Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 11169

B.E./B.Tech. DEGREE EXAMINATION, MAY/JUNE 2014.

Seventh Semester

Computer Science and Engineering

080230056 — DATA WAREHOUSING AND DATA MINING

(Regulation 2008)

Time : Three hours Maximum :100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. What is Multidimensional OLAP?

2. Given 3 dimensions state how many cuboids may be generated in a multidimensional OLAP environment.

3. Comment on replacing missing value using bin means.

4. Give three commonly used statistical measure for the characterization of data dispersion.

5. Prove that all nonempty subsets of a frequent item set must also be frequent.

6. What is FP tree? Give its significance.

7. Define precision and recall.

8. How Baye's theorem may be used for classification?

9. List the features that are inherent features of data a spatial database.

10. How is similarity between text documents determined?

PART B — (5 × 16 = 80 marks)

11. (a) Explain that schema that suits best for holding detailed data in a data-warehousing environment. (16)

Or

(b) Use the various components of the multidimensional warehouse architecture with a suitable diagram. (16)

12. (a) Use a flowchart to summarize the following procedures for attribute subset selection: (16)

(i) Stepwise forward selection

(ii) Stepwise backward elimination

(iii) A combination of forward selection and backward elimination.

Or

(b) Discuss about the various discretization techniques for numerical data with suitable example. (16)

13. (a) Explain how to mine multilevel association rules from Relational Database Environment. (16)

Or

(b) A database has five transactions. Let min sup = 60% and min con f = 80%. Find all frequent item sets using Apriori. (16)

TID	items bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

14. (a) What are decision trees? Explain hoe decision trees may be derived using decision tree induction algorithm. (16)

Or

(b) Explain k-means and medoids clustering algorithms and compare them. (16)

15. (a) Explain the various web mining types with suitable example. (16)

Or

(b) Discuss about mining from text database and the issues related with such approaches. (16)