Code No: 155FN **R18**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

B. Tech III Year I Semester Examinations, March - 2024

INTRODUCTION TO DATA SCIENCE

(Computer Science and Engineering – Data Science)

Time: 3 Hours **Max. Marks: 75**

Note: i) Question paper consists of Part A, Part B.
ii) Part A is compulsory, which carries 25 marks. In Part A, Answer all questions.
iii) In Part B, Answer any one question from each unit. Each question carries 10 marks and may have a, b as sub questions.

PART – A

(25 Marks)

1.a)	Why is fitting a model important in statistical modeling?	[2]
b)	Briefly write about the concept of datafication.	[3]
c)	Define asymmetric attributes.	[2]
d)	How is the mode of a dataset determined?	[3]
e)	What function is used to create a numeric vector in R?	[2]
f)	How can you view the levels of a factor?	[3]
g)	Name one popular R package used for data manipulation.	[2]
h)	Why is package management important in R programming?	[3]
i)	List the advantages of wavelet transforms in data reduction.	[2]
j)	What role does clustering play in data reduction?	[3]

PART – B

(50 Marks)

2.a) Explain the concept of big data, discussing its characteristics and challenges in data science.
b) Explain the importance of navigating through the hype in the field of data science. [5+5]

OR

3.a) Differentiate between populations and samples.
b) Explain the basics of programming with R, emphasizing its applications in data science. [5+5]

4.a) What is an attribute? Define Binary attribute and nominal attributes.
b) Differentiate between discrete and continuous attributes.
c) Discuss the steps involved in calculating the mean of a dataset. [3+3+4]

OR

5. Explain how range, quartiles, variance, and standard deviation measure the dispersion of data. [10]

6.a) Give examples of vector addition, subtraction, multiplication, and division in R.
b) Elaborate on techniques for extracting specific elements, rows, or columns from a matrix. [5+5]

OR

7.a) Explain methods for merging lists and situations where list merging is practical.
b) Explain the process of sorting a data frame based on a specific column. [5+5]

8.a) Explain the concept of function scoping in R.
b) Write a detailed explanation of recursion in R functions, including a practical example. [5+5]

OR

9.a) Elaborate on the steps involved in creating a function in R. Include an example to illustrate.
b) Discuss the utility of the "round" function in R. [5+5]

10.a) Write an overview of data reduction strategies.
b) Explain the challenges and benefits of visualizing complex data and relations. [5+5]

OR

11. Explain the following:
a) Hierarchical visualization techniques
b) Pixel-oriented visualization techniques. [5+5]

---ooOoo---