

III B. Tech II Semester Supplementary Examinations, June-2022
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) Define the concepts of Median and Mode. [2M]
- b) What is the advantage of Pearson's product moment coefficient? [2M]
- c) How do you group ordinal attributes? [2M]
- d) List the characteristics of Nearest Neighbor Classifiers. [3M]
- e) Define the Association rule mining problem. [3M]
- f) List the strengths and weaknesses of the DBSCAN algorithm. [2M]

PART -B

(56 Marks)

2. a) List and explain different types of data with examples. [7M]
- b) Discuss various methods to visualize higher dimension datasets. [7M]
3. a) Illustrate the forms of data preprocessing with a neat sketch. [7M]
- b) Explain the steps to deal with Noisy data using Binning. [7M]
4. a) Discuss the measures for selecting the Best Split. [7M]
- b) Explain the step-by-step process of implementing the Decision tree induction algorithm. [7M]
5. a) Discuss various ways to estimate the conditional probabilities for continuous attributes. [7M]
- b) Predict the class label of a test record X = (Home owner = Yes, Marital Status = Married, Income = \$120K). (See the following Table-1) [7M]

Table-1

Tid	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

6. a) Explain the methods of Frequent Itemset Generation and Rule Generation. [7M]
b) Trace the candidate elimination and pruning methods with an example. [7M]
7. a) Trace the K-means algorithm with a case study. [7M]
b) Compare K-means with DBSCAN clustering algorithm. [7M]

2 of 2