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

BT-4/M-23   44216

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 various tools used for Data Science.

                                                                    7

    (b) Explain Bayes' Theorem with suitable example. 8

2. (a) What is Binomial Distribution ? Explain with example.   7

    (b) Write short notes on the following :   2×4=8

         (i)   Spark

         (ii) Big Data

         (iii) Business Intelligence

         (iv) Random Variable.

Unit II

3. (a) Explain various Operators in R Programming. **7**

   (b) How to deal with Missing Values in R Programming ? Explain with example. **8**

4. (a) Write a R script to design a calculator using various functions in R Programming. **7**

   (b) How to read various file in R Programming ? Explain with suitable example. **8**

Unit III

5. (a) What do you mean by EDA ? How is it implemented in R Programming ? **7**

   (b) Explain Hierarchical Clustering with its characteristics. **8**

6. (a) Explain various Statistical graph used Central Tendency measurement in R Programming. **7**

   (b) What do you mean by Outlier ? Explain the difference between Outlier Detection *vs.* Novelty Detection. **8**

Unit IV

7. (a) What do you mean by Logistic Regression ? Explain its implementation in R Programming. **7**

(b) What is Naive Bayes Algorithm ? Explain its implementation in R Programming. 8

8. (a) What do you know about Hard Margin SVM and Soft Margin SVM ? Explain with example. 7

(b) Prove that in the Bagging method only about 63% of the total original examples (total training set) appear in any of sampled bootstrap datasets. Provide proper justification. 8