

**MODULE 1: INTRODUCTION**

## a) 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

**MODULE 2: R & R STUDIO SOFTWARE**

## a) Introduction to R Programming

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

## b) R Data types

- Vectors
- Lists
- Matrices
- Arrays
- Data Frames

## c) R variables and operators

- Types of operators - arithmetic, relational, logical
- Variable assignment
- Deleting variables
- Finding variables

## d) R Decision Making &amp; Loops

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

## e) Basics, Data Understanding

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

- f) R Vectors
  - Vector creation
  - Vector manipulation
- g) R Arrays
  - Naming columns and Rows
  - Accessing array elements
  - Calculations across arrays
- h) R Factors
  - Factors in data frame
  - Changing order of Levels
  - Generating Factor Levels
- i) Preprocessing of Data
  - Handling Missing Values
  - Changing Data types
  - Data Binning Techniques
  - Dummy Variables
- j) Modeling & Validation
  - Splitting of data – Test & Train
  - Dependent & Independent variables
  - Machine learning Algorithm
  - Error terms calculation
  - Accuracy & Precision
- k) Data Visualization
  - Histograms
  - Bar plots
  - Line graphs
  - Customizing Graphical Parameters
  - Usage of ggplot package

**MODULE 3: DATA EXPLORATION USING STATISTICAL METHODS**

- a) Basic Statistical Concepts
  - Statistic Terminology
  - Measure of Central Tendencies
  - Measure of Dispersion
- b) Central Limit Theorem Basic Probability
  - Probability Terminology
  - Probability Rules
  - Probability Types

- Bayes Theorem
- c) Understanding Distributions
  - Binomial Distribution
  - Poisson Distribution
  - Exponential Distribution
  - Normal/Gaussian Distribution
  - t – Distribution
  - Confidence interval
- d) Advanced Statistical Concepts
  - Hypothesis Testing
  - Chi square testing
  - ANNOVA
  - Z test
  - Correlation & Covariance
  - Multicollinearity
- e) Model Validation/Performance evaluation
  - Confusion matrix
  - Calculation of accuracy, precision, recall
  - ROC and AUC
  - RMSE , MAE

**MODULE 4: MACHINE LEARNING**

- a) 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
- b) Unsupervised Learning
  - Concept of Clustering
  - K means Clustering
  - Hierarchical Clustering
- c) Time Series Analysis
  - Decomposition of Time Series

- Trend and Seasonality detection and forecasting
- Smoothing Techniques
- Understanding ACF & PCF plots
- ARIMA Modeling
- Holt – Winter Method
- d) Optimization & Regularization
  - Gradient descent
  - Simulated Annealing
  - Genetic Algorithm – Basics
  - Dimensionality Reduction – SVD & PCA
- e) Ensemble Method & Association rules
  - Market basket Analysis
  - Ensemble Modeling
- f) Recommendation Engine
  - Developing recommendation engines

## **MODULE 5: TEST MINING**

- a) Introduction to Natural Language Processing
- b) Sentimental Analysis
- c) Text Classification

## **MODULE 6: HADOOP ECOSYSTEMS**

- a) Introduction to Hadoop ecosystems
- b) Map Reduce
- c) Hive & Pig
- d) NoSQL – Hbase
- e) Kafka ,Flume ,Sqoop
- f) Hadoop machine learning : Mahout

## **MODULE 7: PYTHON PROGRAMMING**

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

## **WORKSHOP**

- ✓ REAL TIME LIVE PROJECTS
- ✓ RESUME PREPARATION ASSISTANCE
- ✓ INTERVIEW QUESTION & ANSWER DISCUSSIONS