Python Syllabus

About the Course

Data Analyst the study of the generalize extraction of knowledge from data. Being a data Scientist requires an integrated skill set spanning mathematics, statistics, machine learning, databases and programming languages along with a good understanding of the craft of problem formulation to engineer effective solutions.

Program Highlights

- ✓ Most Comprehensive Curriculum
- ✓ Trained by passionate and Industry experts
- ✓ Each concept will be explained by golden rule

  Theory → Example → Software Implementation (Python)→ Real-Time applicability
- ✓ Designed for the Industry
- ✓ Live Project
- ✓ Placement Assistance

Core Python

Introduction

✔History
✔Features
✔Setting up path
✔Working with Python
✔Basic Syntax
✔Variable and Data Types
✔Operator

Conditional Statements

✔If
✔If- else
✔Nested if-else

Looping

- ✔ For
- ✔ While
- ✔ Nested loops

Control Statements

- ✔ Break
- ✔ Continue
- ✔ Pass

String Manipulation

- ✔ Accessing Strings
- ✔ Basic Operations
- ✔ String slices
- ✔ Function and Methods

Lists

- ✔ Introduction
- ✔ Accessing list
- ✔ Operations
- ✔ Working with lists
- ✔ Function and Methods

Tuple

- ✔ Introduction
- ✔ Accessing tuples
- ✔ Operations
- ✔ Working
- ✔ Functions and Methods

Dictionaries

- ✔ Introduction
- ✔ Accessing values in dictionaries
- ✔ Working with dictionaries
- ✔ Properties
- ✔ Functions

Functions

- ✔ Defining a function
- ✔ Calling a function
- ✔ Types of functions
- ✔ Function Arguments
- ✔ Anonymous functions
- ✔ Global and local variables

Modules

- ✔ Importing module
- ✔ Math module
- ✔ Regular Expression module
- ✔ Calendar
- ✔ Time

Input-Output

- ✔ Printing on screen
- ✔ Reading data from keyboard
- ✔ Opening and closing file
- ✔ Reading and writing files
- ✔ Functions

Exception Handling

- ✔ Exception
- ✔ Exception Handling
- ✔ Except clause
- ✔ Try ? finally clause
- ✔ User Defined Exceptions

Advance Python

Regular expressions

- ✔ Match function
- ✔ Search function
- ✔ Matching VS Searching
- ✔ Modifiers
- ✔ Patterns

Learn scientific libraries in Python

- NumPy,
- SciPy,
- Matplotlib
- Pandas

NumPy and 2D Plotting

- Understanding the N-dimensional data structure
- Creating arrays
- Indexing arrays by slicing or more generally with indices or masks
- Basic operations and manipulations on N-dimensional arrays
- Plotting with matplotlib

Pandas: Python's Workhorse Toolkit for All Things Data Analysis

Accessing Data From Multiple Sources

- Reading and writing data from local files (.txt,.csv,.xls, .json, etc)
- Reading data from remote files
- Scraping tables from web pages (.html)
- Making the most of the powerful read_table method

Data Visualization

- Understanding the structure of a Figure
- Data visualization: scatter plots, line plots, box plots, bar charts,and histograms with matplotlib
- Customizing plots: important attributes and arguments

Data Analysis

- Split-apply-combine with DataFrames
- Data summarization and aggregation methods
- Pandas powerful groupby method
- Reshaping, pivoting, and transforming your data
- Simple and rolling statistics

MACHINE LEARNING – INTRODUCTION

- Introduction to Machine Learning
  - What is Machine Learning?
  - Statistics (vs) Machine Learning
  - Types of Machine Learning
    - Supervised Learning
    - Un-Supervised Learning
    - Reinforcement Learning

SUPERVISED MACHINE LEARNING

- Classification
  - Nearest Neighbor Methods (knn)
  - Logistic
- Tree based Models – Decision Tree
  - Basics
  - Classification Trees
  - Regression
- Trees Probabilistic methods
  - Bayes Rule
  - Naïve
- Bayes Regression Analysis
  - Simple Linear Regression
  - Assumptions
  - Model development and interpretation

- Sum of Least Squares
- Model validation
- Multiple Linear Regression

- Regression Shrinkage Methods
  - Lasso
  - Ridge
- Advanced Models – Black Box
  - Support Vector Machine
  - Neural
- Networks Ensemble Models
  - Bagging
  - Boosting
  - Random
- Forests Optimization
  - Gradient Descent (Batch and Stochastic)
- Recommendation Systems
  - Collaborative filtering
    - User based filtering
    - Item based filtering

UNSUPERVISED MACHINE LEARNING

- Association Rules (Market Basket Analysis)
  - Apriori
- Cluster Analysis
  - Hierarchical clustering
  - K-Means clustering
- Dimensionality Reduction

- Principal Component Analysis
- Discriminant Analysis (LDA/GDA)

MODEL VALIDATION

- Confusion Matrix ROC
- Curve (AUC) Gain and
- Lift Chart
- Kolmogorov-Smirnov Chart Root
- Mean Square Error (RMSE) Cross
- Validation
  - Leave one out cross validation (LOOCV)
  - K-fold cross validation

NATURAL LANGUAGE PROCESSING

- Introduction to Natural Language Processing
- Sentiment Analysis
- Text Similarity