

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.  
2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8 = 50, will be treated as malpractice.



USN

--	--	--	--	--	--	--	--	--	--

10IS74

**Seventh Semester B.E. Degree Examination, July/August 2022**  
**Data Warehousing and Data Mining**

Time: 3 hrs.

Max. Marks:100

**Note: Answer any FIVE full questions, selecting atleast TWO questions from each part.**

**PART - A**

- 1 a. What is ETL? Explain the functions of ETL. (10 Marks)  
b. Define Data Warehousing. Name the modeling technique used in Data Warehouse. (04 Marks)  
c. Suppose a data warehouse consists of the 3 dimensions time, doctor and patient and two measures count and charge, where charge is the fee that a doctor charges a patient for a visit.  
i) Draw a schema diagram for the Data Warehouse.  
ii) Starting with the base cuboid [day , doctor , patient]. What specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2004? (06 Marks)
- 2 a. What are the FASMI characteristics and CODD's OLAP characteristics? Explain in brief. (10 Marks)  
b. Describe Data Cube Operations with example. (10 Marks)
- 3 a. Explain the different types of attributes and give examples for each type of attribute. (10 Marks)  
b. What is Sampling? List the sampling approaches. (04 Marks)  
c. For the following vectors X and Y, calculate i) Cosine ii) Correlation iii) Euclidean distance measures.  $X = (0, -1, 0, 1)$   $Y = (1, 0, -1, 0)$ . (06 Marks)
- 4 a. Define Association Rule. Give example and what are its strength. (04 Marks)  
b. Explain with an example, the Apriori principle. (06 Marks)  
c. Explain Objective measure used to evaluate the quality of Association patterns – their strengths and limitations. (10 Marks)

**PART - B**

- 5 a. Illustrate with an example, how decision tree helps in classification. (10 Marks)  
b. Explain Nearest – Neighbour classifier and explain the Algorithm. (10 Marks)
- 6 a. What is Bayes theorem and show how it is used as the basis of the Naïve Bayes method? Describe the Naïve Bayes method. (10 Marks)  
b. Explain Evaluation criteria for Classification methods. (10 Marks)
- 7 a. Explain with an example, K – means algorithm and write the K – means algorithm and what are its limits. (10 Marks)  
b. What is Hierarchical Clustering Method? What are the methods used for computing the distance between cluster in Hierarchical method with the steps involved in Agglomerative method. (10 Marks)
- 8 Write short notes on :  
a. Web Content Mining  
b. Text Missing.  
c. Text Clustering  
d. Temporal Data Mining Task. (20 Marks)

CMRIT LIBRARY  
BANGALORE - 560 037

\*\*\*\*\*