[07 - 4222]

IV/IV B.TECH. DEGREE EXAMINATION

Computer Science and Engineering

Second Semester

ELECTIVE - II : DATA WAREHOUSING AND DATA MINING

(Effective from the admitted batch of 2006-2007)

Time : 3 hours   Max. Marks : 70

First question is compulsory.
Answer any FOUR from the remaining questions.
All questions carry equal marks.
Answer all parts of any question at one place.

1. Briefly discuss:

   (a) Data transformation

   (b) Snowflake schema

   (c) Metadata Repository

   (d) Frequent Pattern growth

   (e) Gini index

   (f) Rough set approach

   (g) Conceptual clustering

2. (a) Describe the data mining functionalities.

   (b) Discuss about different data clearing methods.

3. (a) What is a data warehouse? Write the differences between operational database systems and data warehouses.

1   (P.T.O.)

[07 - 4222]

(b) Explain how to index OLAP data by bitmap indexing and join indexing.

4. (a) Describe the star-cubing algorithm for computing iceberg cubes.

(b) What is attribute-oriented induction? Discuss about its efficient implementation.

5. (a) Explain the algorithm for mining frequent itemsets without candidate generation. Give relevant example.

(b) Describe the Apriori property and write the Apriori algorithm to find frequent itemsets using an iterative level-wise approach based on candidate generation.

6. (a) Discuss about constraint-based association mining.

(b) Describe a basic algorithm for inducing a decision tree from training tuples.

7. Explain the following:
(i) Naive Bayesian classifier
(ii) K-Nearest-Nieghbor classifiers.

8. (a) Explain Bagging and Boosting strategies for improving classifier and predictor accuracy.

(b) Briefly explain K-means and K-medoids partitioning methods of clustering.

* * * * *