DHARMSINH DESAI UNIVERSITY, NADIAD
FACULTY OF TECHNOLOGY
B.TECH. SEMESTER VII IT

SUBJECT: (IT-704) Data Analysis & Information Extraction

Examination	**: Second sessional**	**Seat No.**	**: _____**
Date	**: 09/09/2016**	**Day**	**: Thursday**
Time	**: 2:15 to 3:30**	**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 association analysis? Explain with an appropriate example. **[2]**
(b) Why should we perform the attribute relevance analysis? **[2]**
(c) What do you mean by a market basket analysis? Explain in brief. **[2]**
(d) Explain anti-monotone property of the apriori algorithm. Give a suitable example. **[2]**
(e) What is a "lift" in the terms of correlation analysis? Give the equation of the correlation coefficient. **[2]**
(f) 'Strong rules are not necessarily interesting." – State true or false. Justify your answer. **[2]**

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

(a) Explain the different approaches to mining multilevel association rules. **[6]**
(b) How can we improve the efficiency of the Apriori algorithm? **[6]**
(c) Explain attribute oriented induction in brief. Give an appropriate example. **[6]**

Q.3 (a) The following contingency table summarizes supermarket transaction data: **[6]**

	Hotdogs	(Hotdogs)'	∑row
Hamburgers	2000	500	2500
(Hamburgers)'	1000	1500	2500
∑col	3000	2000	5000

Based on the given data, is the purchase of hotdogs independent of the purchase of hamburgers? If not, what kind of correlation relationship exists between the two? Give importance of Co-relation in Association Rule.

(b) A database has four transactions. Let min_sup = 60% and min_conf = 80%. **[6]**

TID	Items_bought
T100	{K,A,D,B}
T200	{D,AC,E,B}
T300	{C,A,B,E}
T400	{B,A,D}

a) Find all frequent itemsets using Apriori.
b) List all strong association rules(with support s and confidence c) matching the following metarule, where X is a variable representing customers, and item i denotes variables representing items(e.g. "A", "B", etc.):

$\forall x \in$ transaction, buys(X, item1) $\wedge$ buys(X, item2) => buys(X, item3). [s,c]

OR

Q.3 (a) Suppose that the following table is derived by attribute-oriented induction: **[6]**

Class	Birth_place	Count
Programmer	Canada	180
	Others	120
DBA	Canada	20
	Others	80

Transform the table into a cross tab showing the associated t-weights and d-weights. What is importance of T-weight & d-weight?

(b) A database has 4 transactions. Let min_sup = 60% and min_conf = 80%. **[6]**
Find all frequent item sets using FP- growth algorithm.

TID	Date	Items brought
T100	10/15/99	{1,3,4}
T200	10/15/99	{2,3,5}
T300	10/19/99	{1,2,3,5}
T400	10/22/99	{2,5}