



MANIPAL INSTITUTE OF TECHNOLOGY
 (Constituent Institute of Manipal University)
 MANIPAL-576104



VII SEMESTER B.TECH.(COMPUTER SCIENCE AND ENGINEERING) DEGREE
END-SEMESTER EXAMINATION-NOVEMBER/DECEMBER 2014
SUBJECT: DATAWAREHOUSE AND DATAMINING (CSE 433)
DATE: 26-11-2014

TIME: 3 HOURS

MAX.MARKS: 50

Instructions to Candidates

- **Note:** Answer any **FIVE** full questions.

1.A. Explain the following.

i)Wavelet transformation ii)Principal Component Analysis iii)Attribute subset selection.

1.B. How do you compute dissimilarity for attributes of mixed types.

1.C. Generate frequent item sets for the database shown in Table Q.1C.
 using Apriori. Use minimum support as 40%. ((1+2+1)+3+3)

2.A. “Strong association rules are not necessarily interesting” justify this statement with proper example.

2.B.Why do you require pruning in decision trees? Discuss the common approaches used for pruning.

2.C. List and explain any five important requirements of clustering in data mining. (2+3+5)

3.A. How do you measure clustering quality? Explain.

3.B. Suppose that a data warehouse consists of the four dimensions, date, 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 date. Spectators may be students, adults, or seniors, with each category having its own charge rate.

i. Draw a star schema diagram for the data warehouse.

ii. Starting with the base cuboid [date, spectator, location, game], what specific OLAP operation should one perform in order to list the total charge paid by student spectators at GM Place in 2004?

3.C. What is attribute oriented induction? Explain with an example working set how it is performed. (3+4+3)

4.A. Find frequent itemsets for database in Table Q.4A using FP-Growth. Minimum support count is 2.

4.B. For the neural network consisting of three inputs, one output and two hidden units give sample calculations for learning by the back propagation algorithm. Initial weights and bias values are shown in Table Q. 4B.

Table Q. 4B

x_1	x_2	x_3	w_{14}	w_{15}	w_{24}	w_{25}	w_{34}	w_{35}	w_{46}	w_{56}	θ_4	θ_5	θ_6
1	0	1	0.2	-0.3	0.4	0.1	-0.5	0.2	-0.3	-0.2	-0.4	0.2	0.1

(5+5)

5.A. Classify the unknown sample with $k=3$ using K-NN classifier.

<age <40; income = low; student = Yes, credit_rating=fair>

Training set is as shown in Table Q.5A.

5.B. Give the algorithm of k-medoid clustering. Compare it with k-Means.

5.C. Explain the following ensemble techniques.

i)Bagging ii)Boosting

(3+3+(2+2))

6.A. With necessary examples, explain the Iceberg Cube and Closed Cube options in Cube materialization.

6.B. Explain the three types of web mining.

((2+2)+6)

Table Q.1C.

Tid	Items	Tid	Items
1	1,2,3,4	6	1,2,3,4
2	1,2,3,4,5	7	1,3,4,5
3	1,2	8	3,4,5
4	1,3,4,5	9	3,4,5
5	1,2,3,4	10	2,5

Table Q. 4A.

Tid	Items	Tid	Items
1	1,2,5	6	2,3
2	2,4	7	1,3
3	2,3	8	1,2,3,5
4	1,2,4	9	1,2,3
5	1,3		

Table Q. 5A

Age	Income	Student	Credit_rating	Buys_comp
<40	low	no	fair	yes
>40	medium	yes	fair	yes
<40	high	yes	fair	yes
>40	low	no	excellent	no
<40	low	yes	excellent	no
<40	low	yes	fair	yes
<40	low	yes	fair	yes

GoodLuck