

Code No.: DS504PC

R20

H.T.No.

8 R

**CMR ENGINEERING COLLEGE: : HYDERABAD  
UGC AUTONOMOUS**

**III-B.TECH-I-Semester End Examinations (Regular) - January- 2024  
INTRODUCTION TO DATA MINING  
(CSD)**

[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.

**PART-A**

**(20 Marks)**

1. a) List the methods of filling missing values. [2M]
- b) Define data mining. On what kind of data, data mining can be performed. [2M]
- c) Define the two step process of association rule mining. [2M]
- d) What is the need of confidence measure in association rule mining? [2M]
- e) Quote an example for Bayesian belief network. [2M]
- f) List the characteristics of k-nearest neighbor algorithm. [2M]
- g) What are the weaknesses of hierarchical clustering? [2M]
- h) Which method is more robust among K -means and K -medoids clustering? Justify your answer. [2M]
- i) Define Text Clustering. [2M]
- j) What is web Structure Mining? [2M]

**PART-B**

**(50 Marks)**

2. Why do need to transform data? Explain the different approaches for data transformation. [10M]  
For the given data normalize the data 400 using min max normalization and decimal scaling method to the range 0 to 1.  
Data: 200, 300, 400, 600, 800, 1000
- OR**
3. List and explain the processing steps that may be applied to the data for data mining (KDD process). [10M]
4. Can we design a method that mines the complete set of frequent item sets without candidate generation? If yes, explain with an example. [10M]
- OR**
5. Discuss the FP-Growth algorithm with an example. [10M]
6. Explain in detail about Attribute Selection methods in Classification. [10M]
- OR**
7. Explain the algorithm for constructing a decision tree from training samples. [10M]
- 8.a) Give a brief note on PAM Algorithm. [4M]
- b) What is the drawback of k-means algorithm? How can we modify the algorithm to diminish that problem? [6M]
- OR**
9. What are outliers? Discuss the methods adopted for outlier detection. [10M]
10. Compare and contrast text mining with web content mining using lucid examples. [10M]
- OR**
11. Discuss various kinds of patterns to be mined from web/server logs in web usage mining. [10M]

\*\*\*\*\*