2022

BCA (Hons)

B.Sc. Fifth Semester End Examination - 2022

PAPER - C12T

Introduction to Data Science

Full Marks : 40   *Time : 2 hours*

The figures in the right-hand margin indicate marks.
Candidates are required to give their answers in their own
words as far as practicable.
Illustrate the answers wherever necessary.

Group-A

Attempt any five questions   **5×2=10**

1. a) How does Map Reduce framwork work in Hadoop?

    b) What is *NaN* and *Inf* represent in R programming?

    c) How does *lapply ()* differ from *sapply()* function?

d) What do you mean by inner join?

e) What is root cause analysis?

f) What do you mean by A/B testing?

g) What is the difference between box-plot and histogram?

h) How to create a data frame in R?

Group -B

Attempt any four questions   **4×5=20**

2. What is data science? Explain briefly the important characteristics of big data.   2+3

3. What is versioin control? How does GitHub help in managing version control?

4. Write a R program to read a vector from keyboard in run time and find it's mean and median.

5. What are the main components in Data Science?

6. What do you mean by cross-validation training? Explain k-fold cross validation method.

7. What are the steps to build and evaluate the knn classifier model in R?

Group-C

Attempt any four questions   **1×10=10**

8. a) Write a short note on tidyr package and it's usage. 5

   b) What is multiple regressioin? How can we know the existing datasets already available in R? How can we read from a .csv file?   2+1+2

9. Explain different Exploratory Data Analysis techniques. Give any five features of R language related to data science functionality?   5+5

———