

Total number of printed pages-3

3 (Sem-6/CBCS) CSC HE 4

2024

COMPUTER SCIENCE

(Honours Elective)

Paper : CSC-HE-6046

(Data Mining)

Full Marks : 60

Time : Three hours

**The figures in the margin indicate
full marks for the questions.**

1. Answer the following : 1×7=7
- (a) Define data mart.
 - (b) Define OLAP.
 - (c) What is data cube ?
 - (d) State the application of cluster analysis.
 - (e) What is KDD in data mining ?
 - (f) What are different tasks of data mining ?
 - (g) What is classification ?

Contd.

2. Answer the following questions : $2 \times 4 = 8$

- (a) Explain the difference between data mining and data warehousing?
- (b) What are different types of data mining?
- (c) What is clustering algorithm in data mining?
- (d) What are the characteristics of data in data warehouse?

3. Answer **any three** of the following questions : $5 \times 3 = 15$

- (a) Explain the basic workings of the apriori algorithm.
- (b) Define association rules in data mining and explain their significance.
- (c) Explain Euclidean distance, Cosine similarity, and Jacquard coefficient briefly.
- (d) Describe common pre-processing methods used in data mining to prepare data for analysis.
- (e) Define predictive and descriptive data mining techniques and provide examples of each.

4. Answer **any three** of the following questions : $10 \times 3 = 30$

- (a) Explain the significance of maximal frequent sets and closed frequent sets in association rule mining. Provide a brief overview of how they contribute to pattern discovery. $6 + 4 = 10$
- (b) Briefly explain the k-means, k-medoids, PAM, CLARA and CLARANS algorithms. $2 + 2 + 2 + 2 + 2 = 10$
- (c) Define the DBSCAN algorithm. Explain its approach to identifying clusters. $3 + 7 = 10$
- (d) Define scalability in the context of data mining algorithms. Identify one scalability challenge and propose a solution. $4 + 6 = 10$
- (e) Write a computer program to implement the pincer-search algorithm.
- (f) Describe the salient features of CURE clustering techniques.