

Python and Preprocessing of Data

No of days - 3

Topics Covering -

- Python Basics
 - Numpy and Pandas
 - Machine Learning Introduction
 - Data Preprocessing - Imputation, EDA, Feature Engineering
-

DAY 1

INTRODUCTION

- Why Python
- Applications
- Features
- Flavours & Versions
- Comments
- Literals
- Reserved Words
- input function

DATA TYPES

- int
- float
- bool
- str

DECISION CONTROL FLOWS

- If / Nested If / If-else /If-elif-else statement
- While, for loop, While- else, for - else

OPERATORS

- Arithmetic Operators
- Relational or Comparison Operators
- Logical operators
- Bitwise operators
- Assignment operators
- Ternary Operator

LIST

- Ways of Accessing Values
- Traversing elements
- List Operations
- List Methods
- Membership Operator in List
- List Comprehension

TUPLES

- Creating Tuples
- Ways of Accessing Values
- Tuple Vs immutability
- Tuple Comprehension

SETS

- Creating Sets
- Ways of Accessing Values
- Manipulating and Accessing Sets
- Set Operations

DICTIONARY

- Why Dictionary
- Creating a Dictionary
- Accessing Values
- Update Dictionaries
- Functions of dictionary

FUNCTIONS AND METHODS

- System defined Functions
- User Defined Functions (UDFs)
- Defining a Function
- Structure of function
- Types of Function Arguments
- lambda Function
- Why Use Lambda Functions?

DAY 2

FILE HANDLING

- TYPES OF FILES
- OPENING AND CLOSING FILES
- WRITING, APPENDING AND READING FILES
- With keyword

STRING

- String Literals
- Single (') & Double Quotes(")
- Triple Quotes (''')
- Raw String ("...\'.....")
- Operations on Strings

NumPy package

- Introduction of NumPy
- Install Numpy
- Array Creation
- Array Reshaping
- Indexing
- Operation

Pandas

- Introduction
- Slicing Dataframe
- Filtering Dataframe
- Transforming Dataframe
- Advanced indexing

DAY 3

Introduction to Machine Learning

Data pre-processing

- Is your data clean?
- What is Data Pre processing ?
- Data cleaning techniques
- Handling Missing data
- Handling Categorical data

Exploratory Data Analysis (EDA)

- Introduction
- 2D Scatter-plot
- 3D Scatter-plot
- Pair plots
- Univariate, Bivariate and Multivariate
- Histogram
- Box-plot
- Variance, Standard Deviation
- Median
- IQR (InterQuartile Range)

Feature Engineering

- Introduction
- Need for Feature Engineering in Machine Learning
- Steps in Feature Engineering
- Feature Engineering Techniques

Performance Metrics

- Confusion Matrix
- ROC Curve
- Cross Validation in Machine Learning
- K fold Cross Validation & Grid search