Reg. No.												

Question Paper Code	13689

M.E - DEGREE EXAMINATIONS, APRIL / MAY 2025

Second Semester

Big Data Analytics

24PBDPC201 - FOUNDATIONS OF DATA SCIENCE

Regulation - 2024

Duration: 3 Hours | Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	List the key roles in a data science project.	2	K1	CO1
2.	Mention two SQL commands used for data retrieval.	2	K1	CO1
3.	Define clustering in machine learning.	2	K1	CO2
4.	Name two methods of validating machine learning models.	2	K1	CO2
5.	Define probability distribution in R with an example.	2	K1	CO3
6.	Explain the process of reading a CSV file in R.	2	K1	CO3
7.	Define Combiner in a MapReduce job.	2	K1	CO4
8.	Name two programming languages that can be used to write Hadoop MapReduce programs.	2	K1	CO4
9.	Name two file formats supported for exporting graphs in R.	2	K1	CO5
10.	Display multiple plots in one window in R with example.	2	K1	CO5

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

			Marks	K – Level	CO
11.	a)	Analyze SQL and NoSQL databases based on their advantages, disadvantages, and use cases.	13	K4	CO1
		OR			
	b)	Demonstrate the major data cleaning techniques? Apply the techniques on real-world examples to show how missing values, duplicates, and inconsistencies are handled.	13	K4	CO1
12.	a)	Describe the process of choosing a machine learning model. How do you select an appropriate model based on problem type and data characteristics?	13	K2	CO2
		OR			
	b)	Explain the Naïve Bayes classifier? Explain its working mechanism, assumptions, and real-world applications.	13	K2	CO2

13.	a)	Illustrate the R programming on a sample dataset to create arrays, and matrices. Demonstrate how to manipulate and perform operations on arrays and matrices.	*13*	*K3*	*CO3*
		OR			
	b)	Demonstrate the methods in R that are used to visualize and analyze data distribution in R.	*13*	*K3*	*CO3*
14.	a)	Apply your understanding of Hadoop by writing a basic MapReduce program to solve a real-world problem.	*13*	*K4*	*CO4*
		OR			
	b)	Draw the Diagram and explain its architecture, components, and key features that make it suitable for handling big data applications.	*13*	*K4*	*CO4*
15.	a)	Describe key techniques for creating effective data presentations in R.	*13*	*K2*	*CO5*
		OR			
	b)	List what are matrix plots in R? Illustrate their significance in data analysis with an example.	*13*	*K2*	*CO5*

PART - C (1× 15 = 15 Marks)

16.	a)	Compare and contrast Linear Regression and Logistic Regression based on their mathematical formulations, applications, and assumptions.	*15*	*K3*	*CO2*
		OR			
	b)	Draw and explain the MapReduce architecture in detail. Explain the roles of the JobTracker, TaskTracker, NameNode, and DataNode.	*15*	*K3*	*CO4*