Term Evaluation (Even) Semester Examination March 2025

Roll no.................................

Name of the Course and semester: MCA 4
Name of the Paper: Data Science using R
Paper Code: TMC401

Time: 1.5 hour   Maximum Marks: 50

Note:
- *(i)* Answer all the questions by choosing any one of the sub questions
- *(ii)* Each question carries 10 marks.
- *(iii)* Please specify COs against each question.

Q1. (10 Marks) CO1, CO2
a. What are the key components of the Data Science process, and how do they contribute to data analysis?

OR

b. Explain the significance of Data Science in today's world and provide examples of how it is applied across different industries.

Q2. (10 Marks) CO1, CO2
a. Write short notes on any three
   - (i) R programming language
   - (ii) RStudio
   - (iii) Independent and dependent variables
   - (iv) CSV file.

OR

b. Compare and contrast different types of classifications, such as supervised and unsupervised learning, and explain their respective applications with respect to machine learning in data science.

Q3. (10 Marks) CO1, CO5
a. Explain why data exploration is a critical step before building predictive models.

OR

b. What is statistics? What is the difference between descriptive and inferential statistics with example.

Q4. (10 Marks) CO1, CO5
a. A dataset representing the test scores of 10 students in a data science exam is given below:
Scores: 72, 85, 90, 90, 95, 85, 80, 88, 92, 75 calculate mean, mode and median give its formula also.

OR

b. Explain standard deviation with its formula. The ages of five friends are: 20, 22, 25, 23, 24 calculate standard deviation.

Q5. (10 Marks) CO1
a. What patterns or trends can be observed in the data visualization, and how do different variables interact with each other?

OR

b. What are the common techniques used in data cleaning? Demonstrate how to handle outliers in a dataset.