

No. of Printed Pages : 4

MCS-226

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

Term-End Examination

December, 2025

MCS-226 : DATA SCIENCE AND BIG DATA

Time : 3 Hours

Maximum Marks : 100

Weightage : 70%

***Note :** Question No. 1 is compulsory. Attempt any **three** questions from the rest.*

1. (a) What are the basic methods of data analysis in a Data Science ? Briefly explain each type. 5
- (b) Explain Binomial Distribution with suitable equation and example. 5
- (c) What is a data curation ? Explain the different steps of data curation. 5

- (d) Explain Graph Database. What are the differences between the graph databases–Neo4j and OrientDB ? 5
- (e) What is a Hamming distance measure ? How does it differ from cosine measure ? 5
- (f) Explain PageRank. Justify the statement, “PageRank is trustworthy to retrieve legitimate Webpages”. 5
- (g) What are the lists in R programming ? Give characteristics of lists. 5
- (h) What is the purpose of RMySQL package in R ? What are the packages in R that can scrape data from the web ? 5
2. (a) What is Data Reduction ? What are the methods of Data Reduction ? 5
- (b) What is Heat Map ? How does Heat map differ from Scatter Plot ? Explain the utility of Heat Map in the Data Science. 10

- (c) Differentiate between Apache Hadoop1 and Apache Hadoop2, with suitable explanation for each. 5
3. (a) What is a Big Data System ? What are the components of Big Data System ? 5
- (b) What is MapReduce ? Write the steps for execution of MapPhase and ReducePhase. 5
- (c) What do you understand by Predictive Analysis ? List down the various techniques for the Predictive Analysis. Compare supervised learning and unsupervised learning for predictive analysis, with suitable illustrations. 10
4. (a) Write down the syntax in R to print the multiplication table of 2. 5
- (b) What is HBase ? Explain the components in H-Base with diagram. 5

- (c) What is the Bloom Filter ? What is the need for Bloom Filter ? Discuss the working of Bloom Filter. Explain an illustrative example of Bloom Filter. 10
5. Write short notes on the following : $4 \times 5 = 20$
- (a) Logistic Regression
 - (b) K-means clustering
 - (c) Matrices in R
 - (d) Histogram
 - (e) Entropy *vs.* Information Gain

× × × × ×