Code No.: CS701PC | **R20** | **H.T.No.** | | | **8** | **R** | | | | | | |

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
IV–B.TECH–I–Semester End Examinations (Supply) – April - 2025
DATA MINING
(CSE)

[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) List out major issues in Data Mining. [2M]
   b) Define Data Cleaning. [2M]
   c) Define association rule. [2M]
   d) What is the need of confidence measure in association rule mining? [2M]
   e) Mention the characteristics of K-nearest neighbor classification algorithm. [2M]
   f) Give the construction of naïve Bayesian classification [2M]
   g) List the typical requirements of clustering in data mining. [2M]
   h) List out the requirements of cluster analysis. [2M]
   i) What are the sources of data in web mining? [2M]
   j) Write about Text Clustering? [2M]

PART-B **(50 Marks)**

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

OR

3. What is data mining? Discuss the challenges associated with data mining. [10M]

4. How can we mine closed frequent item sets? Explain. [10M]

OR

5. Consider the following dataset and find frequent item sets and generate association rules for them using Apriori Algorithm. (Minimum support = 2, Minimum confidence = 60%) [10M]

TID	ITEMS
T1	I1,I2,I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

6. Discuss K- Nearest neighbor classification-Algorithm and Characteristics. [10M]

OR

7. Write and explain decision tree classifier with induction algorithm with an example. [10M]

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

OR

9. Explain K-means algorithm with an example. [10M]

10. Explain web structure mining with a suitable algorithm. [10M]

OR

11. Elaborate Text Clustering with an illustrative example. [10M]

****