

No. of Printed Pages : 4

**MCS–226**

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

**Term-End Examination**

**June, 2024**

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

*Time : 3 Hours*

*Maximum Marks : 100*

*(Weightage : 70%)*

---

***Note :** Question No. 1 is compulsory and carries 40 marks. Attempt any **three** questions from the rest.*

---

---

1. (a) Explain the following types of data analysis : 8
- (i) Descriptive Analysis
  - (ii) Exploratory Analysis
  - (iii) Inferential Analysis
  - (iv) Predictive Analysis

**P. T. O.**

- (b) What is meant by ‘Conditional Probability’ ? Explain with the help of a diagram. Also, explain the Bayes theorem. 6
- (c) What is NoSQL ? Explain the features of NoSQL databases. 6
- (d) Define the term similarity. Explain the similarity of documents with the help of an example. 6
- (e) What is meant by Link Analysis ? Explain the rank computation using MapReduce. 6
- (f) Explain the following, with the help of an example, in the context of R-programming : 8
- (i) Strings
  - (ii) Matrices
  - (iii) Bar charts
  - (iv) Scatter plots
2. (a) What are the common misconceptions of data science ? Explain the different stages of data science life cycle with the help of a diagram. 8

- (b) What is the concept of random variable ? Explain the characteristics of binomial and normal distribution. 6
- (c) List the key characteristics of various types of plots for data visualization. Also, define the different methods for collecting, analysing and interpreting the numerical information. 6
3. (a) Explain four V's in the context of big data with the help of an example. 8
- (b) Explain shuffling and sorting using the example of word count. 6
- (c) What is Graph based NoSQL ? Explain when do we need graph database with the help of an example. 6
4. (a) What is meant by Page Rank ? Explain the use of page rank in search engines with the help of an example. 5
- (b) Differentiate between database management system and data stream management system with the help of a diagram. Also, discuss the issues and challenges of data stream. 5

- (c) Explain the use of Bloom filtering with the help of an example. 5
- (d) What is Web Analytics ? Explain the issues in online advertising. 5
5. Explain and write the steps of how R-programming can be used for the following :
- 5×4=20
- (i) Logistic Regression
  - (ii) K-mean Clustering
  - (iii) Random Forest
  - (iv) Association Rules
  - (v) Time Series Analysis