QP CODE: 24803811

24803811

Reg No : .....

Name : .....

INTEGRATED MSC DEGREE EXAMINATION, JULY 2024

Fourth Semester

CORE - ICSC4CR3 - DATA MINING

INTEGRATED MSC COMPUTER SCIENCE-ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING,INTEGRATED MSC COMPUTER SCIENCE-DATA SCIENCE

2020 Admission Onwards

142322DB

Time: 3 Hours Weightage: 30

Part A (Short Answer Questions)

*Answer any **eight** questions.*

*Weight **1** each.*

1. What is the difference between data and information?
2. Explain Qualitative attribute.
3. Name the different data preprocessing techniques.
4. What is data transformation? Name the activities involved in the same.
5. What are the components of apriori algorithm?
6. What are the advantages of KNN?
7. What is meant by k-fold cross validation?
8. What is k-medoids clustering?
9. What is agglomerative hierarchical clustering?
10. What is meant by grid based clustering methods?

(8×1=8 weightage)

Part B (Short Essay/Problems)

*Answer any **six** questions.*

*Weight **2** each.*

11. Give detailed information about Metadata in data warehousing.
12. What is data redundancy and explain the techniques to deal with the issue?
13. Explain the different techniques for dimensionality reduction.

14. Elaborate on market basket analysis with an example.

15. Briefly explain different techniques for measuring classifier performance.

16. Give the general algorithm for decision tree induction.

17. Explain different aspects with which clustering techniques can be compared.

18. Explain the advantages and limitations of DBSCAN.

(6×2=12 weightage)

Part C (Essay Type Questions)

*Answer any **two** questions.*

*Weight **5** each.*

19. Explain the various data mining issues.

20. Consider the below the transaction database and generate the frequent itemset using FP-Growth algorithm where the minimum support count is 2 .

TID	Items
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	{a,b,c}
6	{a,b,c,d}
7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}

21. Describe issues in decision tree induction in detail.

22. Explain distance based outlier detection in detail.

(2×5=10 weightage)