

Name : .....

Roll No. : .....

Invigilator's Signature : .....

**CS/M.Tech (CSE)/SEM-2/CST-623/2011**

**2011**

**DATA MINING & KNOWLEDGE MANAGEMENT**

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.*

**GROUP – A**

Answer Question No. 1 and any *two* from the rest.

1. a) Suppose that you have data describing the closing prices of the stock you own for the last 1000 days. Suppose you are interested in generating all rules which tell you about chances of your stock going up on a given day provided you know the pattern ( up or down ) on  $K$  preceding days, with some minimum support and minimum confidence defined. How would you model this problem as association rule mining problem ? Is there a way to represent this as transactions with binary attributes like in the supermarket case ?

- b) Why is data mart used ?

4 + 1

30088 (M.Tech.)

[ Turn over



2. a) Explain some factors that one must take into account when selecting data mining software.

b) Explain the steps of Apriori algorithm.

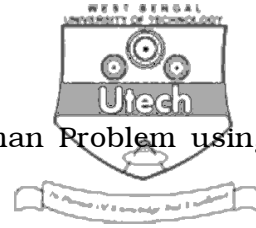
c) A database has five transactions. Let  $\text{min\_sup} = 60\%$  and  $\text{min\_conf} = 80\%$ .

TID	Items Bought
T100	M, O, N, K, E, Y
T200	D, O, N, K, E, Y
T300	M, A, K, E
T400	M, U, C, K, Y
T500	C, O, K, I, E

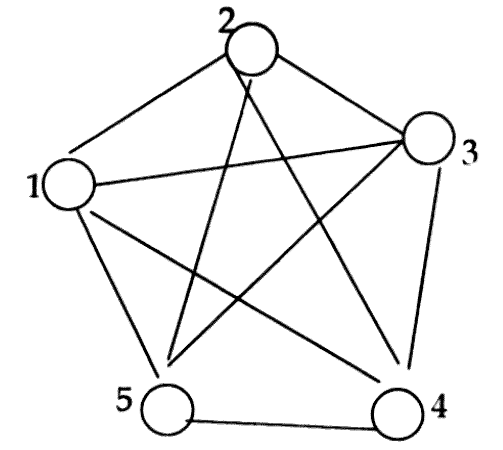
Use Apriori to find frequent item sets of first four transactions. Mark non-frequent items. 4 + 5 + 6



3. a) Suppose that a data warehouse consists of the four dimensions date, spectator, location and game and a measure charge. Charge is the fare that a spectator pays when watching a game on a given date. Spectators may be students, adults, or seniors with each category having its own charge rate.
- b) Draw a fact constellation schema diagram for the data warehouse.
- c) 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 12.1.2010.  $5 + 4 + 3 + 3$
4. a) What is the difference between Genetic Algorithm and Conventional Algorithm ?
- b) Discuss the different stages of Genetic Algorithm in Genetic Cycle.
- c) What are Crossover Rate and Mutation Rate ?



- d) Solve the following Travelling Salesman Problem using Genetic Algorithm.

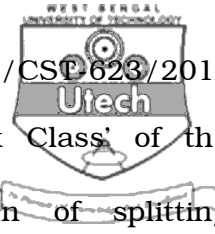


The distance matrix is given below :

Source ↑	1					
	2	0				
	3	8	11	0		
	4	5	1	4	0	
	5	6	12	9	13	0
		1	2	3	4	5
	Destination →					

$$3 + 3 + 2 + 7$$

5. a) What is the difference between supervised and unsupervised learning method ?
- b) What is the shortcoming of decision tree ?



- c) Construct the decision tree for 'Risk Class' of the following data set through selection of splitting attributes by Information Gain Computation or Gini Index Computation.

Own Home ?	Married	Gender	Employed	Credit Rating	Risk Class
Yes	Yes	Male	Yes	A	B
No	No	Female	Yes	A	A
Yes	Yes	Female	Yes	B	C
Yes	No	Male	No	B	B
No	Yes	Female	Yes	B	C
No	No	Female	Yes	B	A
No	No	Male	No	B	B
Yes	No	Female	Yes	A	A
No	Yes	Female	Yes	A	C
Yes	Yes	Female	Yes	A	C

4 + 3 + 8

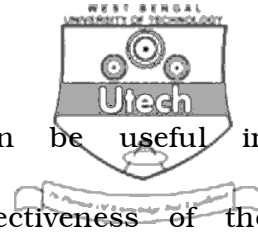
### GROUP – B

Answer Question No. 6 and any *two* from the rest.

6. a) Draw PTPO model for knowledge management.
- b) What is an activation function ? Explain the different types of activation function.

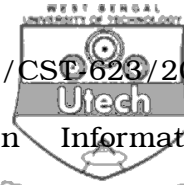
3 + 2

CS/M.Tech (CSE)/SEM-2/CST-623/2011



7. a) How Knowledge Management can be useful in increasing the efficiency and effectiveness of the Organization ?
- b) What is the physical significance of  $RKI = -1$  ?
- c) "If bits are transported, there is no need for transporting man, machine and material." In light of this justify the quantitative definition of RKI.
- d) What is the Relative Knowledge Index between two knowledge organization  $A$  and  $B$  when :
- i)  $A$  gives 10760 bits of info to  $B$  in 1 sec and  $B$  gives 102490 bits to  $A$  in 3 secs
- ii) Link between  $A$  and  $B$  is 58 Kbps
- iii) Bit value of  $A = \text{Re. } 0.96$ , that of  $B = \text{Re. } 0.106$
- e) In the above problem, how is the price of knowledge from  $A$  to  $B$  compared to that from  $B$  to  $A$  ?

$$3 + 2 + 5 + 3 + 2$$



8. a) Cite Claude-Shannon's equation on Information Entropy with proper and brief explanations on the theory of information.
- b) Quoting Stonier's equation on the information content of an organization, state whether this equation agrees with Claude-Shannon's entropy theory or not. Justify your answer.
- c) Briefly explain GNP and HDI with examples. 5 + 5 + 5
9. a) Mention, with explanations, the kinds of changes Masuda and Rada perceive as inevitable for a society transiting from an industry to an information-based socio-political and economic structures.
- b) "Explicit knowledge can be static or dynamic." Explain.
- c) Write down the innovative differences between Knowledge Management and Information Management. 8 + 2 + 5

=====