

GUJARAT TECHNOLOGICAL UNIVERSITY

MCA – SEMESTER – IV • EXAMINATION – SUMMER 2018

Subject Code: 3640006

Date: 28-May-2018

Subject Name: Data Mining

Time: 10.30 am to 1.00 pm

Total Marks: 70

Instructions:

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

- Q.1 (a)** Explain following terms. **07**
- 1) Data discrimination
 - 2) Data cube
 - 3) Roll-up operation
 - 4) Neural Network
 - 5) Outliers
 - 6) Web mining
 - 7) Define full forms : DIANA, AGNES
- (b)** Differentiate between **07**
- 1) Supervised learning and unsupervised learning
 - 2) Eager learner classification and lazy learner classification
- Q.2 (a)** What is data mining? Describe the steps involved in data mining when viewed as a process of knowledge discovery. **07**
- (b)** Explain the various data pre-processing methods in brief. **07**
- OR**
- (b)** Discuss Frequent Pattern Mining: A Road Map. **07**
- Q.3 (a)** What is Apriori property? How the Apriori property is used in finding frequent itemset. Explain Join and Prune step. **07**
- (b)** Describe Decision Tree Induction algorithm. You can describe it with the help of an example. **07**
- OR**
- Q.3 (a)** Explain the terms “Tree Induction” and “Tree Pruning” in detail. **07**
- (b)** Discuss following in brief: **07**
- 1) Information Gain
 - 2) Gain Ratio
 - 3) Gini Index
- Q.4 (a)** Explain the following classification methods in brief. **07**
- 1) k-Nearest-Neighbor classifier
 - 2) Case Based Reasoning classifier.
- (b)** Write down the k-mediod algorithm for clustering and explain its working. How it is better than k-mean clustering algorithm? **07**
- OR**
- Q.4 (a)** What do you mean by cluster analysis? What are the requirements of cluster analysis? **07**
- (b)** List all categorization of major clustering methods and explain any three clustering methods. **07**
- Q.5 (a)** Explain about mining complex data types. **07**

- (b) Discuss Data Mining for Biological Data Analysis with DNA sequence and corresponding amino acid sequence. **07**

OR

- Q.5** (a) Explain how data mining application is useful for Telecommunication Industry. **07**
(b) Discussion the application of data mining in Retail and Marketing. **07**
