B. Tech.
(SEM. VII) EXAMINATION, 2008-09
DATA MINING AND DATA WAREHOUSING

Time : 3 Hours] [Total Marks : 100

Note : Attempt all questions.

1 Attempt any four parts : 5×4=20

(a) What do you mean by data mining? Differentiate between data mining technique and data mining strategy.

(b) Define the term data cleaning with suitable example.

(c) Write short notes on the following:

(i) Data mining metrics
(ii) Social implications of data mining.

(d) Define KDD. Identify and describe the phases in the KDD process.

(e) Differentiate between the following:

(i) Data warehouse and operational database.
(ii) Intrinsic value and actual value.

(f) Write short note on dimensionality reduction.

0105] [Contd...
Attempt any two parts:

(a) (i) Define the terms data generalization and analytical characterization with example.

(ii) Given the following set of values \{1, 3, 9, 15, 20\}, determine the jackknife estimate for both the mean and standard deviation of the mean.

(b) (i) Describe mining association rules in large databases.

(ii) Data quality can be assessed in terms of accuracy, completeness and consistency. Propose two other dimensions of data quality.

(c) Describe the following:

(i) Mining single dimensional Boolean association rules from transactional databases.


Answer any two parts:

(a) Describe various issues regarding classification and prediction.

(b) Explain the algorithm for classification by decision tree induction.

(c) Write short note on the following:

(i) Bayesian classification

(ii) Back propagation algorithm.
(c) (i) What do you mean by clustering? Explain. Also give data types in cluster analysis.
(ii) Write short note on divisive hierarchical clustering.

4 Answer any two parts: 10 \times 2 = 20

(a) What do you mean by Data Warehouse? Discuss its key features with suitable examples.

(b) (i) Briefly describe the multi dimensional data model.
(ii) Describe data cubes with suitable example.

(c) Describe the following with example:
(i) Concept hierarchy
(ii) 3 Tier Architecture

5 Answer any four of the following: 5 \times 4 = 20

(a) OLAP functions and tools.
(b) MOLAP Vs. HOLAP
(c) Data mining interface
(d) Backup and recovery
(e) Testing data warehouse
(f) Efficient Processing of OLAP queries.