

Course Title: Advanced Python for Data Science

Duration: 3 Days

About Program:

This training program will develop understanding of participants about applications of various Machine Learning algorithms in the banking sector. Data Analysis and Mining process has been explored in detail with modules focusing on data extraction, cleaning and manipulation. With various examples each concept will be demonstrated with focus on query resolution and further learning options being discussed as well.

This is a 3-day program aimed at developers, database administrators and data modelers who want to incorporate data science in their everyday work flow and project requirements by using *python* as the platform of choice. This program is an ideal blend of hands-on training experience along with required theoretical fundamentals

Pre-requisites for candidates

- Familiarity with programming/ scripting terminologies
- Basic understanding of command line/ terminal to execute scripts
- Basic idea of Machine learning and data science fundamentals

What Candidates will gain from this training?

- Detailed understanding of libraries in python used for Data analysis and Machine learning
- Hands on experience of Data Analysis, pre-processing, and statistical operations using python
- Understanding Data Visualization with various types of graphs, modifying aesthetics, parameters of plots, layouts, etc
- Exposure to various machine learning algorithms with suitable examples.

Course Outline

- **Numpy library in Python**

- Introduction to features of *numpy*
- Creating multi-dimensional arrays using *numpy*
- Indexing and slicing of *numpy* arrays
- Important functions in *numpy* (Sorting, arithmetic operations, Broadcasting ,etc)
- Comparisons, masks, boolean logics in *numpy*

- **Getting started with Pandas library**

- Reading Data from various sources(CSV, Excel, Databases, etc)
- Creating data frames from python objects
- Series and Data Frame basics
- Extracting rows, columns from data frames
- Functions to obtain results of common statistical operations(mean, standard deviation , etc)
- Applying conditions to extract matching rows and columns from data frames

- **Data Manipulation using Pandas library**

- (Includes operations like grouping of data, sorting of data based on a certain column, applying a function to all values in a data set, etc)
- Pandas *sort_values()* method to sort data records with examples
- Using *applymap()* on data set to perform a common operation on multiple elements
- Using pandas grouping functions to separate data into groups based on conditions
- Multiple examples to demonstrate filtering options in pandas
- Changing data types of objects in pandas data frame

- **Exploratory Data analysis using *Pandas* in Python**

- (Includes topics related to data cleaning, data description and overview, visualization of correlation between data attributes, etc)
- Understanding the EDA and data-driven approach for data analysis
- Extracting relevant columns and rows from data set
- Functions in pandas library to describe the data set(columns, rows, data types, etc)
- Handling missing values from data sets using *fillna()* method, interpolation techniques, dropping rows and columns with missing values, dropping duplicate records, etc
- Concatenation & Merging data frames using pandas

- **Exploratory Data analysis using *Pandas* in Python**

- Using *seaborn* library to plot histograms, charts, box plots, etc on cleaned data
- Visualization of correlations between data variables using *heatmaps* and *scatterplots*
- Plotting various types of graphs including :
 - Line Plots
 - Scatterplots
 - Violin Plots
 - Box plots
 - Swarm Plots
- Using *Seaborn* to change template styles and contexts for plots
- Addition features of *Seaborn* – x and y axis ticks, orientation , color palette , etc
- Adjusting figure size, width, height, legend , etc in graphs
- Using facet grids to plot multiple graphs in grids

- **Predictions using regression algorithm**

- Introduction to regression & its fundamentals
- Using linear regression for predicting stock prices
- Multi-variate regression for predictions
- Visualizing & analysis of results

- **Credit risk analysis by customer classification**

- (Includes analysis of credit risk using Support Vector Machines algorithm and Decision Tree algorithms and generating credit risk prediction models)
- Understanding classification techniques in Machine Learning
- Fundamentals of Decision Tree algorithm(nodes, labels, entropy, etc)
- Example for performing classification task on customers' data set using Decision tree algorithm
- Fundamentals of SVM algorithm(margin, hyperplane, support vectors, kernel trick, etc)
- Example of using SVM to classify customers based on credit risk

- **Customer Segmentation using ML algorithms**

- (Includes examples of credit card customer segmentation using K-means and K-medoids partitioning
- partitioning algorithms)
- Understanding the concept of clustering in Machine Learning
- Fundamentals of *K-means* algorithm
- Example of using k-means to segment credit card customers.
- Fundamentals of k-medoids algorithm
- Performing segmentation task on credit card customers' data set using k-medoid algorithm

- **Saving Machine Learning models**
 - (This includes approaches to data serialization and saving trained models for future use)
 - Understanding data *serialization* and *deserialization*
 - Using *pickle* module in python to serialize a trained model
 - Saving model objects in files and recovering them from files
 - Using *joblib* module -alternative to pickle module
 - Understanding the working of modules with various examples