PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING
END TERM EXAMINATION - JAN 2023

Semester : Semester V - 2020
Course Code : CSE2021
Course Name : Sem V - CSE2021 - Data Mining
Program : B.Tech. COM/CSE/CBC/CBD/CSD/CST/ISE/IST

Date : 13-JAN-2023
Time : 9.30AM - 12.30PM
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.

PART A

ANSWER ALL THE TEN QUESTIONS   **10 X 2 = 20M**

1. What are the main goals of data mining?

(CO1) [Knowledge]

2. List important data mining techniques with example ?

(CO1) [Knowledge]

3. What are the different methods to fill the missing values for an attribute?

(CO2) [Knowledge]

4. What is the purpose of chi square test ?

(CO2) [Knowledge]

5. Write the apriori algorithm for discovering frequent itemsets for mining.

(CO3) [Knowledge]

6. Explain FP growth in detail ?

(CO3) [Knowledge]

7. Explain the working steps of naive bayes classification?

(CO4) [Knowledge]

8. What is the importance of info gain value in selecting root node with formula?

(CO4) [Knowledge]

9. Define centroid and Eucledean Distance.

(CO5) [Knowledge]

10. How does the cluster continue to grow in Density based method?

(CO5) [Knowledge]

PART B

ANSWER ALL THE FIVE QUESTIONS   5 X 10 = 50M

11. Describe in detail the KDD process with neat diagram?

                                                        (CO1) [Comprehension]

12. Briefly discuss in detail Euclidean Distance also find Euclidean distance between the objects represented by the tuples (10, 0, 36) and (12, 1, 32)

                                                        (CO2) [Comprehension]

13. For the given transactional data find the interesting pattern singApriori algorithm have min sup= 2 and min –conf= 70%,

TID	LIST OF Item Sets
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

                                                        (CO3) [Comprehension]

14. John is a student in computer science that loves listening to music. Sometimes, he has homework to do, it can be programming homework, or else. We have some examples of the type of music he listens to, according to some features.

Time of Day	Homework Due	Programming	Music Type
Morning	Yes	No	Classical
Morning	No	No	Pop
Morning	No	Yes	Classical
Morning	Yes	No	Classical
Afternoon	Yes	Yes	Pop
Afternoon	No	No	Pop
Evening	No	**Yes**	Pop
Evening	Yes	**Yes**	Classical

Assume you saw John in the morning, he had homework to do that doesn't require programming. What kind of music would Naive Bayes predict?

                                                        (CO4) [Comprehension]

15. Explain about Partationing clustering with algorithm and slove the problem by using ecludiean measure:

S.No	X	Y	Z
1	16	16	17
2	20	20	21
3	21	22	23
4	30	36	41
5	42	43	44

                                                        (CO5) [Comprehension]

ANSWER ALL THE TWO QUESTIONS   **2 X 15 = 30M**

16. A database has five transactions. Let min sup = 60% and min confidence= 80%.

Transaction ID	Items Bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Construct FP-growth tree and find Conditional Pattern Base and Conditional FPTree.

(CO5,CO3) [Application]

17. Construct decision tree for the following training data set also validate Q1, Q2 and Q3.
    Training data

Instance	A	B	C	Class
11	0	0	0	+
12	0	0	1	+
13	0	1	0	+
14	0	1	1	−
15	1	0	0	+
16	1	0	0	+
17	1	1	0	−
18	1	0	1	+

Validation data

Instance	A	B	C	Class
Q1	0	0	0	+
Q2	0	1	1	+
Q3	1	0	1	−

(CO4) [Application]

* * * * *