

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
III–B.TECH–I–Semester End Examinations (Supply) - June- 2025
DATA MINING
(Common for IT, CSD, CSM)

[Time: 3 Hours]

[Max. Marks: 60]

Note: This question paper contains two parts A and B.

Part A is compulsory which carries 10 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

(10 Marks)

1. a) What is an Asymmetric Data attribute? Give an Example? [1M]
- b) List various Data preprocessing steps. [1M]
- c) What is the need of confidence measure in association rule mining? [1M]
- d) What is correlation analysis? [1M]
- e) Write an example for Bayesian belief network. [1M]
- f) List the measures for selecting best split in decision tree construction. [1M]
- g) What are the limitations of single linkage algorithm? [1M]
- h) List the typical requirements of clustering in data mining. [1M]
- i) Give an example for unstructured text. [1M]
- j) List the applications of web usage mining. [1M]

PART-B

(50 Marks)

- 2.a) What are the advantages and disadvantages of Knowledge Discovery process? [5M]
 - b) Describe the data similarity and data dissimilarity metrics used in data mining. [5M]
- OR**
- 3.a) Describe various methods of Data Cleaning. [5M]
 - b) Discuss about the Issues to be considered during Data Integration. [5M]
4. Explain market basket analysis and its relevance to association rule. Explain the Apriori algorithm using the following transactional data assuming that the support count is 22%. Illustrate with an example. [10M]

TID

List of Items

001	Milk, dal, sugar, bread
002	Dal, sugar, wheat, jam
003	Milk, bread, curd, paneer
004	Wheat, paneer, dal, sugar
005	Milk, paneer, bread
006	Wheat, dal, paneer, bread

OR

5. What is the need of multidimensional rule? Describe the approaches to multilevel association rule mining. [10M]
- 6.a) Demonstrate decision tree induction. [5M]
 - b) Illustrate rule based classification. [5M]
- OR**
- 7.a) Describe the accuracy metrics used for classification. [5M]
 - b) Explain about support vector machine (SVM). [5M]

8. Identify two clusters in the given dataset using K-Means clustering algorithm. [10M]

X	167	182	176	173	172	174
Y	51	62	69	64	65	56

OR

- 9.a) Differentiate Agglomerative and Divisive Hierarchical Clustering. [5M]
b) What are the challenges of outlier analysis? [5M]

- 10.a) What is the difference between Web content mining and web structure mining? [4M]
b) Illustrate any one of the special classification algorithms. [6M]

OR

11. Explain the following:
a) Temporal association rule mining. [5M]
b) Mining sequence patterns in transactional databases. [5M]
