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

Examination	**: Third Sessional**	**Seat No.**	**: _____**
Date	**: 10/10/2013**	**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. **[2]**
(e) Describe what is clustering? State some of the applications. **[2]**
(f) State the various types of the variations of separate components of the time series. **[2]**

Q.2 Attempt ***Any Two*** from the following questions. **[12]**
(a) Explain the algorithm of decision tree induction with example. **[6]**
(b) Write a short note on the fuzzy set approach. **[6]**
(c) Explain BIRCH method of clustering approach. **[6]**

Q.3 (a) What is a naïve bayes classifier? Explain with example. **[6]**
(b) The following data shows the number of Lincoln Continental cars sold by a dealer in Queens during the 12 months of 1994. **[6]**

Month	Number sold	Month	Number sold
Jan	52	July	54
Feb	48	Aug	65
Mar	57	Sept	70
Apr	60	Oct	80
May	55	Nov	90
June	62	Dec	75

a) Calculate the three month moving average for this data.
b) Calculate the five month moving average for this data.
c) Which one of these two moving averages is a better smoothing technique and why?

OR

Q.3 (a) Explain the DBSCAN method of clustering. **[6]**
(b) The following data shows the sales revenues for sales of used cars sold by Atlantic Company for the months of January to April in 1995. **[6]**

Month	Sales($100,000)
Jan	95
Feb	105
Mar	100
April	110

Find the error between the actual value and the forecast value for the months of February, March and April of 1995, using exponential smoothing method with $\alpha = 0.6$.

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