TB256256M

Reg. No :.....

Name :.....

BACHELOR'S DEGREE (C.B.C.S) EXAMINATION, MARCH 2025

2018, 2019, 2020, 2021 ADMISSIONS SUPPLEMENTARY

SEMESTER VI - CHOICE BASED CORE (COMPUTER APPLICATIONS & B.VOC S.W.D)

CA6B12AB18 - Data Mining

Time : 3 Hours **Maximum Marks : 80**

Part A

I. Answer any Ten questions. Each question carries 2 marks **(10x2=20)**

1. Define Data mining.
2. Out of Sample of 120 persons in a village, 76 were administered a new drug for preventing influenza and out of them 24 persons were attacked by influenza. Out of those who were not administered the new drug , 12 persons were not affected by influenza. Prepare a 2X2 table showing the actual frequency.
3. Out of Sample of 130 health workers in a village, 86 were administered a new drug for preventing influenza and out of them 24 persons were attacked by influenza. Out of those who were not administered the new drug , 12 persons were not affected by influenza. Prepare a 2X2 table showing the actual frequency.
4. Appraise data quality.
5. Define Data mart.
6. Define Enterprise warehouse.
7. Define Entropy in the Decision tree.
8. Explain Frequent Patterns.
9. List out fundamental clustering methods.
10. Illustrate the functioning of K-means algorithm.
11. Define sequence .
12. List out the applications of spatial data mining.

Part B

II. Answer any Six questions. Each question carries 5 marks **(6x5=30)**

13. Prepare several disciplines that strongly influence the development of data mining methods.
14. For the one dimensional data set 7,10,20,28,35 perform hierarchical clustering and plot the dendogram to visualize it.
15. Discuss multidimensional data model schema.
16. Discuss Baye's Theorem and Naive Baye classifier.
17. Use MinMax and Z-Score Method to Normalize the given data 200,300,400,600,1000(Set min=0 and max=1)
18. List out the advantages and disadvantages of K- means clustering methods.
19. Illustrate the two statistical distance methods used in K-means clustering algorithm.
20. Discuss complex data types.
21. Appraise the application of Data Mining for Intrusion Detection and Prevention.

Part C

III. Answer any Two questions. Each question carries 15 marks **(2x15=30)**

22. Explain data mining as a step in the process of knowledge discovery.
23. Discuss typical OLAP operations.

24. Find the association rule for the following Database D using Apriori algorithm.

Min support = 2 Min. confidence = 50%

Database D

TID	Items
T1	I1, I2, I5
T2	I2, I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

25. Find the value of decision class for the given Combination (Red, SUV, Dom) from the given training data using Bayesian Classification.

Car No	Color	Type	Origin	Stolen
1	Red	Sports	Dom	Yes
2	Red	Sports	Dom	No
3	Red	Sports	Dom	Yes
4	Yellow	Sports	Dom	No
5	Yellow	Sports	Imp	Yes
6	Yellow	SUV	Imp	No
7	Yellow	SUV	Imp	Yes
8	Yellow	SUV	Dom	No
9	Red	Sports	Imp	No
10	Red	SUV	Imp	Yes