Reg No.										

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

VII SEMESTER B.E. (CS&E)
End-Semester Examination November - 2009

SUBJECT: DATA WAREHOUSE AND DATA MINING (CSE 405.3)
(Revised Credit System)

TIME: 3 HOUR MAX.MARKS: 50

Instructions to Candidates
- Answer **ANY FIVE FULL** questions.
- Missing data may suitably assumed.

1. a) What is Data Warehousing? Give a brief description of the different warehouse schemas. – 5 M

   b) Describe the 4 basic OLAP operations with suitable example. – 5 M

2. a) What is Data Mining? Explain the various stages of Knowledge Discovery in Databases. – 4 M

   b) How is association rules mined from large databases? Write Apriori algorithm to discover all frequent itemsets from large databases. – 6 M

3. a) For the table given below (TABLE-1) find all frequent itemsets using the above algorithm. – 2 M
   List all strong association rules with respect to one large frequent itemset (set containing max. number of elements). Let min_sup = 60% and min_conf=80%. – 4 M

TABLE : 1

TID	Items Brought
T100	{K,A,D,B}
T200	{D,A,C,E,B}
T300	{C,A,B,E}
T400	{B,A,D}

   b) What are Bayesian Classifiers? Discuss Bayes Theorem for classification problems. - 4 M

4.a) What is a decision tree? Write an algorithm to generate a decision tree from the given training data. - 4 M

b) Using a multilayer feed-forward neural network, show the calculations for learning by the backpropagation algorithm. Let the learning rate be 0.9. The initial weight and bias values of the network are given in Table – 2. The first training sample, X=(1,0,1), whose class lable is 1. - 6 M

Table - 2

W_{14}	W_{15}	W_{24}	W_{25}	W_{34}	W_{35}	W_{46}	W_{56}	Θ_4	Θ_5	Θ_6
0.2	-0.3	0.4	0.1	-0.5	0.2	-0.3	-0.2	-0.4	0.2	0.1

5.a) Write a typical k-medoids algorithm for partitioning based on medoid. Also, discuss 4 cases of the cost function for k-medoids clustering with help of diagram. - 5 M

b) What are the different categories of Hierarchical Clustering? Explain BIRCH technique of cluster representation in large databases with example. - 5 M

6. Write short notes on. - 3+3+4 M

i) FP(Frequent Pattern) – Tree

ii) Prediction Models

iii) Web Mining

<<<<>>>>