Roll No											

PRESIDENCY UNIVERSITY
BENGALURU

SET A

SCHOOL OF ENGINEERING
END TERM EXAMINATION - JAN 2024

Semester : Semester VII - 2020
Course Code : CSE2021
Course Name : Data Mining
Program : B.Tech. Computer Science and Engineering

Date : 03-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 the Issues of KDD.

(CO1) [Knowledge]

2. Name the measures of data quality.

(CO2) [Knowledge]

3. Define association rule

(CO3) [Knowledge]

4. Compare supervised and unsupervised learning.

(CO4) [Knowledge]

5. List the types of hierarchical clustering.

(CO5) [Knowledge]

PART B

ANSWER ALL THE QUESTIONS **5 X 10M = 50M**

6. i) Mention the applications of data mining.
ii) Demonstrate the issues of data mining.

(CO1) [Comprehension]

7. Calculate SMC and JC from the following:
P: 1 0 0 1 0 0 1 0 0 1, Q: 1 1 0 1 1 1 0 1 0 1 1

(CO2) [Comprehension]

8. Determine the association rule using apriori algorithm from the following:

Transaction ID	Purchase Products
T1	Book, Pen, Pencil, Eraser, Sharpner, Campus Box, Color pencil
T2	Pen, Color Pencil, Sharpner, Book
T3	Chocolate, Book, Pen, Candy, Eraser
T4	Sharpner, Campus Box, Color pencil, Pencil
T5	Pen, Pencil, Eraser, Sharpner

(CO3) [Comprehension]

9. Determine the value of D when X1 = 70 and X2 = 80 and K = 3 using KNN algorithm from the given data:

X1	X2	D
40	50	Bad
95	75	Good
60	70	Good
45	35	Bad
80	85	Good

(CO4) [Comprehension]

10. Create 3 clusters using K-means algorithm from the given data:
X: 100, 115, 200, 175, 190, 80, 40 Y: 50, 75, 100, 95, 80, 120, 140

(CO5) [Comprehension]

PART C

ANSWER ALL THE QUESTIONS **2 X 20M = 40M**

11. Create a decision tree by ID3 Algorithm using below table:

Age	Income	Student	Credit_Rating	Buys_Computer
Less than 30	High	No	Fair	No
Less than 30	High	No	Excellent	No
Between 31 to 40	High	No	Fair	Yes
Greater than 40	Medium	No	Fair	Yes
Greater than 40	Low	Yes	Fair	Yes
Greater than 40	Low	Yes	Excellent	No
Between 31 to 40	Low	Yes	Excellent	Yes
Less than 30	Medium	No	Fair	No
Less than 30	Low	Yes	Fair	Yes
Greater than 40	Medium	Yes	Fair	Yes
Less than 30	Medium	Yes	Excellent	Yes
Between 31 to 40	Medium	No	Excellent	Yes
Between 31 to 40	High	Yes	Fair	Yes
Greater than 40	Medium	No	Excellent	No

(CO4,CO3) [Application]

12. Compute Agglomerative clustering using single linkage, complete linkage and average linkage from the given data:
70, 90, 45, 67, 89, 34, 71

(CO5,CO4) [Application]