GUJARAT TECHNOLOGICAL UNIVERSITY
Syllabus for Integrated MSc, 8th Semester
Branch: Information Technology
Subject Name: Data Analytics using R Programming
Subject Code: 1380506

Teaching and Examination Scheme

Teaching Scheme			Credits	Examination Marks				Total Marks
				Theory Marks		Practical Marks		
L	T	P	C	ESE(E)	PA (M)	PA (I)	ESE(V)	
3	0	2	4	70	30	20	30	150

Content:

Sr. No.	Content	Teaching Hours	(%) Module Weightage
1.	**Introduction to R** What is R? – Why R? – Advantages of R over Other Programming Languages - R Studio: R command Prompt, R script file, comments – Handling Packages in R: Installing a R Package, Few commands to get started: installed.packages(), packageDescription(), help(), find.package(), library() - Input and Output – Entering Data from keyboard – Printing fewer digits or more digits – Special Values functions : NA, Inf and –inf.	4	15%
2.	**R Data Types** R Data Types: Vectors, Lists, Matrices, Arrays, Factors, Data Frame – R - Variables: Variable assignment, Data types of Variable, Finding Variable ls(), Deleting Variables - R Operators: Arithmetic Operators, Relational Operators, Logical Operator, Assignment Operators, Miscellaneous Operators - R Decision Making: if statement, if – else statement, if – else if statement, switch statement – R Loops: repeat loop, while loop, for loop - Loop control statement: break statement, next statement. Solving problems from Assignment sheet.	8	20%
3.	**Functions in R-Language:** R-Function : function definition, Built in functions: mean(), paste(), sum(), min(), max(), seq(), user-defined function, calling a function, calling a function without an argument, calling a function with argument values - R-Strings – Manipulating - R Vectors – Sequence vector, rep function, vector access, vector names, vector math, vector recycling, vector element sorting - R List - Creating a List, List Tags and Values, Add/Delete Element to or from a List, Size of List, Merging Lists, Converting List to Vector - R Matrices – Accessing Elements of a Matrix, Matrix Computations: Addition, subtraction, Multiplication and Division- R Arrays: Naming Columns and Rows, Accessing Array Elements, Manipulating Array Elements, Calculation Across Array Elements - R Factors –creating factors, generating factor levels gl().	10	25%
4.	**String Manipulation in R language:** String functions : grep(), nchar() , paste(), sprintf(), substr(), strsplit(), regex() gregexpr(), toupper(), tolower(), paste()	2	10%

5.	**Creating Data Frames and visualization of Data:** Data Frames –Create Data Frame, Data Frame Access, Understanding Data in Data Frames: dim(), nrow(), ncol(), str(), Summary(), names(), head(), tail(), edit() functions - Extract Data from Data Frame, Expand Data Frame: Add Column, Add Row - Joining columns and rows in a Data frame rbind() and cbind() – Merging Data frames merge() – Melting and Casting data melt(), cast(). Loading and handling Data in R: Getting and Setting the Working Directory – getwd(), setwd(), dir() File Handling in R language, -CSV Files - Input as a CSV file, Reading a CSV File, Analyzing the CSV File: summary(), min(), max(), range(), mean(), median(), apply() - Writing into a CSV File – R -Excel File – Reading the Excel file	8	15%
6	**Descriptive Statistics using R:** Descriptive Statistics: Data Range, Frequencies, Mode, Mean and Median: Mean Applying Trim Option, Applying NA Option, Median - Mode - Standard Deviation – Correlation - Spotting Problems in Data with Visualization: visually Checking Distributions for a single Variable - R –Pie Charts: Pie Chart title and Colors – Slice Percentages and Chart Legend, 3D Pie Chart – R Histograms – Density Plot - R – Bar Charts: Bar Chart Labels, Title and Colors. Line Chart, Scatterplot, Developing graphs, Box Plot, Drawing line, circle, rectangle, triangle using R language.	8	15%

Reference Books:

1. "The Book of R" by Tilman M. Davies, no starch press(San Francisco)
2. "The Art of R programming" by Norman Matloff, no starch press(San Francisco)
3. Sandip Rakshit, R Programming for Beginners, McGraw Hill Education (India), 2017
4. Seema Acharya, Data Analytics using R, McGrawHill Education (India), 2018
5. Andrie de Vries, Joris Meys, R for Dummies A Wiley Brand, 2nd Edition, John Wiley and Sons, Inc, 2015

Course Outcome:

After learning the course, the students should be able to:

No.	CO statement
CO-1	Understand comprehend the key concepts of NLP and identify the NLP challenges and issues.
CO-2	Demonstrate understanding of state-of-the-art algorithms and techniques for text-based processing of natural language with respect to morphology.
CO-3	Develop Language Modeling for various text corpora across the different languages.
CO-4	Check the syntactic and semantic correctness of sentences using grammars and labelling.
CO-5	Design and develop applications for text or information extraction / summarization /classification.