PYTHON

|  |
| --- |
| TECHVISION  IT TRAINING AND PLACEMENT SERVICES |

**(DATA SCIENCE – Statistics, Data Visualization, Data Analytics, and Machine Learning)**

**Course Duration:** 40 hours (6 weeks regular, 8 weeks weekend)

**Sessions:** Weekdays**;** Weekends

**Course Outline: Python Data Science Training**

**Module 1- Introduction to Data Science**

* What is analytics & Data Science?
* Common Terms in Analytics
* Analytics vs. Data warehousing, OLAP, MIS Reporting
* Relevance in industry and need of the hour
* Types of problems and business objectives in various industries
* How leading companies are harnessing the power of analytics?

**Module 2 - System defined standard modules and install modules with pip**

* Math module
* Random module
* Time module
* Thread module
* OS module
* Shutil module
* Statistics module
* Introduction of NumPy, SciPy, scikit-learn, Pandas, Matplotlib , seaborn, scipy , requests, beautiful soup
* File and folder Operation (read all files from folder, create new folder and file, rename folder and files, delete folder and files)
* Data extraction and distribution
* Data consolidation

**Module 3 - Mathematical Computing with Python (NumPy)**

* NumPy Overview
* Properties, Purpose, and Types of ndarray
* Class and Attributes of ndarray Object
* Basic Operations: Concept and Examples
* Initializing arrays: random, ones, zeros
* Accessing Array Elements: Indexing, Slicing, Iteration, Indexing with Boolean Arrays
* Shape Manipulation
* Linear Algebra
* Programming examples

**Module 4 - Data Manipulation with Python (Pandas)**

* Introduction to Pandas
* Data Structures
* Series
* Data Frame
* Missing Values
* Data Operations
* Data Standardization
* Pandas File Read and Write Support
* Data Acquisition (Import & Export)
* DB interface and data extraction
* Selection, Filtering, Combining and Merging Data Frames, Normalization method
* Removing Duplicates & String Manipulation

**Module 5 - Accessing/importing and exporting data using python modules**

* Importing Data from various sources (Csv, txt, excel, access etc)
* Database Input (Connecting to database)
* Viewing Data objects - subsetting, methods
* Exporting Data to various formats
* Important python modules: Pandas, beautifulsoup
* Introduction to CSV Module
* Reader Objects
* Reading Data from Reader Objects in for loop
* Writer Objects
* The delimiter and line terminator keyword arguments
* The JSON Module
* Reading JSON with loads() function
* Writing JSON with dumps() function

**Module 6 - Data analysis – Visualization using Matplotlib & Seaborn**

* Introduction exploratory data analysis
* Descriptive statistics, Frequency Tables and summarization
* Univariate Analysis (Distribution of data & Graphical Analysis)
* Bivariate Analysis(Cross Tabs, Distributions & Relationships, Graphical Analysis)
* Creating Graphs- Bar/pie/line chart/histogram/ boxplot/ scatter/ density etc)
* Important Packages for Exploratory Analysis(NumPy Arrays, Matplotlib, seaborn, Pandas and scipy.stats etc)

**Module 7 - Web Crawling**

* Web Scraping with Python Beautiful Soup and Requests
* Scrapy
* Selenium
* Requests Get Method
* Web Scraping with Beautiful Soup
* Web Scraping with Beautiful Soup
* Accessing Tags
* Navigable Strings
* Navigating through Tag Names
* Contents and Children Methods
* Parent Method
* next\_sibling
* previous\_sibling
* next\_siblings & previous\_siblings

**Module 8 - Data Preparation and Exploration**

* Importing the Libraries
* Importing the Dataset
* Variable identification and Data extraction
* Univariate and Multivariate analysis
* Chi Square, Z-test, T-test, Entropy, Gini, ANOVA
* Missing Data
* Categorical Data
* Splitting the Dataset into the Training set and Test set
* Feature Scaling
* Dimensionality Reduction

### **Module 9 - SUPERVISED LEARNING: DECISION TREES**

* Regression Model: Liner Regression , Multiple Regression
* Decision Trees - Introduction - Applications
* Construction of Decision Trees through Simplified Examples; Choosing the "Best" attribute at each Non-Leaf node; Entropy; Information Gain, Gini Index, Chi Square, Regression Trees
* Decision Trees - Validation
* Confusion matrix

### **Module 10 - SUPERVISED LEARNING: SUPPORT VECTOR MACHINES**

* Motivation for Support Vector Machine & Applications
* Support Vector Regression
* Support vector classifier (Linear & Non-Linear)
* Mathematical Intuition (Kernel Methods Revisited, Quadratic Optimization and Soft Constraints)
* Interpretation of Outputs and Fine tune the models with hyper parameters
* Validating SVM models

### **Module 11 SUPERVISED LEARNING: KNN**

* What is KNN & Applications?
* KNN for missing treatment
* KNN For solving regression problems
* KNN for solving classification problems
* Validating KNN model
* Model fine tuning with hyper parameters

### **Module 12 SUPERVISED LEARNING: NAÏVE BAYES**

* Concept of Conditional Probability
* Bayes Theorem and Its Applications
* Naïve Bayes for classification
* Applications of Naïve Bayes in Classifications

### **Module 13 - Logistic Regression, LRGD**

* Problem Analysis
* Sigmoid function
* Mathematical Modelling (maximum likelihood estimation)
* LRGD
* Use Cases
* Digit Recognition using Logistic Regression

### **Module 14 - Comprehensive Classification Models**

* Support Vector Machine (SVM)
* K-Nearest Neighbors (K-NN)
* Kernel SVM
* Naive Bayes
* Multinomial NB
* Decision Tree Classification
* Random Forest Classification
* Implementation of the above models for real-world Dataset

#### **Module 15 - Unsupervised Learning and Clustering**

* Application of Unsupervised learning, examples, and applications
* Clustering
* K-Means Clustering in Python
* K-Means algorithm
* Elbow method
* Hands-on k means and Elbow graph
* Hierarchical Clustering in Python, Agglomerative and Divisive techniques
* SSE and Distortion measurements