Name : …………………………………………………………

Roll No. : …………………………………………………………

Invigilator's Signature : ………………………………………

CS/M.Tech(CSE)/SEM-1/MCSE-105C/2011-12
2011
DATA MINING & DATA WAREHOUSING

Time Allotted : 3 Hours   *Full Marks* : 70

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

Answers should be brief and to the point and may be supplemented with neat sketches.

GROUP – A
Answer any *seven* of the following.   7 × 2 = 14

1. "Metadata is a bridge between the data warehouse and the decision support application". Explain.

2. Compare between a priory algorithm and FP tree algorithm.

3. What do you mean by pruning ?

4. Why is star schema preferred than Snowflake schema ?

5. Compare between ROLAP and MOLAP.

6. Describe the Slicing and Drilling operations with suitable examples.

40326   [Turn over

7. Compare between supervised and unsupervised learning.

8. Define Entropy and Gini index.

9. Compare between support and confidence.

10. Define the following terms with example : 2 × 1

   a) Density reachable

   b) Density connected.

GROUP – B

Answer any *four* of the following. 4 × 14 = 56

11. a) Draw the decision tree for 'Tenured' of the following transaction database through the statistical approach and finally set the rules. 7

Name	Rank	Years	Tenured
Mike	Assistant Prof.	≤ 6	No
Mary	Assistant Prof.	> 6	Yes
Bill	Professor	≤ 6	Yes
Jim	Associate Prof.	> 6	Yes
Dave	Assistant Prof.	≤ 6	No
Anne	Associate Prof.	≤ 6	No
Tom	Assistant Prof.	≤ 6	No
Merlisa	Associate Prof.	> 6	Yes
George	Professor	≤ 6	Yes
Joseph	Assistant Prof.	> 6	Yes

   b) Apply the Bayesian classification method on the above example and predict the given tuple X in which class.

   $X = \{$ Name=Jeff; Rank=Professor; Year=4 $\}$. 7

12. a) Illustrate the working principle of DIC algorithm the following transaction database :   7

TID	A	B	C	D	E	F	G	H	I
1	1	0	0	0	1	1	0	1	0
2	0	1	0	1	0	0	0	1	0
3	0	0	0	1	1	0	1	0	0
4	0	1	1	0	0	0	0	0	0
5	0	0	0	0	1	1	1	0	0
6	0	1	1	1	0	0	0	0	0
7	0	1	0	0	0	1	1	0	0
8	0	0	0	0	1	0	0	0	0
9	0	0	0	0	0	0	0	1	0
10	0	0	1	0	1	0	1	0	0
11	0	0	1	0	1	0	1	0	0
12	0	0	0	0	1	1	0	1	0
13	0	1	0	1	0	1	1	0	0
14	1	0	1	0	1	0	1	0	0
15	0	1	1	0	0	0	0	0	1

Assume σ = 50%. Where σ is the minimum support, A to I be the data items and 1 to 15 be the fifteen transactions in the database.

b) Construct the FP tree for the above database.   7

13. a) What specific OLAP operation should be performed for data cube [date, spectator, location and game and a measure charge] in order to list the total fees collected by each spectator on 20. 4. 2004 ?   2

b) Explain the term 'data warehouse'.   4

c) Briefly explain the architecture of a warehouse.   5

d) Compare between OLAP and OLTP.   3

14. a) Define a Border Set. 2

 b) Describe an algorithm to generate the Border Set. 4

 c) Explain the ε-Neighbourhood and Mean-points of an object. 2

 d) Write short note on DBSCAN. 6

15. Write short notes on any *two* of the following : 2 × 7

 a) Collaborative filtering

 b) Knowledge discovery in database

 c) Genetic algorithm.

16. a) Write an level-wise algorithm to find all the frequent set and say that your algorithm is supervised or unsupervised learning. 8

 b) Compare between DBMS and Data Mining. 6

———————