

**DATA MINING AND WAREHOUSING**

Paper : 6.1.2

Full Marks : 80

Time : Three hours

**The figures in the margin indicate full marks for the questions.**

1. Write true/false of the following :  $1 \times 5 = 5$ 
  - i) Knowledge Discovery in Database (KDD) process is a step in Data Mining.
  - ii) A snowflake schema consists of more than one Fact Table and multiple dimension tables.
  - iii) Drill-up operation deals with switching from an aggregated level to an detailed level within the same classification hierarchy.
  - iv) Divisive clustering techniques take the opposite approach from agglomerative techniques.
  - v) Decision tree methods are able to handle both numerical and categorical attributes.
2. Answer the following questions :  $3 \times 7 = 21$ 
  - i) What are the different application areas of data mining ?
  - ii) How is data mining different from KDD ?
  - iii) Write the difference between R-OLAP and M-OLAP.
  - iv) Describe the concepts of frequent set, maximal frequent set and border set.
  - v) What is supervised and unsupervised learning ?
  - vi) Define the terms — Categorical, Original and Ratio-scaled variables.
  - vii) Distinguish between hierarchical and partitioning clustering.
3. Answer the following questions :  $2 \times 5 = 10$ 
  - i) Define the term dimension modelling.
  - ii) Distinguish between star schema and snowflake schema.
  - iii) What is spatial data mining ?
  - iv) What is centroid ?
  - v) Discuss the advantages of the decision tree method.
4. What is OLAP operation ? Define the following terms — Slicing, Drilling, Dicing and Drill-up.  $2 + 8 = 10$
5. Define support and confidence in transaction. What is upward and downward closure property of item sets ?  $4 + 4 = 8$
6. **Explain the following terms in the context of DBSCAN : (any four)**  $4 \times 3 = 12$ 
  - i) E-neighbourhood
  - ii) Core object
  - iii) Directly density reachable
  - iv) Density reachable
  - v) Density connected.
7. Describe the working of PAM algorithm. 6
8. Write short notes on : (any two)  $4 \times 2 = 8$ 
  - i) Entropy
  - ii) Gini index
  - iii) ID 3
  - iv) Splitting criterion