

Subject Code	Subject Name	Credits
26CS506	PROGRAMMING FOR DATA SCIENCE	4

Course Objectives:

1. To understand Data Science and master Programming in Python
2. To familiarize with the programming environment and other ethics
3. To master student implementation basic ML algorithms

Learning Objectives:

After completion of the course the students will be able to:

1. Apply Python Programming for Data Analysis
2. Perform Data Pre-processing and Visualization

Unit I - Introduction to Data Science and Python Programming (12 Hrs.)

Introduction to Data Science, data science life cycle, applications of data science. Overview of Python for data science, Python environment setup. Basic Python programming: variables, data types, operators, input and output statements, control statements, functions and modules.

Unit II - Python Data Structures and Libraries (12 Hrs.)

Python data structures: lists, tuples, sets, and dictionaries. String operations and manipulation. Introduction to NumPy: arrays, array operations, indexing, slicing, and mathematical operations. Introduction to Pandas: Series and Data Frame, data loading and basic data operations.

Unit III - Data Pre-processing and Data Visualization (12 Hrs.)

Data collection and data cleaning techniques. Handling missing values, filtering and transforming data. Data normalization and standardization. Data visualization using Matplotlib and Seaborn: line chart, bar chart, histogram, pie chart, scatter plot.

UNIT IV - Machine Learning for Data Science (12 Hrs.)

Introduction to machine learning concepts. Supervised learning: classification and regression techniques. Unsupervised learning: clustering techniques. Basic machine learning algorithms using Python libraries such as Scikit-learn.

UNIT V - Data Analysis and Applications (12 Hrs.)

Exploratory Data Analysis (EDA). Statistical analysis using Python. Introduction to big data concepts and tools. Case studies and applications of data science in healthcare, finance, agriculture, and business.

Reference Books:

1. Joel Grus, Data Science from Scratch: First Principles with Python, O'Reilly Media.
2. Jake VanderPlas, Python Data Science Handbook, O'Reilly Media.
3. Wes McKinney, Python for Data Analysis, O'Reilly Media.
4. Cathy O'Neil and Rachel Schutt, Doing Data Science: Straight Talk from the Frontline, O'Reilly Media.
5. Géron Aurélien, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media.

