DHARMSINH DESAI UNIVERSITY, NADIAD
FACULTY OF TECHNOLOGY
B.TECH. SEMESTER V [I.T]
SUBJECT: (IT-704) Data Analysis & Information Extraction

Examination	**: Second Sessional**	**Seat No.**	**: _____**
Date	**: 04/09/2014**	**Day**	**: Thursday**
Time	**: 1:00 to 2:15**	**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) Explain the five parameters of a box plot. **[2]**
(b) What is importance of t-weight and d-weight? Explain with an example. **[2]**
(c) Define the primitives for specifying a data mining task. **[2]**
(d) Mention hash base technique for improving apriori algorithm. **[2]**
(e) Explain certainty and utility interestingness measures with example. **[2]**
(f) Explain market basket analysis with an example. **[2]**

Q.2 Attempt ***Any Two*** from the following questions. **[12]**
(a) Describe various kind of concept hierarchy with a proper example. **[6]**
(b) Describe the data mining system on the basis of various data mining architectures. **[6]**
(c) Explain five types of approaches for multilevel association rule mining. **[6]**

Q.3 (a) A database has four transactions.Let *min_sup*=2 & *min_conf* = 80%. **[6]**

TID	***Date***	***Items_bought***
T100	01/09/2003	{K, A, D, B}
T200	01/09/2003	{D, A, C, E, B}
T300	03/09/2003	{C, A, B, E}
T400	08/09/2003	{B, A, D}

List all frequent itemsets using Apriori and list all strong association rules for given type of patterns. E.g X ^ Y => Z.

(b) The following contingency table summarizes supermarket transaction data. **[6]**

	Hotdogs	_____ hotdogs
hamburgers	2000	500
_____ hamburgers	1000	1500

a) Suppose that the association rule "hot dogs=>hamburger" is mined. Given a minimum support threshold of 25% and a minimum confidence threshold of 50%, is this association rule strong?
b) Based on the given data, is the purchase of hot dogs independent of the purchase of hamburger? If not, what kind of correlation relationship exists between the two?

OR

Q.3 (a) For the database given in Q.3 (a) above let *min_sup*=2 & *min_conf* = 80%. **[6]**
List all frequent itemsets using FP Growth and list all strong association rules for given type of patterns. E.g X ^ Y => Z.

(b) Suppose that we would like to mine the general characteristics describing graduate students at Big-university using analytical characterization. Given are the attributes: **[6]**

Target class		Contrasting class	
Major	Count	Major	Count
Science	16	Science	18
Science	22	Business	20
Engineering	18	Business	22
Science	25	Science	24
Science	21	Engineering	22
Engineering	18	Engineering	24

Perform the analytical characterization (Find the information gain for the attribute major.).