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

Examination	**: Second Sessional**	**Seat No.**	**: _____**
Date	**: 05/09/2013**	**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 a supervised learning? Give an appropriate example. **[2]**
(b) Explain what a tree-pruning is. Also explain why it should be used. **[2]**
(c) What do you mean by a predictive accuracy of a method? Explain with example. **[2]**
(d) Explain the various operators used in a genetic algorithm.
(e) Describe what is clustering? State some of the applications. **[2]**
(f) State the meaning of the high-dimensionality of a clustering algorithm. Give an appropriate example. **[2]**

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

(a) State why attribute relevance analysis should be performed. Describe the method for the same.
(b) Propose a concept hierarchy for the attributes address, status, major and GPA. Also state what kind of concept hierarchy it is. **[6]**
(c) Consider the following data: **[6]**

Target class(graduate students)		Contrasting class(under-graduate students)	
Gpa	Count	Gpa	Count
Very_good	16	Very_good	18
Excellent	12	Fair	20
Excellent	18	Fair	22
Excellent	25	Fair	24
Excellent	21	Very_good	22
Excellent	18	excellent	24

Find out whether the attribute gpa is task relevant or not if the minimum information gain threshold is 0.4.

Q.3 (a) Explain the algorithm of attribute-oriented induction. Give an appropriate example. **[6]**
(b) Consider the following data: **[6]**

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

Transform the table into a cross-tab showing the associated t-weights and d-weights.

OR

Q.3 (a) Describe the data mining system on the basis of various data mining architectures. **[6]**

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

TID	Date	Items_bought
T100	10/15/99	{K,A,D,B}
T200	10/15/99	{D,A,C,E,B}
T300	10/19/99	{C,A,B,E}
T400	10/22/99	{B,A,D}

Find all frequent itemsets using Apriori algorithm. And list all the strong association rules.

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