Roll No. ..... Total Pages : 03

BT-4/M-24 44218

DATA SCIENCE WITH R PROGRAMMING
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) Explain the life cycle of Data Science, detailing each stage from data collection to deployment of predictive models. 10

   (b) What are the advantages of using Hadoop for processing big data ? 5

2. (a) How do you handle missing data in a dataset ? 7

   (b) Explain the difference between the mean() and median() functions in R Programming with example and script in R programming. 5

   (c) Provide examples of how Business Intelligence and Data Science are used in different industries. 3

Unit II

3. (a) Create a vector x containing numbers from 1 to 10. Write an R script to write a loop to calculate the sum of the squares of these numbers. **10**

   (b) Write R code to read data from a CSV file named "data.csv" into a data frame. **5**

4. (a) What are the different types of Operators used in R programming ? Explain them with suitable example. **10**

   (b) Explain the difference between vector and list in R programming. **5**

Unit III

5. (a) What is exploratory data analysis (EDA), and why is it important in data analysis ? Provide examples of EDA techniques in R. **8**

   (b) Explain the concept of outliers in a dataset and how they can affect statistical analyses. **7**

6. (a) Describe a situation where PCA would be useful and how you would implement it in R. **8**

   (b) Explain the purpose of statistical graphs in data analysis and provide examples of common types. **7**

Unit IV

7. (a) How do you interpret a decision tree produced by the CART algorithm in R Programming ? **8**

   (b) Explain the principle behind linear regression. How does it model the relationship between independent and dependent variables ? **7**

8. (a) Describe the Naïve Bayes algorithm. Provide an example of how you would use the Naïve Bayes classifier in R. **8**

   (b) Discuss the advantages and disadvantages of using random forests for machine learning tasks. **7**