KOENIG SOLUTIONS · COURSE OUTLINE

Data Science: Scikit-learn, MLflow & MLOps

Learn to build, evaluate and deploy machine learning models with scikit-learn, MLflow and MLOps tools across the full project lifecycle.

56 Hours	24	127	5
DURATION	MODULES	TOPICS	HANDS-ON ACTIVITIES

COURSE OVERVIEW

The course is a comprehensive program designed to equip learners with a deep understanding of data science and machine learning concepts. It is structured in various modules, starting with an Introduction to Data Science & Machine Learning, covering the essentials such as Analytics types, Project lifecycle, and required skills. The course then delves into practical skills with Python for Data Analysis & Preprocessing, teaching the use of popular libraries and data handling techniques. Subsequent modules focus on Supervised Machine Learning for both Regression and Classification, where learners gain hands-on experience with models like linear Regression, Logistic Regression, SVMs, Decision trees, and more. The course emphasizes the importance of Feature Selection and Dimensionality Reduction, Cross-Validation & Hyperparameter Tuning, and introduces Deep Learning fundamentals. Additionally, learners explore Clustering techniques to uncover patterns in data. By the end of the course, participants will have mastered the key concepts and tools necessary for a career in machine learning, including Python programming, Data preprocessing, Model evaluation, and advanced algorithms. This course offers a blend of theoretical knowledge and practical application, ensuring learners are well-prepared for real-world data science challenges.

WHO THIS COURSE IS FOR

Who Should Attend

This course by Koenig Solutions is designed for professionals seeking advanced knowledge in data science and machine learning techniques.

Data Scientists – Looking to deepen their expertise in machine learning algorithms and techniques

Machine Learning Engineers – Seeking to enhance their skills in building and deploying ML models

Data Analysts – Aiming to leverage ML for data-driven insights and business intelligence

Software Developers – Interested in incorporating machine learning into applications and systems

IT Professionals – Exploring a transition into data science and machine learning roles

Statisticians – Wanting to implement and optimize machine learning models in their work

Business Analysts – Seeking to understand data-driven decision-making through ML

Research Scientists – Utilizing machine learning for academic and industry research

Graduate Students – Pursuing careers in computer science, data science, and related fields

AI Enthusiasts – Passionate individuals eager to explore and apply machine learning concepts

Product Managers – Looking to leverage machine learning in product development and innovation

Technical Managers – Leading data-driven projects and managing ML implementation

LEARNING OBJECTIVES

- Explain core data science and machine learning concepts and lifecycle
- Apply Python libraries for data analysis and pre-processing
- Build regression models including simple and multiple linear regression
- Implement classification models such as logistic regression, SVM and decision trees
- Apply feature selection and dimensionality reduction techniques like PCA
- Evaluate models using cross-validation and hyperparameter tuning
- Develop clustering solutions using K-Means
- Deploy ML models through APIs, containers and CI/CD pipelines

COURSE SYLLABUS

24 Modules, 127 Topics

01

Introduction to Data Science & Machine Learning

TOPICS

- The Need for Data Science and Machine Learning
- Types of Analytics
- The Lifecycle of a Data Science Project
- Essential Skills for a Data Scientist Role
- Types of Machine Learning

02

Python for Data Analysis & Pre-processing

TOPICS

- Python Libraries – NumPy, Pandas, Matplotlib, Seaborn, scikit-learn, TensorFlow, Keras, PyTorch Exploratory Data Analysis (EDA)
- Overview of EDA
- Data Cleaning Techniques – Handling Missing and Categorical Data
- Visualizations: 2D Scatter Plot, 3D Scatter Plot, Pair Plots
- Univariate, Bivariate, and Multivariate Analysis, Box Plots Data Pre-Processing
- Importance of Data Pre-Processing
- Handling Missing Values
- Label Encoding for Categorical Data
- One-Hot Encoding Explained Data Transformation
- The Need for Data Transformation
- Introduction to Data Normalization
- Normalization Techniques – Standard Scaler & Min-Max Scaler
- **Data Splitting:** Train, Test, and Validation

03

Supervised Machine Learning – Regression

TOPICS

- Simple Linear Regression
- Introduction to Linear Regression
- Ordinary Least Squares and Regression Errors
- **Data Processing:** Train-Test Split
- Model Evaluation Metrics – R-Squared, RMSE, Score, and Interpretation
- Prediction Plot and Its Interpretation
- Multiple Linear Regression
- Concept of Multiple Linear Regression
- Degrees of Freedom and Adjusted R-Squared
- Assumptions of Multiple Linear Regression – Linearity, Multicollinearity, Autocorrelation, Endogeneity, Normality of Residuals, Homoscedasticity
- Understanding Time-Lag Data in Autocorrelation
- The Dummy Variable Trap

04

Supervised Machine Learning – Classification

TOPICS

- Logistic Regression
- Introduction to Logistic Regression
- Understanding Stratification
- The Confusion Matrix Explained
- Support Vector Machine (SVM)
- Intuitive Understanding of SVM
- Mathematical Explanation of SVM
- Types of SVM Kernel Functions
- IRIS Classification Problem - Exercise
- Decision Tree Classifier
- Introduction to Decision Trees
- Optimal Model Selection Criteria for Decision Trees
- Random Forest Classifier
- Introduction to Ensemble Learning and Random Forests
- Bagging vs Boosting Techniques
- Evaluation Metrics for Classification Models
- Importance of Model Evaluation and the Accuracy Paradox
- Key Metrics – Accuracy, Precision, Recall, F1 Score
- Adjusting Threshold Values
- AUC-ROC Curve Analysis

05

Feature Selection and Dimensionality Reduction

TOPICS

- Univariate Feature Selection
- Importance of Feature Selection
- Overview of Univariate Feature Selection
- F-Test for Regression and Classification
- Hands-on F-Test (P-value Analysis)
- Chi-Square Test for Classification
- Feature Selection Techniques – SelectKBest, SelectPercentile & Generic Univariate Select
- Hands-on Chi-Square Test (P-value Analysis)
- Recursive Feature Elimination (RFE)
- Introduction to RFE
- Feature Importance Scores and Ranking
- Hands-on RFE
- Principal Component Analysis (PCA)
- The Need for Dimensionality Reduction and Importance of PCA
- Mathematical Concepts and Steps to Perform PCA
- Hands-on PCA (Comparing Models With and Without PCA)

06

Cross-Validation & Hyperparameter Tuning

TOPICS

- Cross-Validation
- Importance of Cross-Validation
- Parameters and Implementation of Cross-Validation
- Hands-on Exercise and Result Interpretation
- Hyperparameter Tuning
- Introduction to Hyperparameter Tuning
- Grid Search vs Randomized Search
- Hands-on GridSearchCV (Analyzing Results)

07

Supervised Machine Learning – Natural Language Processing (NLP)

TOPICS

- Introduction to NLP
- Core Concepts – Tokenization, Stop Words, Stemming, Lemmatization
- TF-IDF Vectorization and Its Mathematical Intuition
- Building a Recommendation System Using NLP

08

Unsupervised Machine Learning – Clustering

TOPICS

- Introduction to Clustering
- Mathematical Intuition Behind Clustering
- The Elbow Method and Its Mathematical Explanation
- K-Means Clustering Implementation (Numerical Data)
- K-Means Clustering Implementation (Text Data Processing)

09

Introduction to Deep Learning

TOPICS

- The Need for Deep Learning and Its Applications
- Working of Artificial Neural Networks (ANNs)
- Understanding Backend (TensorFlow) and Frontend (Keras)
- Concept of Tensors
- Overview of Keras Model Building – Construct, Compile & Evaluate
- Activation Functions Explained
- Loss Functions Overview
- Optimization Techniques in Deep Learning
- Evaluation Metrics for Deep Learning Models

10

Machine Learning Ops Fundamentals

TOPICS

Introduction to MLOps and its significance · Challenges in traditional ML model management · Solutions offered by MLOps

11

MLOps Toolbox

TOPICS

- Applying MLOps tools for end-to-end projects
- **Integration of tools:** DVC, Git, MLFlow, and DagsHub

12

Model Versioning with MLFlow

TOPICS

- Versioning and registering ML models with MLFlow
- MLFlow's role in managing ML lifecycle

13

Data Versioning with DVC

TOPICS

Capturing data and model versions with DVC · On-premises and cloud storage integration

14

Creating Shared ML Repository

TOPICS

- Utilizing DagsHub, DVC, Git, and MLFlow for versioning
- Collaborative ML model management

15

Auto-ML and Low-Code MLOps

TOPICS

- Automation of ML model development with Auto-ML and Pycaret
- Streamlining model versioning, training, evaluation, and deployment

16

Explainability and Auditability

TOPICS

- Understanding model interpretability and explainability
- Monitoring model performance and data drift with SHAP and Evidently

17

Containerized ML Workflow with Docker

TOPICS

- Packaging code and dependencies using Docker
- Efficient distribution of Machine Learning applications

18

Deploying ML via APIs

TOPICS

- Model deployment through API development with FastAPI and Flask
- Deploying APIs on Azure Cloud using containers

19

Deploying ML in Web Applications

TOPICS

- Developing web apps with embedded ML models using Gradio and Flask
- Deploying to production in Azure via Docker containers

20

Automated ML Services with BentoML

TOPICS

- Introduction to BentoML and its role in automated ML service development
- Putting BentoML services into production using Docker
- Integration of BentoML with MLFlow

21

Practical Lab

CI/CD with GitHub Actions and CML

TOPICS

- Introduction to GitHub Actions and Continuous Machine Learning (CML)

22

Model Monitoring with Evidently AI

TOPICS

- Monitoring models and services using Evidently AI
- Identifying data drift and evaluating model quality

23

4 Hands-on Experiences

Model Monitoring with Deepchecks

TOPICS

- **Components of Deepchecks:** checks, conditions, and suites

24

Complete MLOps Project

TOPICS

- Developing an ML model from scratch
- Validating code and preprocessing data
- Versioning with MLFlow and DVC
- Sharing repository with DagsHub and MLFlow
- Building an API with BentoML
- Creating a Streamlit app
- Implementing CI/CD with GitHub Actions

HANDS-ON ACTIVITIES

Practise What You Learn

#	MODULE	ACTIVITY
21	**CI/CD with GitHub Actions and CML**	**Practical Lab:** GitHub Actions for MLOps CI/CD
23	**Model Monitoring with Deepchecks**	HANDS-ON EXPERIENCE - Data Integrity Suite - Train Test Validation Suite - Model Evaluation Suite - Custom Performance Suite

WHAT IS INCLUDED

Everything You Need to Succeed

- Koenig course material for every module
- Delivery by certified instructors and subject matter experts
- Hands-on exercises throughout the course
- Course completion certificate from Koenig

ABOUT KOENIG

Who We Are

Established in 1993, Koenig Solutions is a globally recognised training provider with 35+ technology partnerships, 5,000+ training programmes and a network of certified instructors and subject matter experts, supporting individuals and enterprise teams worldwide.

FLEXIBLE DELIVERY

Learn the Way That Suits You

Classroom Training | Live Online Classes (ILO) | Fly-Me-A-Trainer (FMAT) | Flexi (Self-Paced) | 1-on-1 Training

Ready to Get Started?

Talk to a learning advisor about schedules, group training and customised delivery.

Visit koenig-solutions.com | info@koenig-solutions.com