Complete Data Science Course Details

Preface

In today's rapidly evolving technological landscape, data has emerged as one of the most valuable assets for businesses and organizations across the globe. The IT industry, in particular, has been revolutionized by the application of data science, enabling companies to unlock actionable insights, drive innovation, and enhance decision-making processes.

Data science empowers IT professionals to turn vast amounts of raw data into meaningful information, facilitating advancements in automation, artificial intelligence, cybersecurity, cloud computing, and beyond.

Course Objective

To equip learners with the fundamental and advanced skills required to analyse data, build predictive models, and derive actionable insights using Python.

Course Outline

Introduction to Data Science

- What is Data Science?
- Importance and applications in real-world scenarios
- Data Science lifecycle
- Role of Python in Data Science

Python Programming for Data Science

- Python Basics: Variables, Data Types, Operators
- Control Structures: If, Else, Loops
- Functions and Modules
- Working with Python libraries (pip, virtual environments)

Data Structures and File Handling

Lists, Tuples, Dictionaries, Sets

List comprehensions

Reading/writing files (CSV, Excel, JSON)

Directory and file operations using os and glob

Data Analysis with NumPy and Pandas

Introduction to NumPy: Arrays, Vectorized operations

Pandas: Series and DataFrames

Data selection, filtering, sorting

Handling missing data, duplicates

Grouping and aggregations

Merging and joining datasets

Data Visualization

Matplotlib: Line, Bar, Scatter, Pie charts

Seaborn: Heatmaps, Boxplots, Pair plots

Customizing visualizations

Plotly: Interactive plots

Exploratory Data Analysis (EDA)

Understanding distributions and relationships

Outlier detection and handling

Correlation and covariance

Feature exploration techniques

Statistics and Probability for Data Science

Descriptive statistics: Mean, Median, Mode, Variance, SD

Probability distributions: Normal, Binomial

Inferential statistics: Hypothesis testing, t-tests, chi-square

Confidence intervals and p-values

Introduction to Machine Learning

Supervised vs. Unsupervised learning

Train/test split and cross-validation

Linear Regression and Logistic Regression

Decision Trees, Random Forests, KNN, SVM

Clustering: K-Means, Hierarchical clustering

Feature Engineering and Model Evaluation

Feature scaling: Normalization and Standardization

Encoding categorical variables

Feature selection techniques

Model evaluation metrics: Accuracy, Precision, Recall, F1, ROC-AUC

Introduction to Deep Learning

Basics of Neural Networks

Introduction to TensorFlow/Keras

Building and training a simple ANN

Working with Real-world Datasets

Importing and cleaning large datasets

Case studies from industries (e.g., finance, healthcare, e-commerce)

Handling imbalanced data

Time-series basics (resampling, trends, forecasting)

Introduction to Natural Language Processing (NLP)

Text cleaning and preprocessing

Tokenization, Stopwords, Lemmatization

Vectorization (Bag of Words, TF-IDF)

Simple text classification (e.g., sentiment analysis

Model Deployment

Saving models using pickle/joblib

Building REST APIs with Flask

Introduction to Streamlit for dashboards

Overview of cloud deployment (AWS, GCP basics)

Capstone Project

The Capstone Project allows you to apply the skills you've learned throughout the Data Science course to a real-world problem, demonstrating your ability to collect, clean, analyze, model, and deploy data-driven solutions.

Pre-requisites

- Basic understanding of mathematics and statistics (preferably not mandatory).
- Familiarity with basic programming concepts.

Total Duration of the Course

- Total duration: 3 months.
- Days: Monday to Friday (1.5 hrs daily)
- Mode: Online

Certificate

A course completion certificate will be awarded once you have successfully completed this course.

Wishing you a Smooth Coding ahead!