Code No.: IT621PE	R20	H.T.No.			8	R						

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
III–B.TECH–II–Semester End Examinations (Supply) - December- 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 Data Mining?	[2M]
b)	Define Data transformation.	[2M]
c)	What is meant by association rule mining?	[2M]
d)	What are frequent Item Sets?	[2M]
e)	Why are decision trees useful?	[2M]
f)	Write an example for Bayesian belief network.	[2M]
g)	Define Clustering.	[2M]
h)	Explain Hierarchical Clustering.	[2M]
i)	Give examples for unstructured text.	[2M]
j)	Give the taxonomy of web mining.	[2M]

PART-B **(50 Marks)**

2. Explain the need of data preprocessing and various forms of preprocessing. [10M]

OR

3. a) Differentiate between data mining and data warehouse. [5M]
b) Discuss the major issues in data mining. [5M]

4. Apply FP-Growth algorithm to the following data for finding frequent item sets, consider support threshold as 30%. [10M]

TID	List of Item IDs
1	I1, i2, i4, i5
2	I2, i4, i7
3	I2,i3,i4,i5
4	I1,i3,i4,i7
5	I1,i2,i3,i4,i5
6	I3,i4,i5,i6

OR

5. Apply apriori algorithm to find frequent item sets from the following transactional database. Let min_sup = 30%. [10M]

TID	Items_bought
1	Pen, notebook, ruler
2	Pencil, eraser, sharpener
3	Pen, ruler, chart, sharpener
4	Pencil, clip, eraser
5	Ruler, pin, story book, pen
6	Marker, chart, sketch pens

6. a) Discuss the significance of information gain in decision tree induction. [5M]
b) Explain K-Nearest neighbor algorithm with an example. [5M]

OR

7. Apply Naive-Bayesian classifier to identify class label(campus_placement) to the new sample/student < 7 to 8, 'Fair', 'Excellent', 'No'>. [10M]

SID	CGPA	Coding Skills	Soft Skills	Hackathon participation	Campus-placement
1	7 to 8	Excellent	Fair	Yes	Yes
2	8 to 9	Fair	Excellent	Yes	Yes
3	9 to 10	Poor	Fair	No	Yes
4	5 to 6	Poor	Excellent	No	No
5	7 to 8	Excellent	Poor	No	No
6	8 to 9	Fair	Fair	Yes	Yes
7	9 to 10	Poor	Poor	No	No

8. a) Explain Partitioning Clustering Method with an example. [5M]
b) Explain the key issues, strengths and weaknesses of hierarchical clustering algorithm. [5M]

OR

9. Appraise the importance of outlier detection and its application. Explain any one approach for outlier detection. [10M]

10. Compare and contrast text mining with web content mining with an examples. [10M]

OR

11. a) Describe web usage mining. [5M]
b) Explain about Text Clustering with an example. [5M]

****