R22 Regulation *Subject code:4P5EC*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech V Semester Supplementary Examinations, April 2026

FUNDAMENTALS OF DATA SCIENCE

(CSE)

Maximum Marks: 60 Date:27.04.2026 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)	Write any two features of R.	1M	1	L1
b)	Define Exploratory Data Analysis	1M	1	L1
c)	What does the 'k' represent in kNN?	1M	2	L1
d)	State the difference between web scraping and APIs	1M	2	L1
e)	Illustrate the main challenge in user (customer) retention analysis	1M	3	L2
f)	Name the three main categories of feature selection methods.	1M	3	L1
g)	How is user feedback important in building recommendation systems?	1M	4	L1
h)	What is the "curse of dimensionality"?	1M	4	L1
i)	Differentiate between directed and undirected social network graphs	1M	5	L4
j)	Name any two popular tools used for data visualization.	1M	5	L1

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	BTL
2	a) What is the Data Science hype? How can we move beyond the hype? b) Explain the concepts of population and sample with examples.	5M 5M	1	L2
	OR			
3	a) Demonstrate the steps involved in building a statistical model. b) Explain the different stages of the Data Science Process in detail.	5M 5M	1	L3 L2
4	a) Discuss the assumptions of Linear Regression and their significance. b) Describe the advantages and disadvantages of kNN in classification problems.	5M 5M	2	L2 L2
	OR			
5	a) Explain the k-Means algorithm in detail with suitable steps and a diagram. b) What are APIs? Explain how APIs can be used to collect real-time data.	5M 5M	2	L3

6	a) Discuss how domain expertise and creativity contribute to effective feature generation. b) Illustrate the importance of feature generation and feature selection in the data science process.	5M 5M	3	L2
	OR			
7	a) Describe the role of Decision Trees in feature selection. b) Explain how wrapper methods use model performance for selecting the best subset of features.	5M 5M	3	L4 L2
8	a) Explain the concept and importance of recommendation systems in modern applications. b) Outline the challenges caused by high-dimensional data.	5M 5M	4	L2
	OR			
9	a) Demonstrate the important properties of Singular Value Decomposition. b) Describe the steps involved in performing Principal Component Analysis.	5M 5M	4	L2 L4
10	a) Explain how social networks can be represented as graphs with suitable examples. b) Discuss different algorithms for direct discovery of communities in graphs.	5M 5M	5	L2
	OR			
11	a) Describe the steps involved in creating your own visualization for a complex dataset. b) Compare various visualization techniques such as bar charts, heat maps, scatter plots, and network diagrams.	5M 5M	5	L2 L4