

Live Online Instructor-led Training  
on  
**Statistics, Machine Learning and Its Applications**

**Total Duration:** 30 Hours (10-days)  
3-Hours / Per day

### Training Highlights

- 30+ Hours live online Instructor-led Hands-on based learning with Projects.
- **Training includes:** Soft copy of Training material, Training PPT's, Project code & Training Recording.
- Training **certificate of completion** will be provided to each Attendee.

### Projects included in Training

- Hands-on session on Python Libraries & Data processing.
- **Project 1:** Hypothesis Testing Practical Implementation on Dataset.
- **Project 2:** Linear Regression from Scratch In-depth Maths Intuition
- **Project 3:** Car Price Prediction using Linear Regression.
- **Project 4:** Logistic Regression from Scratch In-depth Maths Intuition
- **Project 5:** Applying KNN to Classify Online Shoppers Intention.
- **Project 6:** Telecom Customer Churn Analysis.
- **Project 7:** HR employee Attrition Analysis & Prediction
- **Project 8:** Clustering Credit Card Users K-Means & Hierarchical Clustering
- Applying different algorithms to solve business problems and benchmark the results.

## 10-Days Course Content [30 Hours]

### DAY 1

#### MACHINE LEARNING - INTRODUCTION

- What is Machine Learning
- Quick Comparison - Machine Learning vs. Data Science vs. AI
- Types of Data
- Supervised Learning, Semi-Supervised and Unsupervised Learning
- Types of Machine Learning
- Types of Supervised Learning
- Regression and Classification
- Different phases of a typical Analytics/Data Science projects and role of python
- Machine Learning flow to code
- Features, Labels, Class
- Discussion on Various ML Learnings
- Use Cases
- Why AI & ML are trending now?

#### PYTHON ESSENTIALS (CORE)

- Introduction to installation of Anaconda
- Introduction to Python Editors & IDE's (Anaconda, Spyder, Jupyter, etc.)
- Understand Jupyter notebook & Customize Settings
- Overview of Python- Starting with Python
- Python data types: Primitive
- Core built-in data structures – Lists, Tuples, Dictionaries
- String, String built-in methods
- Data manipulation tools (Operators, Functions, Packages, control structures, Loops, arrays etc)
- Python UDFs – def keywords

## DAY 2

### NUMPY INTRODUCTION

- The Basics (Creating arrays, shape, size, data type)
- Indexing and Slicing NumPy arrays
- Accessing/Changing Specific Elements, Rows, Columns, etc. (slicing)
- Basic Mathematics (arithmetic, trigonometry, etc.)
- Linear Algebra, Statistics
- Reorganizing Arrays (reshape, vstack, hstack)
- Advanced Indexing and Boolean Masking
- Matrix Multiplication

## DAY 3

### ACCESSING/IMPORTING AND EXPORTING DATA USING PYTHON MODULES

- Concept of Packages/Libraries - Pandas
- Importing Data from various sources (CSV, txt, excel, access etc.)
- Exporting Data to various formats
- Data Manipulation steps (Sorting, filtering, duplicates, merging, appending, sub-setting, derived variables, sampling, Data type conversions, renaming, formatting etc.)
- Exporting Data to various formats

### DATA VISUALIZATION TOOLS – MATPLOTLIB & SEABORN

- Introduction to Matplotlib
- Matplotlib interfaces
- Figure and Subplots, First plot with Matplotlib
- Line & Multiline Plots
- Line Plot, Scatter Plot, Histogram, Bar Chart, Horizontal Bar Chart
- Pie Chart, Box Plot, Area Chart
- Adding a Grid, Handling axes, Handling X and Y ticks, Sub Plotting

### DATA MANIPULATION & PREPROCESSING – CLEANSING

- Cleansing Data with Pandas
- Data Manipulation steps (Sorting, filtering, duplicates, merging, appending, sub-setting, derived variables, sampling, Data type conversions, renaming, formatting etc.)
- Stripping out extraneous information
- Scaling and Normalizing data
- Pre-processing and Formatting data
- Feature selection – RFE, Correlation etc.

## DAY 4

### MATHS/STATS/PROBABILITY & HYPOTHESIS TESTING

- Central Tendency (mean, mode, median)
- Measures of Variability (range, interquartile range)
- Variance and Standard Deviation
- Modality, Skewness, Kurtosis
- Binomial Distribution, Normal Distribution, Bernoulli's Distribution
- Normal Distribution, Uniform Distribution, Student's T Distribution, Poisson Distribution
- Hypothesis Testing – T-Test, Z-Test, ANOVA Test, Chi-Square Test

### LINEAR REGRESSION

- Assumptions
- Regression Problem Analysis
- Mathematical modelling of Regression Model
- Gradient Descent Algorithm
- Model Specification
- Optimizers
- Polynomial Linear Regression
- Multivariate Regression Model
- Parameters & Hyperparameters
- Cost Function & Cost Optimizer: Gradient Descent Algorithm
- R Squared & Adj. Squared
- Model Predictions

## DAY 5

### REGRESSION MODEL

- L1 & L2 Regularization
- Ridge, LASSO and Elastic-Net Regression

## DAY 6

### LOGISTIC REGRESSION

- Assumptions
- Reason for the Logit Transform
- Logit Transformation
- Hypothesis
- Variable and Model Significance
- Maximum Likelihood Concept
- Odds Ratio and Interpretation
- Null Vs Residual Deviance
- ROC Curve

- Model Specification
- Case for Prediction Probe
- Model Parameter Significance Evaluation
- Optimization of threshold value
- Estimating the Classification Model Hit Ratio
- Isolating the Classifier for Optimum Results

## **DAY 7**

### **RANDOM FOREST & DECISION TREE ALGORITHM**

- Concept and Working Principle
- Mathematical Modelling
- Optimization Function Formation
- Analysis of Classification Problem case
- Role of Entropy, Gini Index and Information Gain in Decision Trees
- Analysis of Regression Problem case
- Use Cases & Programming using Python
- Decision Trees – CART, ID3
- Overfitting and Pruning
- Ensemble Learning

## **DAY 8**

### **CLASSIFICATION MODEL**

- K-Nearest Neighbor (KNN)
- Naïve Bayes

## **DAY 9**

### **CLUSTERING – K MEANS and HIERARCHICAL**

- Unsupervised Learning
- Clustering Introduction
- K-Means Clustering
- Handling K-Means Clustering
- Maths behind KMeans Clustering – Centroids
- K Means from scratch
- Mean shift Introduction
- Dynamically weight
- Elbow Method – Picking K in K-Means
- Hierarchical Clustering
- Types – Agglomerative and Divisive
- Dendrogram

## DAY 10

### Projects included in Training

- Hands-on session on Python Libraries & Data processing.
- **Project 1:** Hypothesis Testing Practical Implementation on Dataset.
- **Project 2:** Linear Regression from Scratch In-depth Maths Intuition
- **Project 3:** Car Price Prediction using Linear Regression.
- **Project 4:** Logistic Regression from Scratch In-depth Maths Intuition
- **Project 5:** Applying KNN to Classify Online Shoppers Intention.
- **Project 6:** Telecom Customer Churn Analysis.
- **Project 7:** HR employee Attrition Analysis & Prediction
- **Project 8:** Clustering Credit Card Users K-Means & Hierarchical Clustering
- Applying different algorithms to solve business problems and benchmark the results.

### Who can attend?

- Training is best suitable for Engineering college faculty, Research scholar, Student & Working IT Professional.

### EduxLabs Teams

(Esoir Business Solutions Gurugram)

**M: +91-7053133032 | 8318635606**

**Email [info@eduxlabs.com](mailto:info@eduxlabs.com) | [www.eduxlabs.com](http://www.eduxlabs.com)**