DHARMSINH DESAI UNIVERSITY, NADIAD
FACULTY OF TECHNOLOGY
B.TECH. SEMESTER VII[I.T.]

SUBJECT: (IT 704) DATA ANALYSIS AND INFORMATION EXTRACTION

Examination	**: First Sessional**	**Seat No.**	**: _____**
Date	**: 01/08/2013**	**Day**	**: Thursday**
Time	**: 1:00 to 2:15 PM**	**Max. Marks**	**: 36**

INSTRUCTIONS:
1. Figures to the right indicate maximum marks for that question.
2. The symbols used carry their usual meanings.
3. Assume suitable data, if required & mention them clearly.
4. Draw neat sketches wherever necessary.

Q.1 Do as directed.

(a) What is data mining? Why the term data mining is a misnomer? **[2]**
(b) Describe the terms spatial databases and temporal databases in brief. **[2]**
(c) Explain various types of the data mining tasks. **[2]**
(d) Explain the following: **[2]**
Age(X, "20…29") and income(X,"20K…29K") => buys(X, "CD player")
[support = 2%, confidence = 60%]
(e) What is an interestingness measure? Explain its types. **[2]**
(f) Describe various types of the exception. **[2]**

Q.2 Attempt ***Any Two*** from the following questions. **[12]**

(a) Write down and explain various steps of the KDD process. **[6]**
(b) Perform the comparison between OLTP and OLAP systems. **[6]**
(c) Define the fact and dimension tables for the sales database in DMQL. **[6]**

Q.3 (a) Explain various OLAP operations in brief. **[6]**
(b) Suppose that a data warehouse for Big-University consists of the following four dimensions: student, course, semester and instructor, and two measures count and avg_grade. Draw a snowflake schema diagram for the data warehouse. **[6]**

OR

Q.3 (a) Suppose that a data warehouse consists of the four dimensions date, spectator, location and game, and the two measures count and charge. **[6]**

a) Draw a star schema diagram for the data warehouse.
b) Starting with the base cuboid [date, spectator, location, game], what specific OLAP operations should one perform in order to list the total charge paid by student spectators at GM_Place in 2000?

(b) Suppose that the data for analysis include the attribute age. The age values for the data tuples are(in increasing order): 13,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,36,40,45,46,52,70. **[6]**

a) Use min-max normalization to transform the value 35 for age onto the range [0.0, 1.0].
b) Use z-score normalization to transform the value 35 for age, where the standard deviation of age is 12.94 years.

**********All the best**********