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

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
III–B.TECH–II–Semester End Examinations (Regular) - May- 2023
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) List the methods of filling missing values. [2M]
   b) Interpret the dimensionality reduction? [2M]
   c) Interpret the support and confidence rules for item set A and item set B? [2M]
   d) Define Maximal frequent item set. [2M]
   e) Compare information gain and gini index. [2M]
   f) How will you solve a classification problem using decision trees? [2M]
   g) Differentiate agglomerative and divisive hierarchical clustering? [2M]
   h) How can we make k-means algorithm more scalable? [2M]
   i) What is meant by web content mining? [2M]
   j) Give the taxonomy of web mining. [2M]

PART-B **(50 Marks)**

2. What is data mining? Discuss the challenges associated with data mining. [5M]
   Illustrate any three measures for dissimilarity of numeric data. [5M]

OR

3. Describe the five primitives for specifying a data mining task? [10M]
4. Discuss about FP-growth algorithm for the following given example. [10M]
   {M,O,N,K,E,Y} {D,O,N,K,E,Y} {M,A,K,E} {M,U,C,K,Y} {C,O,O,K,I,E},
   Support= 60 %, Confidence = 80 %.

OR

5. A database has four transactions. Let min_sup=60% and min_conf=80%. [10M]

TID	*date*	*items_bought*
100	10/15/2020	{K, A, B, D}
200	10/15/2020	{D, A, C, E, B}
300	10/19/2020	{C, A, B, E}
400	10/22/2020	{B, A, D}

   Find all frequent items using Apriori Algorithm.
6. Explain about Attribute Subset Selection Measures with an example. [10M]

OR

7. Discuss about Naïve-Bayes classification technique with an illustrative example. [10M]
8. How efficient is the K-medoids algorithm on large data sets? Illustrate with example? [10M]

OR

9. How to evaluate clustering algorithms? Provide illustrations. [10M]
10. Explain briefly about Web mining. Discuss the applications of web usage mining. [10M]

OR

11. How to convert unstructured text in to features in text mining? [5M]
    Explain web structure mining with a suitable algorithm. [5M]

****