Data Science with Python

Professional Course Outline

Course Overview

This course provides a comprehensive introduction to Data Science using Python. It is designed to equip students with the practical skills required to collect, clean, analyze, visualize, and model data. The course combines statistical concepts, programming skills, and machine learning techniques to help learners solve real-world data problems. Students will work with real datasets and complete projects that simulate industry-level data science workflows.

Target Audience

- Students interested in data science and analytics
- Aspiring data analysts and data scientists
- Professionals who want to transition into data-driven roles
- Anyone with basic programming knowledge who wants to work with data

Course Objectives

- Understand the data science workflow
- Develop Python skills for data analysis
- Perform data cleaning and preprocessing
- Create meaningful visualizations from data
- Apply machine learning algorithms for prediction and classification
- Interpret and communicate insights from data
- Build real-world data science projects

Course Modules

Module 1: Introduction to Data Science

- What is Data Science
- Applications of Data Science
- Data Science lifecycle
- Roles in data science (Analyst, Scientist, Engineer)
- Tools used in data science

Module 2: Python for Data Science

- Python basics for data science
- Working with Jupyter Notebook
- Python libraries overview
- NumPy fundamentals
- Python data structures for data analysis

Module 3: Data Collection and Data Sources

- Types of data
- Structured vs unstructured data
- Data collection methods
- Working with CSV, Excel, and JSON
- Introduction to APIs

Module 4: Data Cleaning and Preprocessing

- Handling missing values
- Removing duplicates
- Data transformation
- Feature engineering basics
- Data normalization and scaling

Module 5: Data Analysis with Pandas

- DataFrames and Series
- Filtering and selecting data
- Grouping and aggregation
- Data merging and joining
- Exploratory data analysis

Module 6: Data Visualization

- Importance of visualization
- Matplotlib basics
- Seaborn for statistical visualization
- Creating dashboards and plots
- Visual storytelling with data

Module 7: Statistics for Data Science

- Descriptive statistics
- Probability basics
- Distributions
- Hypothesis testing
- Correlation and regression concepts

Module 8: Machine Learning Fundamentals

- Introduction to machine learning

- Supervised vs unsupervised learning
- Training and testing models
- Feature selection
- Model evaluation concepts

Module 9: Supervised Learning

- Linear regression
- Logistic regression
- Decision trees
- Random forest
- Model performance evaluation

Module 10: Unsupervised Learning

- Clustering concepts
- K-Means clustering
- Hierarchical clustering
- Dimensionality reduction
- Principal Component Analysis (PCA)

Module 11: Model Evaluation and Optimization

- Accuracy, precision, recall
- F1 score
- Confusion matrix
- Cross-validation
- Hyperparameter tuning

Module 12: Data Science Project

- Problem definition
- Dataset selection
- Data preprocessing
- Model building
- Result interpretation and presentation

Course Outcomes

- Ability to analyze and interpret real-world datasets
- Strong foundation in Python for data analysis
- Understanding of machine learning concepts
- Experience working with real data science workflows
- Completion of an end-to-end data science project

Tools & Technologies Covered

- Python

- Jupyter Notebook
- NumPy
- Pandas
- Matplotlib
- Seaborn
- Scikit-learn
- VS Code