DHARMSINH DESAI UNIVERSITY, NADIAD
FACULTY OF TECHNOLOGY
B.TECH. SEMESTER VII[I.T.]
SUBJECT: (IT 704) DATA ANALYSIS AND INFORMATION EXTRACTION

Examination	**: Block exam**	**Seat No.**	**: _____**
Date	**: 18/10/2013**	**Day**	**: Friday**
Time	**: 3:00 to 4:15 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 association analysis? Explain with an appropriate example. **[2]**
(b) Define what a grain is in the business process is. **[2]**
(c) Every key structure in the data warehouse contains, either implicitly or explicitly, an element of time. : Justify. **[2]**
(d) State the difference between simplicity and a certainty. **[2]**
(e) What is a posterior probability in the bayes theorem? Give the formula for the Bayesian classifier. **[2]**
(f) Explain what are a lazy learners or instance-based classifiers. **[2]**

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

(a) What is an interestingness measure? Explain its types with example. **[6]**
(b) Write a short note on the market basket analysis. **[6]**
(c) Explain the partitioning method k-medoid and CLARANS with example. **[6]**

Q.3 (a) Explain various techniques of data transformation during data preprocessing. **[6]**

(b) The following contingency table summarizes supermarket transaction data, where the transactions are given as below: **[6]**

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

Suppose that the association rule "hotdogs => hamburgers" is mined. Given a minimum support threshold of 25% and a minimum confidence threshold of 50%, is this association rule strong?

OR

Q.3 (a) Give a short example to show that items in a strong association rule may actually be negatively correlated. **[6]**

(b) The following data represent the sales(in hundreds of thousands of dollars) for two outdoor furniture outlets for the last ten years: **[6]**

Year	Outlet(A)	Outlet(B)	Year	Outlet(A)	Outlet(B)
1	118	95	6	143	145
2	114	100	7	147	160
3	130	118	8	158	181
4	125	124	9	149	190
5	140	130	10	161	205

(a) Calculate the regression coefficients for data for both outlets.
(b) How does the average yearly change in sales differ from one outlet to another?

(c) Plot the sales against time for both outlets.
(d) Which one of the two regression lines seems to be better fit with the given data?

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