



Printed Pages : 3

TCS-31

(Following Paper ID and Roll No. to be filled in your Answer Book)

PAPER ID : 0105

Roll No.

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

- What do you mean by data mining ? Differentiate between data mining technique and data mining strategy.
- Define the term data cleaning with suitable example.
- Write short notes on the following :
 - Data mining metrics
 - Social implications of data mining.
- Define KDD. Identify and describe the phases in the KDD process.
- Differentiate between the following :
 - Data warehouse and operational database.
 - Intrinsic value and actual value.
- Write short note on dimensionality reduction.



2 Attempt any two parts : $10 \times 2 = 20$

- (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.
- (ii) The Apriori algorithm : Finding frequent item sets using candidate generation.

3 Answer any two parts : $10 \times 2 = 20$

- (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.

