USN | | | | | | | | | | |

06IS74

Seventh Semester B.E. Degree Examination, December 2010
Data Mining

Time: 3 hrs. Max. Marks:100

Note: ***Answer any FIVE full questions, selecting atleast TWO questions from Part – A and Part – B.***

PART – A

1 a. What is data mining? Explain the challenges that motivated the development of data mining. **(10 Marks)**

b. Explain the different types of data sets with examples. **(10 Marks)**

2 a. Describe the various approaches for feature selection. **(06 Marks)**

b. Explain with examples the following : i) Simple matching coefficient ii) Jacquard coefficient iii) Cosine similarity. **(06 Marks)**

c. Discuss the measures of proximity between objects that involve multiple attributes. **(08 Marks)**

3 a. What is classification? Explain the two classification models with example. **(06 Marks)**

b. Consider the training examples, shown in table Q3(b) for a binary classification problem.

i) What is the entropy of this collection of training examples with respect to the positive class?

ii) What are the information gains of a1 and a2 relative to these training examples? **(08 Marks)**

Table Q3(b)

Instance	a_1	a_2	a_3	Target class
1	T	T	1.0	+
2	T	T	6.0	+
3	T	F	5.0	-
4	F	F	4.0	+
5	F	T	7.0	-
6	F	T	3.0	-
7	F	F	8.0	-
8	T	F	7.0	+
9	F	T	5.0	-

c. Distinguish between Rule based ordering scheme and class based ordering scheme. **(06 Marks)**

4 a. A data base has four transactions. Let min_Sup = 40% and min_conf = 60%.

TID	DATE	ITEMS BOUGHT
100	01.01.01	{K, A, D, B}
200	01.01.10	{D, A, C, E, B}
300	01.15.10	{C, A, B, E}
400	01.22.10	{B, A, D}

Find all frequent item sets, using Apriori and FP growth algorithms. Compare the efficiency of the two meaning processes. **(10 Marks)**

b. Explain various alternative methods for generating frequent item sets. **(10 Marks)**

PART – B

5 a. What is Apriori algorithm? Give an example.
A Data base has six transactions of purchase of books from a book shop as given below.
t_1 = {ANN, CC, TC, CG}, t_2 = {CC, D, CG}
t_3 = {ANN, D, CC, TC}, t_4 = {ANN, CC, D, CG}
t_5 = {ANN, CC, D, TC, CG} , t_6 = {C, D, TC}
Let X = {CC, TC} and Y = {ANN, TC, CC}. Find the confidence and support of the Association rule X → Y and inverse rule Y → X. **(10 Marks)**

b. Explain the various properties of objective measures. **(10 Marks)**

6 a. What is cluster analysis? Explain different types of clusters. **(10 Marks)**

b. Explain the hierarchical clustering, with example. **(06 Marks)**

c. Explain DBSCAN algorithm. **(04 Marks)**

7 a. Discuss the use of data mining application for telecom industry. **(10 Marks)**

b. What are the trends in data mining? **(10 Marks)**

8 Write short notes on :
a. K – means algorithm.
b. Outlier analysis.
c. Spatial data mining.
d. Social impact of data mining. **(20 Marks)**

*****