

19/4

Code No.: CS701PC

R20

H.T.No.

8

R

**CMR ENGINEERING COLLEGE: : HYDERABAD**  
**UGC AUTONOMOUS**  
**IV-B.TECH-I-Semester End Examinations (Supply) – April - 2024**  
**DATA MINING**  
**(CSE)**

**[Time: 3 Hours]****[Max. Marks: 70]****Note:** This question paper contains two parts A and B.

Part A is compulsory which carries 20 marks. Answer all questions in Part A.

Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

**PART-A****(20 Marks)**

1. a) What is data cleaning? [2M]
- b) What is Data Discretization? [2M]
- c) Define frequent itemset. [2M]
- d) What closed itemset? [2M]
- e) Define the decision tree. [2M]
- f) Describe information gain. [2M]
- g) Define Clustering. [2M]
- h) State the K-medoids algorithm. [2M]
- i) What is text mining? [2M]
- j) Describe web structure mining. [2M]

**PART-B****(50 Marks)**

2. What is Data Mining? Explain the process of knowledge discovery in database. [10M]
- OR**
3. Explain the techniques in dimensionality reduction. [10M]
  4. Describe various types of association rules. [10M]
- OR**
5. Describe the FP growth algorithm with an example. [10M]
  6. Explain K-Nearest-Neighbor Classification algorithm. [10M]
- OR**
7. What is tree pruning? Explain the Naïve Bayes Classification. [10M]
  8. Explain in briefly about K-means algorithm with an example. [10M]
- OR**
9. Describe the different types of hierarchical clustering methods. [10M]
  10. Explain Web Content Mining in detail. [10M]
- OR**
11. Discuss the techniques and algorithms used to analyze and extract useful information from unstructured text data. [10M]

\*\*\*\*\*