Code: CS6T4

III B.Tech-II Semester–Regular/Supplementary Examinations–March 2018

DATA WAREHOUSING AND DATA MINING
(COMPUTER SCIENCE & ENGINEERING)

Duration: 3 hours   Max. Marks: 70

PART – A

Answer *all* the questions.   All questions carry equal marks

11x 2 = 22 M

1. a) What is data mart?

   b) What is a data cube?

   c) List the steps in the knowledge discovery process.

   d) What is data integration?

   e) Define support and confidence measures related to Association rule mining.

   f) What are the various attribute selection measures in the classification?

   g) What are the requirements for cluster analysis?

   h) What is the aim of probabilistic hierarchical clustering?

   i) What are Outliers?

   j) List the categories of sequence data.

   k) What is star schema in multidimensional data model?

PART – B

Answer any *THREE* questions. All questions carry equal marks.

$$3 \text{ x } 16 = 48 \text{ M}$$

2. a) Differentiate between OLTP and OLAP.   8 M

   b) What are various schemas for multidimensional data
      models?   8 M

3. a) What are the major issues in data mining?   8 M

   b) In real world data tuples with missing values for some
      attributes are common occurrence. Describe various
      methods for handling this problem.   8 M

4. a) How can you find frequent itemsets using candidate
      generation?   8 M

   b) Explain the classification by decision tree induction. 8 M

5. a) What is meant by clustering? Explain the partitioning
      methods with an example.   8 M

   b) What are the algorithms used in grid –based methods?
      Explain.   8 M

6. a) What is meant by outlier Analysis? What are various
   methods for finding outliers?   8 M

   b) Write detailed notes on regression and trend analysis in
   time-series data.   8 M