Reg. No ..... Name **25U419**

END SEMESTER EXAMINATION - MARCH 2025

SEMESTER 4 : INTEGRATED M.Sc. PROGRAMME COMPUTER SCIENCE - DATA SCIENCE

COURSE : **21UP4CRMCP11 : DATA MINING**

(For Regular 2023 Admission and Improvement/Supplementary 2022/2021 Admissions)

Time : Three Hours Max. Weightage: 30

PART A

Answer any 8

1. List any two reasons that would cause data to be incomplete.
2. State the basic idea behind grid-based clustering methods.
3. Find the midrange from the following data:
   89, 77, 88, 91, 88, 93, 99, 79, 87, 84, 86, 82, 88, 89, 78
4. List any two techniques used for correlation analysis.
5. State how a market basket is represented for analysis.
6. State any two real life examples of data outliers.
7. Name any two algorithms that are used for classification and prediction.
8. Define a closed itemset.
9. To improve the efficiency of the level-wise generation of frequent itemsets, an important property called the Apriori property is used to reduce the search space. Define the Apriori property.
10. Name any two algorithms that are used for decision tree induction.

(1 x 8 = 8 Weight)

PART B

Answer any 6

11. When a decision tree is built, many of the branches will reflect anomalies in the training data due to noise or outliers. Explain how the problem can be addressed.
12. Explain the various aspects that should be considered for data integration.
13. Explain how association rules are generated from frequent itemset.
14. Explain how histograms can be used in outlier detection.
15. Present an example where data mining is crucial to the success of a business, by briefing its importance.
16. Discuss the basic terminologies in association analysis.
17. Write short note on Information Gain of an attribute.
18. Write short notes on linear regression with respect to data reduction strategies.

(2 x 6 = 12 Weight)

PART C

Answer any 2

19. With an example, explain a centroid-based technique in cluster analysis.
20. With an example, explain how Chi-square test helps in data integration.
21. With an example, explain the k-nearest neighbor algorithm.
22. Make a detailed note on enhancing the performance of Apriori algorithm.

(5 x 2 = 10 Weight)