R20 Regulation *Subject code:3E8EA&3E8HA*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech VIII Semester Supplementary Examinations, May 2026

DATA SCIENCE

(Common to CSE & CSE(DS))

Maximum Marks: 70 Date: 04.05.2026 Duration: 3 hours

Note: **1.This question paper contains two parts A and B.**
2. Part A is compulsory which carries 20 marks. Answer all questions in Part A.
3. Part B consists of 5 Units. Answer any one full question from each unit which carries 10M.
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 (10X2M=20 Marks)		Marks	CO	BTL
1	What is Data Science?	2M	1	L1
2	What are the assumptions of linear regression?	2M	1	L1
3	How do you choose the optimal value of k in k-NN?	2M	2	L1
4	What are the basic tools used in exploratory data analysis (EDA)?	2M	2	L1
5	How can domain knowledge influence the process of feature generation for user retention?	2M	3	L1
6	Define filters and wrappers.	2M	3	L1
7	What is feature selection?	2M	4	L1
8	Define clustering of graphs.	2M	4	L1
9	What is data visualization	2M	5	L1
10	Define Neighborhood properties in graphs.	2M	5	L1

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	Bloom Tx
11	Explain the principles of populations and samples within statistical inference, illustrating how these concepts are applied in data analysis and hypothesis testing.	10M	1	L2
	OR			
12	Explain the concept of probability distributions and their role in statistical inference.	10M	1	L2
13	Explain the concept of the elbow method in the context of k-Means clustering? analyze the rationale behind using the elbow method.	10M	2	L2
	OR			
14	List out the advantages and disadvantages of using APIs for data collection? Explain.	10M	2	L2
15	Compare and contrast filter-based feature selection methods with wrapper-based methods in the context of user retention prediction.	10M	3	L2

	OR			
16	Discuss the importance of creativity and imagination in the process of feature generation for user retention.	10M	3	L2
17	Discuss the trade-offs involved in using dimensionality reduction techniques such as SVD and PCA in recommendation systems.	10M	4	L2
	OR			
18	How would you evaluate the concept of implicit feedback and its significance in the design of recommendation engines?	10M	4	L2
19	Build the visualization for a complex dataset, Explain the steps and considerations involved in creation.	10M	5	L2
	OR			
20	Discuss the importance of user retention for Real Direct and how brainstorming sessions can help identify relevant features using filters, wrappers, decision trees, and random forests.	10M	5	L2