TB206270W **Reg. No :.....**

Name :.....

B. Sc. DEGREE (C.B.C.S.) EXAMINATION, MARCH 2023

(2020 Admission Regular, 2019, 2018 Admissions Supplementary)

SEMESTER VI - CHOICE BASED CORE (COMPUTER APPLICATIONS)

(Common for Computer Applications & B. Voc. SWD)

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 discrimination.
2. Suppose that the data for analysis includes the attribute age. The age values for the 10 data tuples are (in increasing order) 13, 15, 20,21, 22, 22, 25, 25, 33, 35, 52, 70. What is the mean of the data? What is the median?
3. In a certain sample of 2000 senior citizens, 1800 senior citizens are vaccinated, out of 1400 males, 1236 male senior citizens were vaccinated.Prepare a 2X2 table showing the actual frequency.
4. Recall metadata Repository.
5. Define Data Integration.
6. Appraise data quality.
7. Elaborate on Bayesian Classifiers.
8. Illustrate parameter tuning in KNN algorithm.
9. Illustrate two commonly used partitioning methods.
10. Explain the general characteristics of hierarchical methods.
11. Illustrate Graph Pattern Mining.
12. Discuss multimedia mining.

Part B

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

13. Explain different forms of data integration.
14. In a certain sample of 2000 families, 1200 students are consumers of milk, out of 1600 males , 1096 male students consumes milk. Use chi-square test and state whether there is any significant difference between consumption of milk among male and female students. [Table value of χ^2 at 5% level of significance for degrees of freedom 1 is 3.84]
15. Discuss ETL.
16. Compare and contrast KNN and SVM algorithm.
17. Formulate the application of Market Basket analysis in Marketing.
18. Justify the application of clustering in the field of outlier analysis.
19. Compare and contrast Hierarchical and Density-based clustering methods.
20. Analyse the application of Data mining in computer science.
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. Describe major challenges to data mining regarding user interaction issues.
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. Using K-Means Clustering Algorithm divide the given dataset into two clusters.

ID	X	Y
1	1	1
2	1.5	2
3	3	4
4	5	7