(NO. OF PAGES: 2)

(FOR THE CANDIDATES ADMITTED 24UDA412

DURING THE ACADEMIC YEAR 2024ONLY) **REG.NO. :**

N.G.M.COLLEGE (AUTONOMOUS) : POLLACHI

END-OF-SEMESTER EXAMINATIONS :MAY-2026

BSC CS WITH DA **MAXIMUM MARKS: 75**

SEMESTER: IV **TIME : 3 HOURS**

PART - III

24UDA412– DATA MINING AND WAREHOUSING

SECTION – A **(10 X 1 = 10 MARKS)**

ANSWER THE FOLLOWING QUESTIONS.

1. Which of the following is an example of a data mining application?

A) Word Processing B) Financial Fraud Detection

C)Image Editing D) Operating System Design

2. Which measure is commonly used to compute similarity between two numeric vectors?

A) Hamming Distance B) Euclidean Distance

C) Chi-square Test D) Entropy

3. Which of the following is an OLAP operation?

A) Classification B) ClusteringC) Roll-up D) Regression

4. In the Apriori algorithm, candidate itemsets are generated using:_____

A) Divide and Conquer B) Greedy Strategy

C) Join and Prune Strategy D) Random Selection

5. In K-Means clustering, the number of clusters is:_____

A) Determined automatically B) Predefined by the user

C) Equal to the number of attributes D) Equal to the number of records

ANSWER THE FOLLOWING IN ONE (OR) TWO SENTENCES

6. Define Data Mining.

7. What is meant by noisy data?

8. What is ETL in data warehousing?

9. What is a frequent itemset?

10. Define hierarchical clustering. **(CONT…2)**

Ethical paper

SECTION – B (5 X 5 = 25 MARKS)

ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS.

11. a)Explain the different kinds of data that can be mined in data mining.

(OR)

b)Discuss various real-world applications of data mining with suitable examples.

12. a)Explain different types of attributes used in data mining with examples.

(OR)

b)Describe methods for handling missing values in data preprocessing.

13. a)Differentiate between an operational database system and a data warehouse.

(OR)

b)Explain the ETL process in data warehousing.

14. a)Explain the basic concept of frequent itemset mining.

(OR)

b)Describe the steps involved in decision tree induction.

15. a)Explain the working principle of the K-Means algorithm.

(OR)

b)Discuss the differences between agglomerative and divisive hierarchical clustering.

SECTION – C (5 X 8 = 40 MARKS)

ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS.

16. a)Explain the different kinds of patterns that can be discovered in data miningwith examples.

(OR)

b)Discuss the major issues and challenges in data mining andsuggest possible solutions to overcome them.

17. a)Explain the major tasks involved in data preprocessing. Discuss how data quality affects the mining results.**(OR)**

b)Explain different measures of similarity and dissimilarity for numeric and categorical data with examples.

18. a)Explain the architecture of a data warehouseand the role of each component.

(OR)

b)Describe the data cube and various OLAP operations used in multidimensional data analysis.

19. a)Explain the Apriori algorithm with a suitable example. Discuss how association rules are generated from frequent itemsets. **(OR)**

b) Compare Bayes classification and rule-based classification methods with examples.

20. a)Compare K-Means and K-Medoid clustering methods with its advantages and limitations.

(OR)

b)Explain various data mining applications in financial analysis, retail, and intrusion detection systems. Evaluate their impact on decision-making.

Ethical paper

****