Code No.: R22CS504PC

R22   H.T.No. | | 8 | R | | | | |

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
III–B.TECH–I–Semester End Examinations (Regular) - December- 2024
DATA MINING
(CSE)

[Time: 3 Hours]   [Max. Marks: 60]

Note: This question paper contains two parts A and B.
  Part A is compulsory which carries 10 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   (10 Marks)

1. a) Define Knowledge Discovery in Databases (KDD).   [1M]
   b) What is dimensionality reduction?   [1M]
   c) Define a frequent item set.   [1M]
   d) What is the APRIORI principle?   [1M]
   e) What are the basic steps in constructing a decision tree?   [1M]
   f) What is the Naive Bayes classifier?   [1M]
   g) Name two applications of clustering in real-world scenarios.   [1M]
   h) How does PAM differ from K-Means in partitioning   [1M]
   i) What is web mining?   [1M]
   j) What is the difference between structured and unstructured text? [1M]

PART-B   (50 Marks)

2. Discuss the key techniques used in data mining and provide examples of their   [10M]
   applications.

OR

3. Explain the different types of data transformation techniques and their applications.   [10M]

4. Describe the association rule problem in data mining and its significance in   [10M]
   discovering patterns.

OR

5. Compare and contrast the FP-Growth algorithm with the APRIORI algorithm,   [10M]
   focusing on performance and memory requirements.

6. Describe the structure of Bayesian Belief Networks and their use in classification.   [10M]

OR

7. Explain the K-Nearest Neighbor algorithm in detail with an example.   [10M]

8. Illustrate an overview of clustering techniques and their applications.   [10M]

OR

9. Explain the methods used for outlier detection in clustering and their significance in   [10M]
   data mining.

10. Explain the process of text clustering and its importance in text mining.   [10M]

OR

11. Describe web structure mining and explain how it is used to analyze hyperlink   [10M]
    structures.

****