Code No.: DS504PC | R20 | H.T.No. | | | 8 | R | | | | | |

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
III–B.TECH–I–Semester End Examinations (Supply) - June- 2024
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 is data mining? List the types of data that can be mined. [2M]
   b) List the distance measures for dissimilarity computation. [2M]
   c) Define closed frequent item set. [2M]
   d) How to compute confidence of an association rule? Give example. [2M]
   e) Why KNN is called as instance-based learning and lazy learning? [2M]
   f) List out various classification techniques. [2M]
   g) What is grid-based clustering? [2M]
   h) What are the key issues in hierarchical clustering? [2M]
   i) Give the taxonomy of web mining. [2M]
   j) What is web structure mining? [2M]

PART-B (50 Marks)

2. Discuss the steps in knowledge discovery process and compare it with data access and information retrieval. [10M]

OR

3. What is Data Cleaning? Describe various methods of Data Cleaning. [10M]

4. A database has five transactions. Let *min sup* D 60% and *min conf* D 80%. [10M]

TID	*items_bought*
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y }
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Find all frequent itemsets using FP-growth algorithm.

OR

5. Explain apriori algorithm with an example. [10M]

6. What is a decision tree? Explain decision tree induction algorithm. [10M]

OR

7.a) Write an algorithm for k-nearest-neighbor classification given k, the nearest number of neighbors, and n, the number of attributes describing each tuple. [10M]
   b) Write the advantage and disadvantages of naïve-bayes classifier.

8. Use the K-means clustering algorithm and Euclidean distance to cluster the following eight examples into three clusters: [10M]
A1= (2, 10), A2= (2, 5), A3= (8, 4), A4= (5, 8), A5= (7, 5), A6= (6, 4), A7= (1, 2), A8= (4, 9). Find the new centroid at every new point entry into the cluster group. Assume initial cluster centers as A1, A4 and A7.

OR

9. Given the following distance matrix, construct the dendogram using single linkage clustering algorithm. [10M]

Item	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. Compare and contrast text mining with web content mining using lucid examples. [10M]

OR

11. How to apply mining techniques to unstructured text database? Explain with example. [10M]

****