring Tools
■ Experimentation in Model Development	■ Introduction to Model Serving	■ Instrumenting the ML Model for Monitoring
■ Collaborative Development in MLOps	■ Deployment Strategies	■ Code Implementation - Exposing Metrics for Prometheus
	■ Containerization of ML Models	■ Visualizing Metrics in Grafana
	■ Serving Models with Docker	
	■ Model Serving Frameworks	
	■ Deploying Models on Cloud Platforms	

Further Information:

For More information, or to book your course, please call us on Head Office 01189 123456 / Northern Office 0113 242 5931

info@globalknowledge.co.uk

www.globalknowledge.com/en-gb/

Global Knowledge, Mulberry Business Park, Fishponds Road, Wokingham Berkshire RG41 2GY UK