Reg No.										

प्रज्ञानं ब्रह्म
Manipal
INSPIRED BY LIFE

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

KNOWLEDGE IS POWER
MANIPAL INSTITUTE OF TECHNOLOGY

VII SEMESTER B.E. (CSE) DEGRE END SEMESTER MAKE UP EXAMINATION
DATA WAREHOUSE AND DATA MINING (CSE- 405.2)
DATE: 03-01-2013

TIME: 3 HOURS MAX.MARKS: 50

Instructions to Candidates

- Answer **any five** full questions.
- Missing data can be suitably assumed.

1A. How does a data cube model n dimensional data? Explain with suitable example. (3M)

1B. Distinguish between Virtual Data warehouses and Enterprise Data Warehouses. (3M)

1C. Suppose that a DW consists of three dimensions spectator , location, and game and the two measures count and charge, where charge is the fare that a spectator pays for watching a game, count is number of spectators watching a game. Spectator may be a child, middle aged, or senior person with different charges for each category. There are 4 different games played at 5 different locations.

a) Draw a multidimensional data cube
b) Write the sequence of OLAP operations performed to display total charge collected for child and senior spectators watching two different games, g1, g2 at three different locations l1, l2, l3. (4M)

2 A. With the aid of neat diagram, explain different steps of KDD. (3M)

2 B. Why is Lift of an association rule significant? How do you calculate it? (3M)

2 C. Write the following routines used in Apiori algorithm with proper inputs and outputs. (2+2M)

i) Candiate_ generate(…) ii Prune(…)

3 A. When is pincer search algorithm better compared to Apriori algorithm? Explain the concept involved in Pincer search algorithm. (1+3M)

3 B. Write FP_Tree Construction algorithm. How many data base scans does FP growth algorithm need? (3.5+0.5M)

3 C. Justify the name given to Dynamic Itemset counting algorithm for finding frequent item sets. (2M)

4 A. What is supervised learning and how is it applied to classification? (1M)

4 B Classify the tuple (age < 30, income = medium, student = yes, credit_rating = fair) using naïve Bayesian classifier with the training set given below.

rid	age	income	student	credit_rating	Class: buys_computer
1	<30	high	no	fair	no
2	<30	high	no	excellent	no
3	30-40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	30-40	low	yes	excellent	yes
8	<30	medium	no	fair	no
9	<30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<30	medium	yes	excellent	yes
12	30-40	medium	no	excellent	yes
13	30-40	high	yes	fair	yes
14	>40	medium	no	excellent	no

(Training Set) (4M)

4 C. Find the best attribute split (First level) for the training set given in question 4 B using GINI Index method. (5M)

5 A. Give the reasons for iterative development of a data warehouse. (2M)

5 B. List and explain three levels of data modeling. (3M)

5 C. Discuss the following with respect to Data ware house design. (2+3M)

i) Granularity

ii) Language interface

6. Write short note on the following. (3+4+3M)

a) Prediction

b) Data warehouse Schemas

c) Back propagation neural network learning algorithm