

- Q.5 i. Define clustering. **3**
 ii. You have a dataset with the following data points and want to perform K-medoid clustering with K=2. Starting with two random initial medoids, perform one iteration of the K-medoid algorithm. Calculate the new medoids. **7**
 Data Points: (2, 3), (3, 2), (5, 8), (6, 6), (8, 9), (10, 10)
 Initial Medoids: M1(3, 2), M2(8, 9)
- OR iii. You have a dataset with the following data points in a 2D space: **7**
 Data Points: (2, 3), (3, 2), (5, 8), (6, 6), (8, 9), (10, 10). Perform K-means clustering with K=2. The distance function is Euclidean distance.
- Q.6 Attempt any two:
 i. What is the primary purpose of reporting and query tools and applications in a business context? **5**
 ii. Explain OLAP and types of OLAP. **5**
 iii. Describe basic analytical operations of OLAP. **5**
 (a) Roll-up (b) Drill-down
 (c) Slice and dice (d) Pivot (rotate)

Total No. of Questions: 6

Total No. of Printed Pages:4

Enrollment No.....



Faculty of Engineering
 End Sem Examination Dec-2023
 IT3ED02 Data Mining & Warehousing

Programme: B.Tech.

Branch/Specialisation: IT

Duration: 3 Hrs.

Maximum Marks: 60

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

- Q.1 i. In a star schema, what is the central table called? **1**
 (a) Fact table (b) Dimension table
 (c) Data mart (d) Master data table
- ii. What is the primary function of a data mart in a data warehousing environment? **1**
 (a) To store historical data
 (b) To serve as a temporary data storage area
 (c) To provide subject-specific data for a specific group or department
 (d) To manage real-time data processing
- iii. What is the primary goal of data mining? **1**
 (a) Data collection (b) Data analysis
 (c) Data storage (d) Data visualization
- iv. In data mining, what is "association rule mining" primarily used for? **1**
 (a) Predictive modelling
 (b) Clustering
 (c) Finding interesting relationships in data
 (d) Data cleansing
- v. What is the primary goal of the Apriori algorithm in data mining? **1**
 (a) Clustering (b) Classification
 (c) Association rule mining (d) Regression analysis

[2]

- vi. What do you mean back propagation? **1**
 (a) It is the transmission of error back through the network to adjust the inputs.
 (b) It is the transmission of error back through the network to allow weights to be adjusted so that the network can learn.
 (c) It is another name given to the curvy function in the perceptron
 (d) None of these
- vii. What is clustering? **1**
 (a) A supervised learning technique
 (b) A type of unsupervised learning technique
 (c) A regression technique
 (d) An anomaly detection method
- viii. In K-means clustering, what does "K" represent? **1**
 (a) The total number of data points
 (b) The number of clusters to be created
 (c) The total variance of the data
 (d) The distance between data points
- ix. MOLAP stands for- **1**
 (a) Multidimensional Operational Analytic Processing
 (b) Multidimensional Online Analytical Processing
 (c) Mining Online Analytical Program
 (d) Mining Operational Analytical Processing
- x. What is the primary purpose of reporting and query tools in data mining? **1**
 (a) Data storage (b) Data visualization
 (c) Data collection (d) Data pre-processing
- Q.2 i. What is the primary purpose of a data mart? **2**
 ii. Explain the concept of data integration in the context of data warehousing. Why is it important? **3**
 iii. Discuss the ETL (Extract, Transform, and Load) process in data warehousing. What challenges might you encounter during this process, and how can they be addressed? **5**
- OR iv. Define star and snowflake schema for multidimensional database with diagram. **5**

[3]

- Q.3 i. Define knowledge discovery process. **3**
 ii. Explain data mining architecture. Write some application of data mining. **7**
- OR iii. Describe challenges to data mining regarding data mining methodology and user interaction issues. **7**
- Q.4 i. What is Apriori algorithm? **2**
 ii. With the help of decision tree find means of predicting which company profiles will lead to an increase or decrease in profits based on the following data: **8**

| Age | Competition | Type | Profit |
|-----|-------------|----------|--------|
| Old | Yes | Software | Down |
| Old | No | Software | Down |
| Old | No | Hardware | Down |
| Mid | Yes | Software | Down |
| Mid | Yes | Hardware | Down |
| Mid | No | Hardware | Up |
| Mid | No | Software | Up |
| New | Yes | Software | Up |
| New | No | Hardware | Up |
| New | No | Software | Up |

Profit is class attribute.

- OR iii. Find all frequent item sets or frequent patterns in the following database using FP-growth algorithm. **8**

| T_ID | ITEMS |
|------|------------|
| T1 | b, d, c, a |
| T2 | e, d, c |
| T3 | a, b |
| T4 | a, c, d |
| T5 | f, g, d, b |

Take minimum support as 2.