

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS

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

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) What do you mean by data cleaning? [2M]
- b) What is Data Binaryzation? [2M]
- c) What are frequent patterns? Give an example. [2M]
- d) What are closed frequent item sets? Give an example. [2M]
- e) Define entropy and Information gain. [2M]
- f) What is overfitting in Decision trees? [2M]
- g) Define Clustering. List different types of data in cluster analysis. [2M]
- h) What are the advantages of PAM Method? [2M]
- i) What are the features of Unstructured text mining? [2M]
- j) Define web content mining. [2M]

PART-B

(50 Marks)

2. Describe Data Mining along with its functionalities and issue. In your answer explain the following: [10M]
 - i) Is it another hype?
 - ii) Is it simple transformation of technology developed from databases, statistics and machine learning?
 - iii) Explain how the evolutions of database technology lead to data mining.
 - iv) Describe the steps involved in data mining when viewed as knowledge discovery process.

OR

- 3.a) “Data preprocessing is necessary before data mining process”. Justify your answer. [5M]
- b) Enumerate feature subset selection methods. [5M]
4. Can we design a method that mines the complete set of frequent item sets without candidate generation? Explain with example. [10M]

OR

5. Apply FP-Growth algorithm to the following transactional data to find frequent itemsets. [10M]
 List all frequent itemsets with their support count.

TId	List of Item IDs
1	i1,i3,i5,i7
2	i2,i4,i6,i8
3	i1,i3,i5,i7
4	i9,i7,i5,i1
5	i2,i4,i6,i7
6	i1,i2,i3,i4
7	i3,i4,i5,i6
8	i7,i8,i6,i1
9	i8,i5,i3,i2
10	i1,i3,i4,i6

6. Explain decision tree induction algorithm for classifying data tuples with suitable example.

OR

7. Describe the data classification process with a neat diagram. How does the Naive Bayes classification works? Explain. [10M]
8. Write K-means clustering algorithm. Apply K-means clustering algorithm on the following data. Use $C_1(2,4)$ and $C_2(6,3)$ as initial cluster centers. Data: a(2,4), b(3,3), c(5,5), d(6,3), e(4,3), f(6,6).

OR

9. Given the following distance matrix, construct the dendrogram using single linkage, and complete linkage clustering algorithms. [10M]

	A	B	C	D	E
A	0	2	3	3	4
B	2	0	3	5	4
C	3	3	0	2	6
D	3	5	2	0	4
E	4	4	6	4	0

10. Discuss the following [10M]
i) Web structure mining.
ii) Web usage mining.

OR

11. Discuss in detail about text clustering. [10M]
