



R22 Regulation **Subject code:4P5EC**
TKR COLLEGE OF ENGINEERING AND TECHNOLOGY
(Autonomous, Accredited by NAAC with 'A+' Grade)

B.Tech V Semester Regular Examinations, December 2024
Fundamentals of Data Science
(CSE)

Maximum Marks: 60

Date:08.01.2025

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		CO	Bloom Tx
All the following questions carry equal marks (10X1M=10 Marks)			
1.a)	What is Data Science?	1	L1
b)	What is the purpose of Exploratory Data Analysis (EDA) in the Data Science process?	1	L1
c)	What is the role of 'K' in k-Nearest Neighbors method?	2	L1
d)	Name two popular tools for web scraping and briefly describe their usage.	2	L2
e)	How does domain expertise contribute to the process of feature generation?	3	L1
f)	Describe the difference between feature selection and feature extraction.	3	L2
g)	What is the primary goal of a recommendation system?	4	L1
h)	What is Singular Value Decomposition?	4	L1
i)	What is a social network graph?	5	L1
j)	Name two commonly used tools for creating data visualizations.	5	L2
Part-B			
Answer All the following questions. (5X10M=50Marks)		CO	Bloom Tx
2	a) Explain the process of Data science? [5M] b) What do you think of Hype about data science? [5M]	1	L2 L3
OR			
3	a) What is Big Data? Discuss the challenges and opportunities presented by Big Data. [5M] b) Analyze the importance of R programming language and explain its role in Data Science. [5M]	1	L1 L3
4	a) Explain the working of linear regression with an example. Discuss its advantages, and limitations. [5M] b) Compare and contrast APIs and web scraping for data extraction. [5M]	2	L2 L2
OR			
5	a) Illustrate the k-means clustering algorithm with an example. Explain the steps involved and discuss challenges. [5M] b) Discuss the challenges faced in developing a spam filtering system using machine learning algorithms. [5M]	2	L2 L2
6	a) Analyze the importance of feature generation in machine learning. Explain how it can be performed. [5M]	3	L3

	b) Analyze how Random Forests enhance the feature selection process compared to single decision trees. [5M]		L3
	OR		
7	Explain in detail about the filter, wrapper, and embedded methods for feature selection. [10M]	3	L2
8	a) Describe the workflow of a recommendation system. [5M] b) Discuss Principal Component Analysis (PCA) as a dimensionality reduction technique [5M]	4	L1 L2
	OR		
9	a) Explain the role of dimensionality reduction in recommendation systems. [5M] b) Analyse different algorithmic ingredients of a Recommendation Engine. [5M]	4	L2 L3
10	a) Explain how social network graphs represent relationships and interactions among entities and discuss their significance in real-world applications. [5M] b) Analyze the role of data visualization and discuss how it enhances data understanding. [5M]	5	L2 L3
	OR		
11	a) Explain different neighborhood properties in graphs. [5M] b) Discuss different Ideas and tools for data visualization [5M]	5	L2 L1