INSTITUTE OF AERONAUTICAL ENGINEERING
(AUTONOMOUS)

Code No: **BCSB06**

MODEL QUESTION PAPER - II

M. Tech I Semester

FOUNDATIONS OF DATA SCIENCE
(Computer Science and Engineering)

Time: 3 hours   Max. Marks: 70

Answer ONE Question from each Unit
All Questions Carry Equal Marks
All parts of the question must be answered in one place only

UNIT - I

1. (a) State how R programming is different from other programming languages? List the features of R languages? Write a R script to act as Basic calculator with variables and without variables. [7M]

    (b) Describe the multiple ways to read and write data from disc and web. Elaborate R functions to choose character, numerical input dynamically from user. Compute the given mathematical formula and display on console in R. $X=a^8$ $\log 32+4\sqrt{300}+\sin 63*e^8$. [7M]

2. (a) Identify the different ways to access the R objects. List the different data types in R with suitable example. [7M]

    (b) List the inbuilt summary functions to apply on vectors. Create a vector, matrix and array data object and apply inbuilt functions on it. [7M]

UNIT - II

3. (a) Write the R script which include relevant packages and procedure to access .csv and .exl files. Elaborate with suitable example. [8M]

    (b) How to apply the SQL queries on data frames. List out the No SQL databases and provide the example for each. Elaborate the process with XML data. [6M]

4. (a) Discuss the multicollinearity. Describe the procedure for finding hidden relations among attributes in the given dataset. [7M]

    (b) Generate prediction model using linear regression for finding relative relation among variables. Write a R script to get a linear equation y=mx+c form for the heart weight and body weight in cats dataset. [7M]

UNIT - III

5. (a) Outline about the learning of a model? Write any four learning techniques and in each case give the expression for weight- updating. [8M]

    (b) Describe the limitations of the perception model. How to create and evaluate a data model. Describe with one case study. [6M]

6. (a) List the different types clustering. Write about k-nn algorithm. Write a R script to cluster the mtcars dataset using k-nn algorithm. [7M]

   (b) State the suitable classification algorithm for pima Indian diabetes dataset classification. Improve the accuracy by performing multiple preprocess steps. [7M]

   Calculate the Minkowski distances among objects for X=

UNIT - IV

7. (a) Describe the basic structure of back propagation ANN algorithm. Elaborate each layer importance and error propagation function. [7M]

   (b) How to evaluate hypothesis of the given problem. Describe the basic principle of sampling theory. [7M]

8. (a) Describe the null and alternative hypothesis with examples. What is p-value and give its importance. [7M]

   (b) State different types of learning algorithms with suitable example. Elaborate lazy learning algorithms. [7M]

UNIT - V

9. (a) Summarize the importance of visualization of different type of data in exploration data analysis. [8M]

   (b) List out the different plots with relevant package to explore and summarize the numerical and text data in R. [6M]

10. (a) How to partition the window to get more number of plots. Discuss on single and multi-object plots in R. [9M]

    (b) Discuss about the residuals with respect to observed values? State a case study to show the fitted line and residuals in logistic regression. [5M]