

4

Roll. No									
----------	--	--	--	--	--	--	--	--	--

**B.E/B.TECH. (FULL TIME) DEGREE END SEMESTER EXAMINATIONS, APRIL/MAY 2014.**

**COMPUTER SCIENCE AND ENGINEERING BRANCH**

**EIGHTH SEMESTER - (REGULATIONS 2004)**

**CS 514 – DATA WAREHOUSING AND DATA MINING**

**TIME: 3 HOURS**

**MAX. MARKS: 100**

**ANSWER ALL QUESTIONS**

**Part – A (10 x 2 = 20 Marks)**

- 1 What are the data mining task primitives?
- 2 What is metadata? Explain its types.
- 3 Distinguish between noise and outliers. Give suitable examples.
- 4 What are the advantages of dimensional modeling?
- 5 How do you perform discretizations of numeric attributes?
- 6 What is lazy learner? Give an example.
- 7 What is rule based classification? Give an example.
- 8 Difference between density and grid clustering.
- 9 What is precision and recall?
- 10 What are the applications of data mining?

**Part – B (5 x 16 = 80 Marks)**

11. (i) What is data mining? Explain in detail with the KDD process. (10)
- (ii) How data warehouses systems different from operational database systems and how are they similar? (6)
- 12 (a)
  - (i) What is data warehouse? Explain three-tier data warehouse architecture with a neat sketch. (8)
  - (ii) A data warehouse can be modeled by either a star schema or a snowflake schema. Briefly describe the similarities and the differences of the two models, and the analysis their advantages and disadvantages. Give which might be more empirically useful and state the reasons. (8)

**(OR)**

- 12 (b)
  - (i) How are concept hierarchies useful in OLAP? Explain OLAP operations in the multidimensional data model. (8)
  - (ii) Suppose that a data warehouse consists of the four dimensions data,

spectator, location, and game, and the two measures count and charge, where charge is the fare that a spectator pays when watching a game on a given data. Spectator may be students, adults, or seniors, with each category having its own charge rate.

(1) Draw a star schema diagram for the data warehouse. (4)

(2) Starting with the base cuboid [data, spectator, location, game], what specific OLAP operations should one perform in order to list the total charge paid by student spectators GM\_Place in 2000. (4)

13 (a) (i) List and discuss the steps for integrating a data mining system with a data warehouse. (10)

(ii) A data set for analysis includes only one attribute  $X = \{7, 12, 5, 18, 5, 9, 13, 12, 19, 7, 12, 12, 13, 3, 4, 5, 13\}$  (6)

(1) What is the mean and median of the data set X?

(2) Find the standard deviation for X.

(3) Find outliers in the data set X. Discuss the results.

(OR)

13 (b) (i) What is an association rule mining? Apply the FP-Growth algorithm to find the strong association rules for the transactional database given below. (12)

TID	ITEMS
01	A, B, C, D
02	A, C, D, F
03	C, D, E, G
04	A, D, F, B
05	B, C, G
06	D, F, G
07	A, B, G

Analyze misleading associations for the rule set obtained in the above transactional database.

(ii) What are the essential difference between association rules and decision rules? (4)

14 (a) (i) What are the prediction techniques supported by a data mining system. (4)

(ii) State the ID3 algorithm. Apply the same to construct a decision tree for the data given below. (12)

A	B	C	CLASS
15	1	A	C1
20	3	B	C2
25	2	A	C1
30	4	A	C1
35	2	B	C2
25	4	A	C1
15	2	B	C2
20	3	B	C2

(OR)

- 14 (b) The data mining task is to cluster the following eight points into three clusters: (16)  
 $A_1(2,10), A_2(2,5), A_3(8,4), B_1(5,8), B_2(7,9), B_3(6,4), C_1(1,2), C_2(5,9)$ . The distance function is Euclidean distance. Suppose initially we assign  $A_1, B_1$  and  $C_1$  as the center of each cluster, respectively. Apply the K-means algorithm to show the three cluster centers after the first round execution and the final three clusters.
- 15 (a)  
 (i) What is web mining? Differentiate between web content mining, web structure mining, and web usage mining. What is trend analysis? (12)  
 (ii) What are the differences between visual data mining and data visualization? (4)

(OR)

- 15 (b) What are the major challenges faced in bringing data mining related to market? Illustrate data mining issues that may have a strong impact on the market and society. Discuss how to approach such issue. (16)