Data Science / Machine Learning using R

Duration: 40-50 Hours

Prerequisites

- Basic knowledge of Programming Language.
- Basic knowledge of Database and files.

Course Content

1. **Fundamentals of Statistics**
2. **Introduction to Statistics**
   - Types of data
   - Measures of central tendency and dispersion
   - Statistical Graphics
3. **Probability and Probability Distributions**
   - Binomial Distribution
   - Poisson Distribution
   - Normal Distribution
4. **R Programming Basics**
   - Introduction to R
   - Data Types
   - Reading data, Subsetting Data
   - Visualizing the Data
   - Input Output Sub setting
   - Control structure
   - Functions
   - Data Exploration
   - Data Harmonization
5. **Descriptive & Inferential Statistics**
   - Estimation Theory
     - Sampling Distribution
     - Point Estimation
     - Interval Estimation
   - Sampling Distribution
   - Test of Hypothesis
     - Inference about one population means
     - Inference about two populations means
     - Analysis of Variance Concept
   - Inference about one & two population (Means & Proportion)
   - Analysis of Variance (1 Way & 2 Way)
6. **Machine Learning: Supervised Algorithms**
   - Introduction to Machine Learning
   - Naïve Bays Algorithm
   - K-Nearest Neighbor Algorithm
   - Decision Tress (SingleTree)
   - Regression
     - Correlation coefficient
     - Simple Linear Regression
     - Multiple Linear Regression
     - Logistic Regression

- Time Series Analysis
  - Moving Average
  - Simple Exponential Smoothening
  - Holt-Winter's Method
  - ARIMA Models
- Support Vector Machines
- Random Forest
- Support Vector Machines
- Model Ensembling
  - Bagging
  - Boosting
  - Stacking

7. **Unsupervised Learning Algorithms**
   - Cluster Analysis
     - Hierarchical Clustering
   - K-means Clustering
   - Association Rules Mining
   - Principal Components Analysis
8. **Natural Language Processing**
   - Term Document Matrix
   - TF-IDF
   - Word Cloud
   - Recommendations Systems
9. **Neural Network**