Data Science with Python

Duration: 5 Days

Module 1: Introduction to Data Science and Python

- Overview of Data Science, its lifecycle and applications
- Data Science workflow and problem-solving approach
- Python environment setup using Jupyter Notebook / Google Colab / VS code
- Python fundamentals for Data Science
- Data structures, functions, loops and modules
- Working with Python libraries and packages

Module 2: Data Manipulation and Analysis Using NumPy and Pandas

- Introduction to NumPy and array operations
- Working with Pandas Series and DataFrames
- Importing and exporting datasets (CSV, Excel, JSON)
- Data selection, filtering, sorting and aggregation
- Handling missing values and duplicate records
- Data transformation, merging and grouping

Module 3: Data Cleaning and Exploratory Data Analysis (EDA)

- Understanding data types and data quality
- Data cleaning and preprocessing techniques
- Identifying outliers and handling inconsistent data
- Descriptive and summary statistics
- Univariate, bivariate and multivariate analysis
- Correlation and relationship analysis
- Hands-on exploratory data analysis using real-world datasets

Module 4: Data Visualization Using Python

- Introduction to Matplotlib and Seaborn
- Creating line charts, bar charts and histograms
- Scatter plots, box plots and heatmaps

- Distribution and categorical data visualization
- Customizing charts, labels, legends and layouts
- Visual storytelling and communicating data insights

Module 5: Statistics and Mathematics for Data Science

- Descriptive statistics: Mean, Median, Mode and Variance
- Measures of dispersion and standard deviation
- Probability concepts and distributions
- Sampling techniques and sampling distributions
- Hypothesis testing and confidence intervals
- Correlation and regression fundamentals

Module 6: Introduction to Machine Learning

- Overview of Machine Learning and its applications
- Types of Machine Learning: Supervised and Unsupervised
- Machine Learning workflow and model development
- Introduction to Scikit-learn
- Feature selection and feature engineering
- Training and testing datasets
- Data preprocessing and feature scaling

Module 7: Supervised Machine Learning Algorithms

- Linear and Multiple Linear Regression
- Logistic Regression
- Decision Trees and Random Forest
- K-Nearest Neighbors (KNN)
- Support Vector Machines (SVM)
- Classification and Regression use cases
- Model training and prediction using Scikit-learn

Module 8: Unsupervised Machine Learning and Model Evaluation

- Introduction to clustering techniques
- K-Means Clustering
- Hierarchical Clustering
- Principal Component Analysis (PCA)

- Model evaluation metrics for regression and classification
- Confusion Matrix, Accuracy, Precision, Recall and F1-score
- Cross-validation and hyperparameter tuning
- Model performance comparison and improvement

Module 9: End-to-End Data Science Project

- Understanding business problems and defining objectives
- Dataset collection and preparation
- Data cleaning and exploratory data analysis
- Data visualization and insight generation
- Feature engineering and model selection
- Building and evaluating a machine learning model
- Generating predictions and communicating findings
- Presenting the end-to-end Data Science solution