

## Introduction to Data Science

- Data Science examples – -Netflix, Money ball, Amazon.
- Introduction to Analytics, Types of Analytics.
- Introduction to Analytics Methodology
- Analytics Terminology, Analytics Tools
- Introduction to Big Data
- Introduction to Machine Learning

## R & R STUDIO SOFTWARE

### Introduction to R Programming

- The importance of R in analytics
- Installing R and other packages
- Perform basic R operations
- R Studio – Install

### R Data types

- Vectors
- Lists
- Matrices
- Arrays
- Data Frames

### R variables and operators

- Types of operators – arithmetic, relational, logical
- Variable assignment

- Deleting variables
- Finding variables

### R Decision Making & Loops

- R- If statement
- R- if....else statement
- R- while loop
- R- for loop

### Basics, Data Understanding

- Built-in functions in R
- Subsetting methods
- Summarize and structure of data
- Head(), tail(), for inspecting data
- Reading and Writing Data

### R Vectors

- Vector creation
- Vector manipulation

### R Arrays

- Naming columns and Rows
- Accessing array elements
- Calculations across arrays

### R Factors

- Factors in data frame
- Changing order of Levels
- Generating Factor Levels

## **Preprocessing of Data**

- Handling Missing Values
- Changing Data types
- Data Binning Techniques
- Dummy Variables

## **Modeling & Validation**

- Splitting of data – Test & Train
- Dependent & Independent variables
- Machine learning Algorithm
- Error terms calculation
- Accuracy & Precision

## **Data Visualization**

- Histograms
- Bar plots
- Line graphs
- Customizing Graphical Parameters
- Usage of ggplot package

## **DATA EXPLORATION USING STATISTICAL METHODS**

### **Basic Statistical Concepts**

- Statistic Terminology
- Measure of Central Tendencies
- Measure of Dispersion

### **Central Limit Theorem Basic Probability**

- Probability Terminology
- Probability Rules
- Probability Types
- Bayes Theorem

### **Understanding Distributions**

- Binomial Distribution
- Poisson Distribution
- Exponential Distribution
- Normal/Gaussian Distribution
- t – Distribution
- Confidence interval

### **Advanced Statistical Concepts**

- Hypothesis Testing
- Chi square testing
- ANNOVA
- Z test
- Correlation & Covariance
- Multicollinearity

### **Model Validation/Performance evaluation**

- Confusion matrix
- Calculation of accuracy, precision, recall
- ROC and AUC
- RMSE , MAE

## MACHINE LEARNING

### Supervised Learning

- Linear Regression
- Logistic Regression
- Nonlinear Regression
- Naïve Bayes Classification
- Neural Network
- Decision Trees
- Support Vector Machines(SVM)
- K Nearest Neighbor(KNN)
- Lasso & Rigid regression

### Unsupervised Learning

- Concept of Clustering
- K means Clustering
- Hierarchical Clustering

### Time Series Analysis

- Decomposition of Time Series
- Trend and Seasonality detection and forecasting
- Smoothing Techniques
- Understanding ACF & PCF plots
- ARIMA Modeling
- Holt – Winter Method

### Optimization & Regularization

- Gradient descent

- Simulated Annealing
- Genetic Algorithm – Basics
- Dimensionality Reduction – SVD & PCA

### Ensemble Method & Association rules

- Market basket Analysis
- Ensemble Modeling

### Recommendation Engine

- Developing recommendation engines

## TEST MINING

- Introduction to Natural Language Processing
- Sentimental Analysis
- Text Classification

## HADOOP ECOSYSTEMS

- Introduction to Hadoop ecosystems
- Map Reduce
- Hive & Pig
- NoSQL – Hbase
- Kafka ,Flume ,Sqoop
- Hadoop machine learning : Mahout

## **PYTHON PROGRAMMING**

- Data types and Data Structures
- Concept of Modules
- Introduction to pandas , scikit – learn , NumPy
- Machine learning in Python

## **RESUME PREPARATION ASSISTANCE**

## **INTERVIEW QUESTION & ANSWER DISCUSSIONS**