INT232:DATA SCIENCE TOOLBOX : R PROGRAMMING

L:0 T:0 P:4 Credits:3

Course Outcomes: Through this course students should be able to

- analyze and configure R software for statistical programming environment and describe generic programming language concepts implemented in a high-level statistical language.
- establish Program in R environment to create custom analytical models to meet the dynamic business needs.
- evaluate and verify the analysis findings by conducting various statistical tests used for hypothesis testing.
- review advanced data science concepts using predictive analytics fundamentals.
- visualize and customize the various graphical packages for creating various types of graphs, plots and charts.
- research on a real-life projects to derive business insights.

Unit I

Installation and development environment overview : downloading and installing R from CRAN, installing R on your windows computer, installation Rstudio, libraries in R and R studio, installing packages,, using R reference card

Introduction to basics : discover the basic data types and operators in R

Unit II

Vectors and matrices : learn how to work with vectors and matrices in R

Factors : R stores categorical data in factors, learn how to create subset and compare categorical data

Data frames : creating, merging, naming, filtering, indexing and selection in data frames

Lists : naming, extracting, adding, deleting components from lists, subsetting a list

Unit III

R syntax : conditional statements, loops, functions and packages in R

Data input and output in R : CSV files, excel files and SQL with R

Unit IV

Advanced R programming : mathematical functions, apply family of functions, regular expressions, dates and timestamps

Data manipulation with R using : data filters, handling missing data, dplyr, tidyr, pipe

Unit V

Text mining in R : Text mining functions, string functions used in R,, analyzing text data for mining

Social media data mining : Facebook data analysis, twitter data analysis

Unit VI

Data visualization in R : ggplot2, histograms, scatterplots, barplots and boxplots

Advanced data visualization in R : themes, coordinates, faceting, interactive visualizations with plotly

Text Books:

1. R IN A NUTSHELL by JOSEPH ADLER, O'REILLY

References:

1. BEGINNING R: THE STATISTICAL PROGRAMMING LANGUAGE by DR. MARK GARDENER, WILEY

2. THE ART OF R PROGRAMMING by NORMAN MATLOFF, WILLIAM POLLOCK