Hands-on Data Science in R

Learn via: **Classroom / Virtual Classroom / Online**

Duration: **1 Day**

https://bilginc.com/en/training/hands-on-data-science-in-r-380-training/

Overview

With the advent of big data, there is an increased focus on data mining and the value that can be derived from large data sets. Data mining is the process of selecting, exploring, and modeling large amounts of data to uncover previously unknown information for business benefit.

R is an open source software environment for statistical computing and graphics and is very popular with data scientists. R is being used for data analysis, extracting and transforming data, fitting models, drawing inferences, making predictions, plotting, and reporting results. Learn how to use R basics, working with data frames, data reshaping, basic statistics, graphing, linear models, non-linear models, clustering, and model diagnostics.

Prerequisites

Attendees should have some coding experience, basic statistics.

Who Should Attend

Anyone interested in learning to use data mining techniques to find insights in data and who has at least some statistical and programming experience.

What You Will Learn

- How to configure the RStudio environment and load R packages
- How to use R basics such as basic math, data types, vectors, and calling functions
- How to use advanced data structures such as data frames, lists, and matrices
- How to use R base graphics
- How to use R basic statistics, correlation, and covariance
- How to use linear models such as simple linear regression, logistic regression
- How to use non-linear models such as decision trees and Random Forests
- How to apply clustering using K-means
- How to complete model diagnostics

Outline

Course Objectives

- How to configure the RStudio environment and load packages
- How to use R basics such as basic math, data types, vectors, and calling functions
- How to use advanced data structures such as data frames, lists, and matrices
- How to use R base graphic packages
- How to do exploratory data analysis
- How to use R to support basic statistics, correlation, and covariance
- How to use linear models such as linear regression and logistic regression
- How to use models such as decision trees, Random Forest, and K nearest neighbor
- How to use clustering models such as K-means

Course Structure

Modules:

- Introduction to RStudio
- R Basics
- Introduction to Data Mining in R

- Classification and Clustering Models in R
- Summary

Module 1 Overview

Topics:

- What is R?
- What is RStudio?
- Why use RStudio?
- Navigating RStudio
- What are packages?
- How to install packages
- Hands-on Exercises

Module 2 Overview

Topics:

- R Math
- Data Types
- Working with Data
- Loading Data
- Writing Data
- Data Structures
- Hands-on Exercises

Module 3 Overview

Topics:

- Overview of Data Mining and Data Science
- Exploratory Data Analysis
- Base Graphics in R
- Linear Regression
- Logistic Regression
- Hands-on Exercises

Module 4 Overview

Topics:

- Decision Trees
- Clustering
- Model Diagnostics
- Hands-on Exercises

Module 5 – Summary

- Skills Review