

[Nov-17]

**[ECS-705A]**  
**M.Tech. Degree Examination**  
**Computer Science & Technology**  
**I SEMESTER**

**DATA MINING AND DATA WAREHOUSING**

(Effective from the admitted batch 2017-18)

**Time: 3 Hours**

**Max.Marks: 60**

-----  
**Instructions:** Each Module carries 12 marks.

Answer all modules choosing one question from each module

All parts of the module must be answered in one place only.

Figures in the right hand margin indicate marks allotted.

-----  
**MODULE-I**

- |    |                                                             |   |
|----|-------------------------------------------------------------|---|
| 1. | a) Explain major requirements and challenges in data mining | 6 |
|    | b) Differentiate Data Matrix and Dissimilarity Matrix       | 6 |

**OR**

- |    |                                                                        |   |
|----|------------------------------------------------------------------------|---|
| 2. | a) Explain data mining as a step in the process of knowledge discovery | 6 |
|    | b) What are the different application areas of data mining             | 6 |

**MODULE-II**

- |    |                                                                                                                                                                                                                                                                                      |   |
|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|
| 3. | a) Suppose that a data warehouse consists of three dimensions time, doctor and patient, and two measures count (the number of patients examined) and charge (fee that a doctor charges a patient for a visit). Draw either a star or a snowflake schema for the above data warehouse | 6 |
|    | b) Explain in detail about the implementation of a data warehousing                                                                                                                                                                                                                  | 6 |

**OR**

- |    |                                                         |    |
|----|---------------------------------------------------------|----|
| 4. | Explain multidimensional data model with a neat diagram | 12 |
|----|---------------------------------------------------------|----|

### MODULE-III

5. a) How does tree pruning work? Explain the enhancements to basic decision tree induction 8  
b) Discuss about Back propagation classification algorithm 4

#### OR

6. a) Explain a classifier which assumes class conditional independence 6  
b) Explain about nearest neighbors classifier 6

### MODULE-IV

7. a) Define frequent itemset, closed itemset and association rule 4  
b) The Apriori algorithm makes use of prior knowledge of subset support properties. Prove that all nonempty subsets of a frequent item set must also be frequent and Prove that the support of any nonempty subset  $s_0$  of item set  $s$  must be at least as great as the support of  $s$  8

#### OR

8. Explain Apriori algorithm with example and how the efficiency of Apriori algorithm can be improved 12

### MODULE-V

9. Write algorithms for k-means and k-medoids and explain how they work with example 12

#### OR

10. a) Given an example of how specific clustering methods may be integrated, for example where one clustering algorithm is used as a preprocessing step for another 6  
b) What is Cluster Analysis? What are some typical applications of clustering 6