R22 Regulation *Subject code:4E3HA*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech III Semester Supplementary Examinations, June 2026

INTRODUCTION TO DATA SCIENCE

(CSE(DS))

Maximum Marks: 60 Date:18.06.2026 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

All the following questions carry equal marks (10X1M=10 Marks)		Marks	CO	BTL
1.a)	Explain the importance of asking interesting questions from data.	1M	1	K2
b)	What is regression	1M	1	K1
c)	What is meant by reading a graph	1M	2	K1
d)	Define interactive visualization	1M	2	K1
e)	What is logistic regression	1M	3	K1
f)	Define nearest neighbor classification	1M	3	K1
g)	What does the example function do in R	1M	4	K1
h)	What are NA values in R	1M	4	K1
i)	Define a data frame	1M	5	K1
j)	Explain how tables are used in handling categorical data	1M	5	K2

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	BTL
2	Analyze the properties of data and explain how they influence the choice between classification and regression techniques. Using appropriate examples, evaluate how asking meaningful questions from data leads to selecting the right analytical approach.	10M	1	K4
	OR			
3	Examine advanced ranking techniques and analyze how they improve upon basic scoring systems. Discuss how normalization and Z-scores affect ranking fairness and interpretability in real-world applications.	10M	1	K4
4	Describe the components of a visualization aesthetic and explain how interactive visualizations enhance understanding of complex datasets. Provide examples showing how poor design choices can mislead interpretation.	10M	2	K2
	OR			

5	Explain model evaluation techniques and discuss why evaluation environments are important for comparing mathematical models. Illustrate how baseline models help in understanding the relative performance of advanced models.	10M	2	K2
6	Discuss the assumptions and workings of logistic regression and explain how it differs from linear regression. Describe how parameter fitting influences model accuracy.	10M	3	K2
	OR			
7	Explain clustering and describe how distance and network methods help identify patterns in data. Illustrate how PageRank uses network structure to determine node importance.	10M	3	K2
8	Apply R programming concepts to describe how startup, shutdown, help, and example functions support effective coding. Demonstrate how these tools assist in learning and applying R data structures.	10M	4	K3
	OR			
9	Apply the concepts of vector declarations, common vector operations, and the use of all() and any() to real data processing scenarios. Show how NA/NULL handling and filtering are incorporated into analytical workflows.	10M	4	K3
10	Explain how matrices and arrays differ in structure and usage. Describe how they are applied to represent multi-dimensional datasets in R.	10M	5	K2
	OR			
11	Describe the purpose and structure of lists, data frames, factors, and tables in R. Explain how each supports the analysis of structured and categorical data.	10M	5	K2