

III B. Tech II Semester Supplementary Examinations, April - 2021
DATA WAREHOUSING AND DATA MINING

(Computer Science and Engineering)

Time: 3 hours

Max. Marks: 70

- Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B**)
2. Answer **ALL** the question in **Part-A**
3. Answer any **FOUR** Questions from **Part-B**
- ~~~~~

PART -A

(14 Marks)

1. a) What is the difference between the Text databases and multimedia databases? [2M]
- b) What is the importance of data preprocessing? [2M]
- c) Discuss about Entropy. [2M]
- d) Write about conditional independence. [3M]
- e) Differentiate between support and confidence. [3M]
- f) Mention the important issues in hierarchical clustering. [2M]

PART -B

(56 Marks)

2. a) Discuss about data similarity and dissimilarity. [7M]
- b) Write short notes on major issues in data mining. [7M]
3. a) Explain about entity identification problem. [7M]
- b) Describe the various approaches to remove the noisy data from the original data. [7M]
4. a) Explain the methods for expressing attribute test conditions. [7M]
- b) Draw and explain a decision tree for mammal classification problem. [7M]
5. a) Discuss the characteristics of Bayesian belief networks. [7M]
- b) Write an algorithm for generating the topology of a Bayesian network. [7M]
6. a) Explain about association rule mining. [7M]
- b) Write an apriori algorithm. [7M]
7. a) Explain about DBSCAN algorithm. [7M]
- b) What are the strengths and weaknesses of a DBSCAN algorithm? [7M]
