Total No. of Questions : 6]

SEAT No. :

P119

[Total No. of Pages : 2

Oct.-16/BE/Insem.- 177

B.E. (Computer Engineering) (Semester - I)

DATA MINING TECHNIQUES AND APPLICATIONS

(2012 Pattern) (Elective - I(d))

Time : 1 Hour] *[Max. Marks : 30*

Instructions to the candidates:

1) ***Answer Q1 or Q2, Q3 or Q4, Q5 or Q6.***
2) ***Neat diagrams must be drawn wherever necessary.***
3) ***Figures to the right side indicate full marks.***
4) ***Assume suitable data if necessary.***

Q1) a) Describe three challenges to data mining regarding data mining methodology. **[6]**

b) Consider the following group of data **[4]**

200, 300, 400, 600, 1000

i) Use the min-max normalization to transform value 600 onto the range [0.0,1.0].

ii) Use the decimal scaling to transform value 600.

OR

Q2) a) What are the major tasks in data preprocessing? Explain them in brief. **[6]**

b) Differentiate between **[4]**

i) Supervised and Semi-supervised learning.

ii) Classification and Regression.

iii) Descriptive and Predictive data mining tasks.

Q3) a) A database has five transactions: **[5]**

TID	Items
1	F,C,A,M,P
2	F,C,A,B,M
3	F,B
4	C,B,P
5	F,C,A,M,P

Assuming the support count 2, construct an FP-tree.

b) State the antimonotonicity property. **[2]**

c) Define k--itemset, support count and strong association rules. **[3]**

P.T.O.

OR

Q4) a) Differentiate between **[4]**

i) Multilevel and multidimensional associations

ii) Pattern-pruning and data-pruning constraints

b) A database has five transactions. Let minimum support is 60%. **[6]**

TID	Items
1	Butter, Milk
2	Butter, Dates, Balloon, Eggs
3	Milk, Dates, Balloon, Cake
4	Butter, Milk, Dates, Balloon
5	Butter, Milk, Dates, Cake

Find all the frequent item sets using Apriori algorithm. Show each step.

Q5) a) Explain the following terms **[6]**

i) Posterior probability

ii) Prior probability

iii) Class-conditional independence

b) Explain with neat diagram confusion matrix for a two class problem. **[4]**

OR

Q6) a) Consider the training examples shown in the table below for a binary classification problem. **[6]**

Al	A2	Class
T	T	Yes
T	T	Yes
T	F	No
F	F	Yes
F	T	No
F	T	No
F	F	No
T	F	Yes
F	T	No

i) Compute the information gain for A1.

ii) Compute the information gain for A2.

iii) What is the best split between A1 and A2 according to Information gain?

b) Define the following terms with respect to Classifier **[4]**

i) Precision ii) Recall iii) Accuracy iv) Misclassification rate.