Reg. No ....   Name **23U418**

END SEMESTER EXAMINATION : MARCH 2023
SEMESTER 4 : INTEGRATED M.Sc. PROGRAMME COMPUTER SCIENCE
COURSE : **21UP4CRMCP11 : DATA MINING**
(For Regular -2021 Admission)

Time : Three Hours   **Max. Weightage: 30**

PART A
Answer any 8

1. Define a closed itemset.
2. Define incremental data mining.
3. Define the concept - web usage mining.
4. State the basic idea behind grid-based clustering methods.
5. State any two drawbacks of decision tree induction algorithm.
6. State the function of ETL tool.
7. Write the formula that calculates the Gini index.
8. List any two real life examples of a frequent itemset.
9. List any two features of Apriori algorithm.
10. Define field overloading of attributes.

(1 x 8 = 8 Weight)

PART B
Answer any 6

11. Prepare a short note on distance-based approaches to outlier detection.
12. Discuss the basic terminologies in association analysis.
13. Write short notes on Information Gain of an attribute.
14. With a graphical representation, explain how the measures of central tendency vary in symmetry.
15. Assuming a numeric attribute, explain how the attribute values can be normalized using decimal scaling.
16. Explain how association rules are generated from frequent itemset.
17. Discuss the need of data integration with an example.
18. There are three possibilities for partitioning tuples based on the splitting criterion in a decision tree. Explain all the three possibilities of splitting.

(2 x 6 = 12 Weight)

PART C
Answer any 2

19. With an example, explain the CART algorithm.
20. With an example, explain the k-medoids method with respect to cluster analysis.
21. Prepare a detailed note on association rule mining.
22. Describe the major tasks involved in data preprocessing.

(5 x 2 = 10 Weight)