

Code No.: AI512PE

R20

H.T.No.

8

R

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
III-B.TECH-I-Semester End Examinations (Supply) - June- 2024
DATA MINING
(CSM)

[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 is an Asymmetric Data attribute? Give an Example [2M]
- b) Differentiate Classification and Clustering [2M]
- c) What is the need of confidence measure in association rule mining? [2M]
- d) Define closed frequent item set [2M]
- e) Quote an example for Bayesian belief network [2M]
- f) What is Decision Tree Pruning [2M]
- g) What are the limitations of partitioning algorithms? [2M]
- h) List Various types of clustering methods [2M]
- i) Give examples for unstructured text [2M]
- j) What is meant by stop words? [2M]

PART-B

(50 Marks)

2. Discuss about the following
 - a) What motivated Data mining? Explain [5M]
 - b) Data mining as a step in the process of knowledge discovery [5M]
- OR
- 3.a) Explain about the Data Cleaning techniques in detail. [5M]
- b) Discuss in detail about data preprocessing [5M]
- 4.a) List out different kinds of Association Rules with an example fo reach. [5M]
- b) Explain about maximal frequent Item set and closed frequent Itemset [5M]
- 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, storybook, pen
6	Marker, chart, sketchpens

- 6.a) Compare the methods of classification and prediction. [5M]
b) How to evaluate performance of classification model? Explain [5M]
- OR
- 7.a) State Bayes theorem. How this concept is used in classification [7M]
b) What is Lazy Learner [3M]
8. Explain the various clustering evaluation methods [10M]
- OR
9. Appraise the importance of outlier detection and its application. Explain any one approach for outlier detection [10M]
- 10.a) Discuss the basic measures for text retrieval [5M]
b) How to convert unstructured text into features in text mining? [5M]
- OR
11. Discuss various kinds of patterns to be mined from web/server logs in web usage mining [10M]
