

Roll No																			
---------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



**PRESIDENCY UNIVERSITY  
BENGALURU**

**SET A**

**SCHOOL OF ENGINEERING  
END TERM EXAMINATION - JAN 2024**

**Semester :** Semester III - 2022

**Course Code :** CSE2021

**Course Name :** Data Mining

**Program :** B.Tech.

**Date :** 08-JAN-2024

**Time :** 9:30AM - 12:30 PM

**Max Marks :** 100

**Weightage :** 50%

---

**Instructions:**

- (i) Read all questions carefully and answer accordingly.
  - (ii) Question paper consists of 3 parts.
  - (iii) Scientific and non-programmable calculator are permitted.
  - (iv) Do not write any information on the question paper other than Roll Number.
- 

**PART A**

**ANSWER ALL THE QUESTIONS**

**5 X 2M = 10M**

1. List any four applications of data mining. (CO1) [Knowledge]
2. Describe attribute and the types of attributes. (CO2) [Knowledge]
3. Outline the advantages of FP growth over the Apriori algorithm. (CO3) [Knowledge]
4. State the Back propagation with suitable diagram? (CO4) [Knowledge]
5. Define Cluster Analysis and distinguish any 2 differences on intra-cluster and inter-cluster. (CO5) [Knowledge]

**PART B**

**ANSWER ALL THE QUESTIONS**

**5 X 10M = 50M**

6. a. Discuss the Patterns can be mined in data mining.  
b. Explain the major issues in data mining on data mining and society. (CO1) [Comprehension]

7. a. Consider a movie recommendation system where users rate movies on a scale. Two users, User A and User B, have provided ratings for a set of movies. Apply the Cosine similarity measure and Jaccard Coefficient to find the similarity between these users.  
 A: 1 5 0 0 2 1 0 3, B: 3 6 0 1 4 2 1 2  
 b. What is Minkowski Distance? Find the similarity metrics using minkowski distance for n=1, 2 using the below data:  
 X: 10, 25, 70, 80, 45, 60 Y: 100, 115, 145, 200, 165, 180.

(CO2) [Comprehension]

8. A database has 5 transactions. Apply FP Growth algorithm to:
1. Generate the frequent itemset
  2. Construct FP tree
  3. Find frequent pattern and Conditional frequent pattern generation.

TID	ITEMS
T1	{M,O,N,K,E,Y}
T2	{D,O,N,K,E,Y}
T3	{M,A,K,E}
T4	{M,U,C,K,Y}
T5	{C,O,O,K,I,E}

Let the minimum support be 3.

(CO3) [Comprehension]

9. Consider Temperature data set, Use Naïve Bayesian Classification for predicting the class for a new sample:  
 Temperature = 26, Humidity = 72, Wind Speed = 8.

Temperature	Humidity	Wind Speed	Class
30	70	5	Hot
25	80	10	Mild
18	60	8	Cool
28	75	12	Hot
22	90	7	Mild
15	50	5	Cool
20	65	9	Mild
32	85	15	Hot
17	55	6	Cool
27	78	11	Hot

(CO4) [Comprehension]

10. a) Apply K-Means Clustering to form 2 Clusters for the given data set : {2,4,6,9,12,16,20,24,26}  
 b) Illustrate about Minpoint, Core Point, Border Point and Noise Point in DBSCAN with an example

(CO5) [Comprehension]

**PART C**

**ANSWER ALL THE QUESTIONS**

**2 X 20M = 40M**

11. The given data set consists 14 customers historical data of buys\_computer depending on age, income, student, credit rating. Apply CART/Gini Index algorithm to find the best splitting attribute to construct decision tree.

Age	Income	Student	Credit_Rating	Buys_Computer
<=30	High	No	Fair	No
<=30	High	No	Excellent	No
31 to 40	High	No	Fair	Yes
>40	Medium	No	Fair	Yes
>40	Low	Yes	Fair	Yes
>40	Low	Yes	Excellent	No
31 to 40	Low	Yes	Excellent	Yes
<=30	Medium	No	Fair	No
<=30	Low	Yes	Fair	Yes
>40	Medium	Yes	Fair	Yes
<=30	Medium	Yes	Excellent	Yes
31 to 40	Medium	No	Excellent	Yes
31 to 40	High	Yes	Fair	Yes
>40	Medium	No	Excellent	No

(CO4) [Application]

12. Apply Single linkage and Complete linkage Agglomerative Clustering for the given distance matrix and visualize the clusters using dendrogram

	A	B	C	D	E	F
A	0					
B	5	0				
C	14	9	0			
D	11	20	13	0		
E	18	15	6	3	0	
F	10	16	8	10	11	0

(CO5) [Application]