

CMR ENGINEERING COLLEGE: : HYDERABAD

UGC AUTONOMOUS

III-B.TECH-II-Semester End Examinations (Supply) - June- 2025

DATA MINING

(IT)

[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 the Curse of Dimensionality? [2M]
- b) Give a brief note on Data Discretization. [2M]
- c) Explain about maximal frequent items set. [2M]
- d) Name the steps in association rule mining. [2M]
- e) Define Conditional Independence. [2M]
- f) List out different Classification Techniques. [2M]
- g) Illustrate the meaning of Cluster Analysis. [2M]
- h) State Bayes theorem. [2M]
- i) What are the features of Unstructured text mining? [2M]
- j) Define Web mining and text mining. [2M]

PART-B**(50 Marks)**

2. Discuss about Feature Subset Selection. [10M]

OR

3. Define data cleaning. Discuss the different techniques for handling missing values. [10M]

4. Consider the Data set D. Given the minimum support=2, apply Apriori algorithm on this dataset. [10M]

Transaction	IDItems
100	A,C,D
200	B,C,E
300	A,B,C,E
400	B,E

OR

5. Generate frequent itemsets for the following data using FP-Growth algorithm with min support=50% [10M]

Transaction	List of items
T1	I1,I2,I3
T2	I2,I3,I4
T3	I4,I5
T4	I1,I2,I4
T5	I1,I2,I3,I5
T6	I1,I2,I3,I4

6. Discuss various measures for selecting best split for different attribute types. [10M]

OR

7a) Explain in detail the Naive-Bayes Classifier. [05M]

b) List the characteristics of K- Nearest neighbor classification. [05M]

8. Explain in detail about K-Means clustering algorithm and discuss its limitations. [10M]

OR

9. Discuss the following methods in finding inter-cluster distance in hierarchical clustering [10M]

i) MIN (or) Single Linkage ii) Complete Linkage iii) Grouping

10. What activities are involved in web usage mining? Explain. [10M]

OR

11. Explain about Episode rule discovery for texts. [10M]
