INT232:DATA SCIENCE TOOLBOX : R PROGRAMMING

L: 2 T: 0 P: 2 Credits: 3

Course Outcomes: Through this course students should be able to

CO1 :: analyze and configure R software for statistical programming environment and describe generic programming language concepts implemented in a high-level statistical language.

CO2 :: establish Program in R environment to create custom analytical models to meet the dynamic business needs.

CO3 :: evaluate and verify the analysis findings by conducting various statistical tests used for hypothesis testing.

CO4 :: visualize and customize the various graphical packages for creating various types of graphs, plots and charts.

CO5 :: review advanced data science concepts using predictive analytics fundamentals.

CO6 :: analyze data for the purpose of exploration using Descriptive and Inferential Statistics.

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 WITH R : Explanation and Implementation of Basic types of graphs (scatter plot, line chart, bar chart, pie chart), Explanation and Implementation of Advanced types of graphs (Word Cloud, Heat Map, Donut Chart), Dynamic Visualization using GGPLOTS, Advanced Visualization using PLOTLY, Implementation of dashboards using RMARKDOWN

List of Practicals / Experiments:

Concepts and Basics of R Programming

- Programs to define basic data types.
- Program to demonstrate different operators.

Data Types

- Program to implement vector and metrices.
- Program to implement factors, data frame and list.

Conditional statements

- Program to demonstrate the type of loops.
- Program to implement the different types of functions.

SQL in R

- Program to perform different SQL queries.

Built-in Functions in R

- Program to demonstrate the different built in statistical, date and timestamp functions.

Data Manipulation and Data Wrangling

- Program to demonstrate the concept of data wrangling.
- Program to perform data manipulation using built in packages.

Data Mining in R

- Program to demonstrate the concept of text mining.
- Program to extract and analyze social media data.

Visualization in R

- Program to demonstrate basic visualization methods.
- Program to implement advanced visualization methods.
- Program to demonstrate dashboard in R.

Text Books:

1. R PROGRAMMING FOR BEGINNERS by SANDIP RAKSHIT, MC GRAW HILL

References:

1. HANDS ON PROGRAMMING WITH R: WRITE YOUR OWN FUNCTIONS AND SIMULATIONS by GARRETT GROLEMUND, O'REILLY