

7E5099

Roll No. _____

[Total No. of Pages : 2]

7E5099

B. Tech. VII Semester (Main) Examination, Nov-Dec-2011

Information Technology

7IT3 Data Mining & Ware Housing

Time : 3 Hours

Maximum Marks : 80

Min. Passing Marks : 24

Instructions to Candidates:

Attempt any **five questions** selecting **one question from each unit**. All questions carry **equal marks**. Schematic diagrams must be shown wherever necessary. Any data you feel missing suitably be assumed and stated clearly. Units of quantities used/calculated must be stated clearly.

Unit - I

1. a) What do you mean by data reduction and what are the different process of data reduction ? (6)
- b) How do you fill the missing values of data. (4)
- c) Explain data transformation methods. (6)

OR

1. a) What do you mean by datamining ? Briefly explain knowledge discovery as a process. (8)
- b) Suppose a group of 12 sales price records has been sorted as follows : (8)
5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215 partition them into three bins by each of the following methods :
 - i) Equal - frequency (equidepth) partitioning
 - ii) Equal - width partitioning
 - iii) Clustering

Unit - II

2. a) Describe in detail the Apriori Algorithm. (8)
- b) Assume the user - specified minimum support is 50 %. The transaction database shown below : (8)

Tid	A	B	C	D	E	F
T1	1	0	1	1	0	0
T2	0	1	0	1	0	0
T3	1	1	1	0	1	0
T4	0	1	0	1	0	1

Apply apriori algorithm to generate frequent item set on the above transaction.

OR

2. a) Explain the following terms with suitable example
i) Support
ii) Confidence
iii) Association rule
iv) IQR
v) Percentile (8)
b) Explain 'AR' algorithm with suitable example. (8)

Unit - III

3. a) What are advantages and disadvantages of decision tree approach over other approaches of data mining. (6)
b) Describe DBSCAN method of cluster generation. (6)
c) Explain the requirements of clustering in datamining. (4)

OR

3. a) Explain Back propagation algorithm with suitable example. (12)
b) Briefly explain Neural network approach. (4)

Unit - IV

4. a) Describe the process architecture of datawarehouse. (8)
b) Write the difference between OLTP and OLAP. (5)
c) How database is different from datawarehouse. Give an example. (3)

OR

4. a) Explain multidimensional data model. (8)
b) Write short notes on : (8)
i) Schema
ii) Dimension
iii) Fact
iv) Dimension table

Unit - V

5. a) Describe datawarehouse recovery models. (5)
b) Define backup and recovery strategy. (5)
c) Explain the tuning steps in datawarehouse. (6)

OR

5. a) Explain the testing terminologies in brief. (6)
b) Write short notes on (any four) (10)
i) Security
ii) Disk configuration
iii) Fixed queries
iv) Aggregation
v) Adhoc queries