Roll No											

PRESIDENCY UNIVERSITY
BENGALURU

SET B

SCHOOL OF ENGINEERING
END TERM EXAMINATION - JAN 2024

Semester : Semester VII -2020
Course Code : CSE2021
Course Name : Data Mining
Program : B.Tech.

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. Define data mining
(CO1) [Knowledge]
2. State the formula of covariance.
(CO2) [Knowledge]
3. Recite the formula of support and confidence.
(CO3) [Knowledge]
4. Why predictions are used in supervised learning? Describe.
(CO4) [Knowledge]
5. Describe DBSCAN Clustering.
(CO5) [Knowledge]

PART B

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

6. Explain the steps of KDD with suitable diagram.
(CO1) [Comprehension]
7. Compute Min-max, Z-Score normalization from the following data:
1500, 3075, 405, 1910, 920, 6450, 725 and 1850.
(CO2) [Comprehension]

8. Prepare the FP-Growth tree from the given table:

Transaction ID	Items
T1	A C D F E L M
T2	N M C A F
T3	A B F E L
T4	C B A M N F
T5	D A C B H L M N

(CO3) [Comprehension]

9. Predict Buys_Computer value using navie bayes classifier, when (age <= 30 , income = medium, student = yes, credit_rating = fair) from the given 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) [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. Construct decision tree by CART Algorithm using the following 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

(CO4,CO3) [Application]

12. Calculate Agglomerative clustering using single linkage, complete linkage and average linkage from the given data:
X: 100, 115, 200, 175, 190, 80, 40 Y: 50, 75, 100, 95, 80, 120, 140

(CO5,CO4) [Application]