

## **Data Science Course Content:**

### **1. Introduction about Data Science:**

- What is data science?
- Need of data science?
- Use cases of Data science
- How is data science different from business intelligence?
- Who are data scientists?

### **2. Statistics:**

#### **Section 1: Descriptive Statistics**

- Intro to Research Methods
- Central tendency & Variability and Central limit theorem
- Variance and Range parameters
- Normal & Sampling distributions

#### **Section 2: Inferential Statistics**

- Intro to Predictive modeling
- Hypothesis testing: t-tests/z-test(1-sample,paired sample)
- Correlation
- Regression
- Chi-square test
- Analysis of Variance (ANOVA)

#### **Section 3: practice lab**

### **3. R & Python Programming:**

#### **Section 1: Environment setup & R/Python language basics:**

- Application of machine learning
- Understand Business Analytics and R, Python
- Knowledge on the R & python language

- Community and ecosystem
- Understand the use of 'R & python' in the industry
- Compare R, Python with other software in analytics
- Install R, Python and the packages useful for the course
- Perform basic operations in R, Python using command line
- Learn the use of IDE R, Python and Various GUI

## **Section 2: Exploratory Data Analysis (EDA) and Data Preprocessing techniques in R/Python language:**

- Data structures & data types(Vectors, Matrices, Lists, Factors & Data frames/Pandas)
- Importing Data(from different File formats, Databases, Stats software and web)
- Exporting Data
- Viewing Data
- Handling Missing Values
- Date & Time
- Understanding the cor() in R & Python
- EDA functions like summarize(), list()
- Multiple packages in R & Python for data analysis
- The Fancy plots like Segment plot
- HC plot in R & Python
- Understanding on Data Visualization
- Graphical functions present in R & Python
- Plot various graphs like table plot, histogram, boxplot
- Customizing Graphical Parameters to improvise the plots
- How to use ggplot2/matplot libraries

## **Section 3: practice lab**

### **4. Machine Learning:**

#### **Section 1: Classification**

- Logistic Regression
- K-Nearest Neighbors (K-NN)
- Support Vector Machine (SVM)
- Naive Bayes
- Decision Tree Classification
- Random Forest Classification
- XG Boost Classification
- Evaluating Classification Models Performance like Confusion matrix, ROC curve, F-Score...

## Section 2: Regression

- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Support Vector Regression (SVR)
- Decision Tree Regression
- Random Forest Regression
- Evaluating Regression Models Performance like R-Squared and Adjusted R-Squared Intuitions

## Section 3: Clustering

- What is Clustering
- K- Means Clustering
- Hierarchical Clustering
- LDA (Latent Dirichlet allocation)

## Section 4: Recommendation

- What is Recommendation
- Association rules using generator
- Content and Collaborative filtering recommendation techniques
- Recommendation engine building using R/Python libraries

## 5. Real Time Projects:

### Section 1: Real time Projects

We will teach 2 Real time end to end projects in Health Care and Retail domain industries.

### Section 2: Tableau

- Introduction about Tableau and building First Bar chart
- Time series, Aggregation, and Filters
- Maps, Scatterplots, and Dashboards
- Joining and Blending Data, PLUS: Dual Axis Charts

### Section 3: Resume Preparation