Code No: R1931051 **R19** **SET - 1**

III B. Tech I Semester Supplementary Examinations, Dec/Jan-2022-23

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) Differentiate between OLAP and OLTP. [8M]
   b) Explain about concept hierarchy with an example. [7M]

(OR)

2. a) Define data warehouse. Give its advantages over databases and mention its characteristics. [8M]
   b) Compare and contrast star schema and snowflake schema. [7M]

UNIT-II

3. a) Elaborate the Knowledge discovery process with a neat diagram. [8M]
   b) Describe the role of data cleaning and dimensionality reduction in data preprocessing. [7M]

(OR)

4. a) Focus on various similarity and dissimilarity measures. [8M]
   b) Outline the concept of principle component analysis. [7M]

UNIT-III

5. Consider the following dataset, find frequent item sets and generate association rules for them. [15M]

TID	items bought
T100	I1,I2,I5
T200	I2,I4
T300	I2,I3
T400	I1,I2,I4
T500	I1,I3
T600	I2,I3
T700	I1,I3
T800	I1,I2,I3,I5
T900	I1,I2,I3

minimum support count is 2 and minimum confidence is 50%.

(OR)

6. a) "Constraint based frequent pattern mining is used for mining in multi-dimensional space". Justify with an example. [8M]
   b) Write short notes on pattern evaluation methods [7M]

UNIT-IV

7. a) Outline the concepts of Gain ratio and Gini index. [8M]
   b) Explain about Bayes theorem. Give its role in classification. [7M]

(OR)

8. a) Briefly explain about support vector machines. Give its variants [8M]
   b) Discuss about decision tree induction algorithm with an example. [7M]

UNIT-V

9. a) Discuss about k-medoids algorithm. Compare it with k-means algorithm.. [8M]
   b) Mention the requirements for cluster analysis [7M]

(OR)

10. What is an outlier? Give its impact on cluster formation. Discuss about various outlier detection methods. [15M]

*****