

Course Catalog

R Programming
Certification TOC

Course Outline

Module 1: Introduction to Business Analytics

- Introduction to terms like Business Intelligence, Business Analytics, Data, Information.
- How information hierarchy can be improved/introduced.
- Understanding Business Analytics and R.
- Knowledge about the R language, its community and ecosystem.
- Understand the use of 'R' in the industry.
- Compare R with other software in analytics.
- Install R and the packages useful for the course.
- Perform basic operations in R using command line.
- Learn the use of IDE R Studio and Various GUI.
- Use the 'R help' feature in R.
- Knowledge about the worldwide R community collaboration.

Module 2: Introduction to R Programming

- The various kinds of data types in R and its appropriate uses.
- The built-in functions in R like: seq(), cbind(), rbind(), merge().
- Knowledge on the various subsetting methods.
- Summarize data by using functions like: str(), class(), length(), nrow(), ncol().
- Use of functions like head(), tail(), for inspecting data.
- Indulge in a class activity to summarize data.
- dplyr package to perform SQL join in R

Module 3: Data Manipulation in R

- The various steps involved in Data Cleaning.
- Functions used in Data Inspection.
- Tackling the problems faced during Data Cleaning.
- Uses of the functions like grepl(), grep(), sub().
- Coerce the data, uses of the apply() functions.

Module 4: Data Import Techniques in R

- Import data from spreadsheets and text files into R.
- Import data from other statistical formats like sas7bdat and spss.
- Packages installation used for database import.
- Connect to RDBMS from R using ODBC and basic SQL queries in R.
- Basics of Web Scraping.

Module 5: Exploratory Data Analysis

- Understanding the Exploratory Data Analysis(EDA).
- Implementation of EDA on various datasets.
- Boxplots, whiskers of Boxplots.
- Understanding the cor() in R.
- EDA functions like summarize(), llist()
- Multiple packages in R for data analysis
- Fancy plots like the Segment plot, HC plot in R.

Module 6: Data Visualization in R

- Understanding on Data Visualization.
- Graphical functions present in R

R Programming Certification TOC

- Plot various graphs like tableplot, histogram, Boxplot.
- Customizing Graphical Parameters to improvise plots.
- Understanding GUIs like Deducer and R Commander
- Introduction to Spatial Analysis.

Module 7: Data mining: clustering techniques

- Introduction to Data Mining
- Concept of Machine Learning
- Understanding Supervised Machine Learning algorithms.
- Understanding unsupervised Machine Learning Algorithms
- K-means Clustering.
- Hierarchical Clustering

Module 8: Data Mining: Association rule mining and Sentiment analysis

- Association Rule Mining.
- User Based Collaborative Filtering (UBCF)
- Item Based Collaborative Filtering (IBCF)
- Sentiment Analysis

Module 9: Linear and Logistic Regression

- What is Regression.
- Linear Regression
- Logistic Regression.

Module 10: Anova and Predictive Analysis and Data Mining

- Decision Trees and Random forest
- Anova
- Sentiment Analysis
- Decision Tree, the 3 elements for classification of a Decision Tree
- Entropy
- Gini Index
- Pruning and Information Gain.
- Bagging of Regression and Classification Trees
- Concepts of Random Forest
- Working of Random Forest, features of Random Forest, among others.