R22 Regulation *Subject code:4P5EC*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

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

B.Tech V Semester Supplementary Examinations, May 2025

FUNDAMENTALS OF DATA SCIENCE

(CSE)

Maximum Marks: 60 Date: 27.06.2025 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	Bloom Tx
1.a)	How does data science differ from traditional data analysis?	1M	1	L1
b)	What is the role of Exploratory Data Analysis (EDA) in the Data Science process?	1M	1	L1
c)	Discuss the Linear Regression model.	1M	2	L2
d)	What is Bayes theorem?	1M	2	L1
e)	Describe the difference between feature selection and feature extraction.	1M	3	L2
f)	Provide an example of a commonly used filter technique for feature selection.	1M	3	L3
g)	What is dimensionality reduction?	1M	4	L1
h)	What are the key algorithmic ingredients of a recommendation engine?	1M	4	L1
i)	How are the nodes and edges represented in Social-Network Graphs?	1M	5	L1
j)	What are the challenges faced when visualizing complex datasets.	1M	5	L1

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	Bloom Tx
2	a) What are different skill sets required for data scientists? b) Discuss the Data Science Process in detail, highlighting its stages and their importance in solving real-world problems.	5M 5M	1	L1 L2
	OR			
3	a) Write a brief introduction to R programming language? Why is R particularly suited for statistical analysis and data visualization? b) Why are the probability distributions are important in statistical modeling? Explain with examples of commonly used distributions in Data Science.	5M 5M	1	L1 L1
4	a) Explain the Naïve Bayes algorithm and its application in spam filtering. b) Compare and contrast APIs and web scraping for data extraction. Discuss the advantages, challenges.	5M 5M	2	L1 L2
	OR			
5	a) Describe the k-NN algorithm in detail, including its training and prediction phases.	5M	2	L2

	b) What is data wrangling? Discuss its steps, tools, and challenges in preparing raw data for analysis.	5M		L2
6	a) Explain the various steps involved in the feature generation process with an example application. b) Describe the structure of decision trees and explain how they inherently perform feature selection.	5M 5M	3	L2 L2
	OR			
7	a) Discuss the importance of feature generation in machine learning, highlighting the role of domain expertise and imagination in identifying meaningful features. b) Compare and contrast the filter, wrapper, and embedded methods for feature selection.	5M 5M	3	L1 L2
8	a) Compare and contrast collaborative filtering and content-based filtering b) Discuss Principal Component Analysis (PCA) as a dimensionality reduction technique.	5M 5M	4	L1 L2
	OR			
9	a) Discuss the building of a User-Facing Data Product in recommender systems. b) What is Singular Value Decomposition (SVD)? Explain how it is applied in recommendation systems.	5M 5M	4	L1 L2
10	a) Analyze the role of graph partitioning in understanding the structure of social networks and its impact on network analysis. b) Discuss different tools and softwares used for data visualization, focusing on their strengths and weaknesses.	5M 5M	5	L1 L2
	OR			
11	a) What is the role of Clustering in graphs? b) Give an Example (industry) project that inspire the data visualization.	5M 5M	5	L1 L3