

Data Analytics with R Programming

Course Summary

Description

R programming is a very popular, open source environment for statistical computing, data analytics and graphics. This course introduces R programming language to students. It covers language fundamentals, libraries, advanced concepts, advanced data analytics and graphing with real world data.

Format: Lectures and hands on labs. (50% lecture + 50% labs). Pace of the class is determined by the students.

Topics

- Scalars, Vectors
- Control statements (if/else, loops)
- Matrices, Factors, lists
- String operations and file IO
- Functions and packages
- Data Frames
- Charting and Graphing in R
- Statistics
- Regressions (Linear and Logistic Regression)
- Distributions
- Text Analytics
- Dplyr package
- Collaborative Filtering
- Clustering (k-means)
- Classifications
- R and Hadoop

Audience

This course is designed for Developers / Data Analytics.

Prerequisites

Prior to taking this course, it is preferred that students have a basic programming background.

Duration

Three days

Data Analytics with R Programming

Course Outline

I. Day One: Language Basics Course

Introduction

- A. About Data Science
 - 1. Data Science Definition
 - 2. Process of Doing Data Science.
- B. Introducing R Language
- C. Variables and Types
- D. Control Structures (Loops / Conditionals)
- E. R Scalars, Vectors, and Matrices
 - 1. Defining R Vectors
 - 2. Matrices
- F. String and Text Manipulation
 - 1. Character data type
 - 2. File IO
- G. Lists
- H. Functions
 - 1. Introducing Functions
 - 2. Closures
 - 3. lapply/sapply functions
- I. DataFrames
- J. Labs for all sections

II. Day Two: Intermediate R Programming

- A. DataFrames and File I/O
- B. Reading data from files
- C. Data Preparation
- D. Built-in Datasets
- E. Visualization
 - 1. Graphics Package
 - 2. plot() / barplot() / hist() / boxplot() / scatter plot

- 3. Heat Map
- 4. ggplot2 package (qplot(), ggplot())
- F. Exploration With Dplyr
- G. Labs for all sections

III. Day 3: Advanced Programming With R Statistical

- A. Modeling With R
- B. Statistical Functions
 - 1. Dealing With NA
 - 2. Distributions (Binomial, Poisson, Normal)
- C. Regression
 - 1. Introducing Linear Regressions
- D. Recommendations
- E. Text Processing (tm package / Wordclouds)
- F. Clustering
 - 1. Introduction to Clustering
 - 2. KMeans
- G. Classification
 - 1. Introduction to Classification
 - 2. Naive Bayes
 - 3. Decision Trees
 - 4. Training using caret package
 - 5. Evaluating Algorithms
- H. R and Big Data
 - 1. Hadoop
 - 2. Big Data Ecosystem
 - 3. RHadoop
- I. Labs for all sections