

No. of Printed Pages : 4

**MCS-226**

**MASTER OF COMPUTER  
APPLICATIONS (MCA) (NEW)**

**Term-End Examination**

**June, 2022**

**MCS-226 : DATA SCIENCE AND BIG DATA**

*Time : 3 Hours*

*Maximum Marks : 100*

*(Weightage : 70%)*

*Note : (i) Question No. 1 is compulsory and carries  
40 marks.*

*(ii) Attempt any three questions from the  
remaining Q. Nos. 2 to 5.*

- 
- 
1. (a) What is the Data Science ? What are its applications ? Explain the terms categorical and quantitative data with the help of an example for each. 5

- (b) Assume that a fair coin is tossed 4 times. What would be the probability distribution of a variable X, which represents the number of tails in four tosses of coin ? What kind of probability distribution is this ? Justify. 5
- (c) What is Big data ? Explain the characteristics of Big data. 5
- (d) What is the map-reduce paradigm ? Explain with the help of an example. 5
- (e) Define the term similarity. Explain the Jaccard similarity of sets with the help of an example. 5
- (f) What is a data stream ? How is data stream processing different from transaction processing ? 5
- (g) Write a program using R to add and subtract two matrices. 5
- (h) Define the term classification. Write the steps about how R programming can be used to perform classification. 5

P. T. O.

2. (a) What are the different measures for defining the spread of a quantitative variable ? Explain. 5
- (b) What is meant by 'Sampling Distribution' ? Explain with the help of an example. 5
- (c) Explain the concept of data pre-processing and data cleaning with the help of an example for each. 5
- (d) What is a box plot ? Draw a sample box plot and explain it. 5
3. (a) Explain the characteristics of HDFS. What are the advantages of HDFS over file allocation table (FAT) based system ? 5
- (b) What is the role of shuffling and sorting in map-reduce ? Explain with the help of an example. 5
- (c) What are the features of SPARK architecture ? Explain. 5
- (d) What are NoSQL databases ? How are they different from relational database system ? 5

P. T. O.

4. (a) What is collaborative filtering ? Explain with the help of an example. 5
- (b) What is the use of Bloom filtering ? Explain with the help of an example. 5
- (c) Explain any page ranking algorithm with the help of an example. 5
- (d) Explain the issues related to mining of social networks. 5
5. (a) Define the following in the context of R programming, with the help of an example : 5
  - (i) Complex data type
  - (ii) % % operator
  - (iii) ← or << – operator
  - (iv) Vectors
  - (v) Strings
- (b) What is the use of bar chart ? How can you draw a bar chart using R programming language ? Explain. 5
- (c) What is linear regression ? Explain how R programming can be used to perform linear regression. 5
- (d) What is clustering ? Write the steps of how R-programming can be used to perform K-mean clustering. What is confusion matrix in this context ? 5

MCS-226