BT-4/M-22   **44216**

DATA SCIENCE R PROGRAMMING
Paper–PC-CS-AIDS-204A

Time : Three Hours]   [Maximum Marks : 75

Note : Attempt *five* questions in all, selecting at least *one* question from each unit. All questions carry equal marks.

UNIT–I

1. (a) Comment on the evolution of data science along with its life cycle.   7.5

    (b) Write short notes on Hadoop and Spark.   7.5

2. (a) Discuss the various measures of central tendency using suitable examples.   7.5

    (b) What is normal distribution? How can you find the mean and variance of normal distribution? State its major characteristics.   7.5

UNIT–II

3. (a) What are the various data types in R? Explain using suitable examples.   7.5

    (b) Explain various types of loops in R using suitable examples.   7.5

4. (a) Discuss the various functions in R for data wrangling.   7.5

(b) How can you perform following in R?

(i) Handling missing data.

(ii) Creating reports.

(iii) Reading from CSV files. 7.5

UNIT-III

5. What are the various type of graphs that can be drawn in R? Explain in detail along with each graph purpose and implementation in R using suitable examples. 15

6. How can you perform following in R?

(i) Outlier detection.

(ii) PCA for dimensionality reduction.

(iii) Logistic Regression. 15

UNIT-IV

7. What is a Random Forest? Explain its significance along with its implementation in R using suitable examples. 15

8. (a) What is meant by linear regression? How can you implement linear regression in R? Explain in brief.

7.5

(b) What is CART? How can you implement this model in R? Explain. 7.5