R22 Regulation *Subject code: 4E3HA*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

(Autonomous, Accredited by NAAC with 'A+' Grade)

B.Tech III Semester Supplementary Examinations, July 2024

INTRODUCTION TO DATA SCIENCE

(CSE (DS))

Maximum Marks: 60 Date:25.07.2024 Duration: 3 hours

Note: **1.This question paper contains two parts A and B.**
2. Part A is compulsory which carries 10 marks. Answer all questions in Part A.
3. Part B consists of 5 Units. Answer any one full question from each unit.
4. Each question carries 10 marks and may have a, b, c, d as sub questions.

	Part-A	CO	Bloom Tx
All the following questions carry equal marks (10×1M=10 Marks)			
1.a)	How does data science differ from traditional computer science,	1	L1
b)	Who is the primary target audience for "The Quant Shop"?	1	L1
c)	What is data visualization and exploratory data analysis?	2	L2
d)	What is the concept of baseline models	2	L2
e)	What is the fundamental concept behind linear regression?	3	L3
f)	How are graphs and networks utilized in data analysis	3	L3
g)	What is the significance of R in data analysis?	4	L4
h)	How does R handle NA and NULL values?	4	L4
i)	How are matrices and arrays commonly used in data science?	5	L5
j)	How do data frames organize data in R,	5	L5
	Part-B		
Answer All the following questions. (5×10M=50Marks)			Bloom Tx level
2	Can you provide an overview of "The Quant Shop" and its main focus in the realm of data science? [10M]	1	L1
	OR		
3	What are the common challenges and pitfalls associated with Z-scores and normalization, and how can they be addressed? [10M]	1	L1
4	What tools or software do you recommend for effective exploratory data analysis, and why are they beneficial? [10M]	2	L2
	OR		
5	Discuss the different philosophies or approaches to mathematical modeling, and how they influence the development and interpretation of models? [10M]	2	L2
6	a) What methods are commonly used for measuring distances in data analysis? [5M] b) How does nearest neighbor classification work, and in what scenarios is it advantageous? [5M]	3	L2

	OR		
7	a) What is the primary purpose of clustering in data analysis? [5M] b) What are common challenges in implementing clustering algorithms, and how can they be addressed? [5M]	3	L3
8	a) What processes occur during the startup and shutdown of R, and why are they important for users? [5M] b) Explain the concept of vectorized operations in R and their advantages in data analysis tasks? [5M]	4	L3
	OR		
9	How are scalars, vectors, arrays, and matrices declared in R, and what distinguishes them from each other? [10M]	4	L2
10	a) Describe the structure of a data frame. How does it differ from a matrix? [5M] b) How can you create a data frame in R. [5M]	5	L3
	OR		
11	a) How do you create factors, and what role do they play in statistical analysis? [5M] b) Explain the difference between ordered and unordered factors. [5M]	5	L5