III B. Tech I Semester Regular Examinations, February-2022
DATA WAREHOUSING AND DATA MINING
(Computer Science and Engineering)

Time: 3 hours   Max. Marks: 75

Answer any **FIVE** Questions **ONE** Question from **Each unit**
All Questions Carry Equal Marks
*****

UNIT-I

1. a) Compare OLAP & OLTP systems. [8M]
   b) Illustrate indexing methods used for OLAP data. [7M]

(OR)

2. a) Explain data cube computation. What is the need for partial materialization? [8M]
   b) Suppose that a data warehouse consists of the three dimensions time, doctor and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit. Draw a Snowflake schema for this data warehouse. 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 2010? [7M]

UNIT-II

3. a) Explain the process of knowledge discovery. [8M]
   b) Discuss applications of data mining. [7M]

(OR)

4. a) Summarize visualization techniques. [8M]
   b) Illustrate data discretization techniques. [7M]

UNIT-III

5. a) Explain apriori algorithm with an example for mining frequent item sets. [8M]
   b) Explain market basket analysis. [7M]

(OR)

6. a) Describe how to mine multidimensional association rules. [8M]
   b) Construct an FP-tree for the dataset given below: [7M]

Tran-ID	List of item_IDs	Tran-ID	List of item_IDs
T_1	I_1, I_2, I_5	T_6	I_2, I_3
T_2	I_2, I_4	T_7	I_1, I_3
T_3	I_2, I_3	T_8	I_1, I_2, I_3, I_5
T_4	I_1, I_2, I_4	T_9	I_1, I_2, I_3
T_5	I_1, I_3	T_{10}	I_1, I_2, I_3, I_4, I_5

UNIT-IV

7. a) Explain basic algorithm for inducing a decision tree from training samples. [8M]

b) Distinguish supervised learning from unsupervised learning. [7M]

(OR)

8. a) Explain Naïve Bayesian classification technique with example. [8M]

b) Distinguish between Lazy learners and Eager learners. [7M]

UNIT-V

9. a) Illustrate K-medoids algorithm. [8M]

b) Explain DBSCAN algorithm used for clustering. [7M]

(OR)

10. a) Compare hierarchical clustering methods. [8M]

b) Explain how to compare the clusterings generated by different methods. [7M]

*****