RAMAIAH
Institute of Technology

Department of Computer Science & Engineering

QUESTION BANK FOR VII Semester (Term: Aug-Dec 2020)

Data Analytics Laboratory(CSL75)

I.A. Marks : 50	**Exam Hours: 03**
Credits: 0:0:1:0	**Exam Marks: 50**

Develop and Execute the following programs in R Language using R-studio

1.	Write a R script to handle the following errors and apply the following data preprocessing over student.csv (Fields in student.csv file : stud-name,usn,address,dept,cgpa)
	
	    a. Remove numbers in names col
	    b. Remove special char in names col
	    c. In cgpa , if NA is there fill it(avg)
	    d. Handle outliers in cgpa
	    e. Handle impossible values in CGPA for stud data set(handle percentage in cgpa)
	    a. Redundant Whitespace deletion in stud name
	    b. Fixing capital letter mismatches in stud name

2.	Write a R script to handle the following errors and apply the following data preprocessing over Faculty.csv (Fields in Faculty.csv file : faculty-name, faculty-no, address, dept, designation, salary)
	
	    a. Remove numbers in names col
	    b. Remove special char in names col
	    c. In salary , if NA is there fill it(avg)
	    d. Handle outliers in salary
	    e. Handle impossible values in salary for faculty data set(handle character values in slary)
	    c. Redundant Whitespace deletion in faculty name
	    d. Fixing capital letter mismatches in faculty name

3.	Write a R script to handle the following errors and apply the following data preprocessing over dept.csv (Fields in dept.csv file : dept-name, dept-no, no-of-staff-in-dept,avg-no-of-publications)
	
	    a. Remove numbers in dept-name col
	    b. Remove special char in names col
	    c. In no_of_staff_in_dept , if NA is there fill it(avg)
	    d. Handle outliers in avg_no_of_publications column

HOD, Dept. of CSE

	e. Handle impossible values in no_of_staff_in_dept , in dept data set(handle character values in no_of_staff_in_Dept column)
	e. Redundant Whitespace deletion in dept name
	f. Fixing capital letter mismatches in dept name

4.	Write a R script for performing simple Linear Regression over advertisement dataset to predict sales based on the past history of sales and budget spent for advertisement. Advertisement data has to be stored in csv file and accessed with fields (Budget, sales). Do the predictions :
	     a. Using predefined functions.
	     b. Without using predefined functions

5.	Write a R script for Multiple linear regression analysis over advertisement dataset to predict sales based on the past history of advertisement data with different medias, budget-spent-for-advertisement and sales units (media,budget,sales). Do the predictions :
	     a. Using predefined functions.
	     b. Without using predefined functions

6.	Write a R script for KNN algorithm implementation for loan defaulters classification without using predefined functions. Compare your results with predefined functions. Consider K=1 and do the classification. Then you have to classify for k=2, K=3. Loan-defaulters.csv file has the following fields : age,loan,defaulter-status.

7.	Write a R script for performing simple linear regression over faculty dataset to predict their publications based on their past performance in terms of number of publications and experience in years. Faculty.csv file should have following field : (no-of-publications,experience-in-yrs). Do the predictions :
	     c. Using predefined functions.
	     d. Without using predefined functions

8.	Write a R script to perform logistic regression over Bank data set with the fields (qualification, credit card, bank balance, default details). Do the predictions of defaulter: a. Using predefined functions.

	b. Without using predefined functions
9.	Write a R script to Demonstrate Linear Discriminate Analysis & QDA over Bank data set with (qualification, credit card, bank balance, default details), using predefined functions. Do the predictions of defaulter: a. Using predefined functions.
10.	Write a R script for visualizing the following. a. Consider students marks scored in Physics,Chemistry,Maths,FOC,CAD. Draw the pie chart for the same. b. Consider different party candidates' votes received from public in election. Show the votes received in percentage in pie chart. c. Consider the students' performance in different subjects. (Like student roll.no, their marks, gender). Show the pie chart for girls and boys students' performance.
11.	Write a R script for visualizing the following. a. Consider students marks scored in Physics,Chemistry,Maths,FOC,CAD. Draw the bar chart for the same. b. Consider different party candidates' votes received from public in election. Show the votes received in percentage in bar chart. c. Consider the students' performance in different subjects. (Like student roll.no, their marks, gender). Show the bar chart for girls and boys students' performance.
12.	Write a R script for visualizing the following. a. Consider students marks scored in (Physics,Chemistry,Maths,FOC,CAD). Draw the box plot for the same. b. Consider different party candidates' votes received from public in election. Show the votes received in percentage in box plot. c. Consider the students' performance in different subjects. (Like student roll.no, their marks, gender). Show the box plot for girls and boys students' performance.

HOD, Dept. of CSE

Note:
- **Student is required to solve one problem from the give question bank.**
- **The questions are allotted based on lots.**

Marks Distribution:

Write-Up	Execution	Viva	Change of Program	Total
10	30	10	-10 Marks	50 Marks

HOD, Dept. of CSE