Reg. No.										

Manipal Institute of Technology, Manipal

(A Constituent Institute of Manipal University)

VII SEMESTER B.TECH (COMPUTER SCIENCE AND ENGINEERING)

END SEMESTER EXAMINATIONS, NOV/DEC 2015

SUBJECT: DATA WAREHOUSING AND DATA MINING [CSE 433]

REVISED CREDIT SYSTEM

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Answer **ANY FIVE FULL** questions.
- Missing data, if any, may be suitably assumed.

1A.With a neat diagram explain data warehouse architecture. 3M

1B.Discuss data warehouse back-end tools and Utilities. 2M

1C.Define data warehouse. Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit. Starting with the base cuboid{day, doctor, patient}, what specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2014? Also draw the Star Schema. 5M

2A. Find frequent item sets and association rules with min. confidence = 60% and min. support = 40% for the sales data given below using Apriori algorithm:

T ID	Item Set
100	Milk, Bread, Jam
101	Bread, Butter, Juice
102	Soda, Bread, Butter
103	Bread, Juice, Soda
104	Milk, Juice

4M

2B. How correlation analysis is done using lift and x^2? 2M

2C. Explain the pincer-search method to find frequent itemsets. 4M

3A.Normalize the following group of data using

i. min-max normalization by setting min=0 and max=1.

ii. Z-score normalization.

200,300,400,600,1000 5M

3B. How do you handle missing values? 2M

3C. Draw the FP tree of the below transactional database with minimum support of 40% . The presence of an item is indicated by '1' and absence of an item is indicated by 0.

TID	A	B	C	D	E
T1	1	1	1	0	0
T2	1	1	1	1	1
T3	1	0	1	1	0
T4	1	0	1	1	1
T5	1	1	1	1	0

3M

4A.Write an algorithm to generate decision tree and explain the partitioning scenarios. 5M

4B.Explain the working of Naïve Bayesian classifier. 5M

5A.With the help of an example explain rule accuracy and coverage also give the formulae. 5M

5B. Give an overview of basic clustering methods. 5M

6A. Write the following algorithms.

i. K-means

ii. K-medoids 5M

6B.Explain eager and lazy learners. Give an example algorithm for each. 5M