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

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
III–B.TECH–I–Semester End Examinations (Supply) - May- 2023
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? What are the functionalities of data mining? [2M]
   b) List the major tasks in data pre-processing. [2M]
   c) What are maximal frequent item sets? Give an example. [2M]
   d) How to represent Frequent item set in compact format? [2M]
   e) State Bayes theorem. [2M]
   f) What do you mean by lazy classification? [2M]
   g) Differentiate between AGNES and DIANA algorithms. [2M]
   h) What are the key issues in Hierarchical clustering? [2M]
   i) What is the purpose of web usage mining? [2M]
   j) Define web structure mining. [2M]

PART-B **(50 Marks)**

2.a Define data cleaning? Explain different techniques in handling the missing values? [6M]
   b What is KDD? Explain about data mining as a step in the process of knowledge discovery. [4M]

OR

3.a Write about Dimensionality reduction methods. [6M]
   b Discuss about different measures of similarity and dissimilarity. [4M]

4. Explain about the Apriori algorithm for finding frequent item sets with an example. [10M]

OR

5. Illustrate the FP-growth algorithm with an example. What are the advantages of FP-Growth algorithm? [10M]

6. Explain Naïve Bayes Classification. Given the data below, predict the output (Flu?) for the following new instance using Naïve Bayes algorithm. [10M]
   X: (Chills = N; Runny Nose=N, Headache=Strong, Fever=Y)

Chills	Runny Nose	Headache	Fever	Flu
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

OR

7. a What are Bayesian Belief Networks? Explain the concept of inference using BBN with an example. [5M]
   b Describe K-nearest neighbor algorithm. Why is it called instance based learning? [5M]

8. Write agglomerative clustering algorithm. Given the following distance matrix, construct the dendogram using complete linkage clustering algorithm. [10M]

	1	2	3	4	5
1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	11	10	2	8	0

OR

9.a What is outlier detection? Explain distance based outlier detection. [6M]

b Give a brief note on PAM Algorithm. [4M]

10. Explain the process of mining the World Wide Web. [10M]

OR

11. Write a note on text mining. Discuss about episode rule discovery for texts. [10M]

****