



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

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

Examination	: Block examination	Seat No.	: _____
Date	: 21/10/2016	Day	: Friday
Time	: 3:00 to 4:15	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) What is data mining? Explain with example. [2]
- (b) What are the different types of the measures in data mining? [2]
- (c) Describe what is operation-derived concept hierarchy with an appropriate example. [2]
- (d) What do you mean by indexing in data mining? Explain its types. [2]
- (e) What are the draw-backs of apriori algorithm? [2]
- (f) What is the difference between PAM and CLARA? Which one is better? [2]

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

- (a) What are the various ways of coupling or integrating DM system and a DB/DW system? [6]
- (b) Define the requirements of the a clustering algorithm in the data mining. [6]
- (c) Enumerate three classes of schema that are popularly used for modeling the Data Warehouse. [6]

Also draw a schema diagram (using any one schema) for a D/W consists of:
3 dimensions: time, doctor, patient,
2 measures: count and charge.

- Q.3**
- (a) Give a short example to show that items in a strong association rule may actually be negatively correlated. [6]
 - (b) Discuss the classification by decision tree induction with example. [6]

OR

- Q.3**
- (a) Why naive Bayesian classification is called “naive”? Briefly outline the major ideas of naive Bayesian classification. [6]
 - (b) Identify dimensions and measures. Draw star schema diagram and 3-D CUBE for Hospital management system. [6]