DECEMBER 2023 **P/ID 17621/PCAE5**

Time : Three hours Maximum : 80 marks

PART A — (10 × 2 = 20 marks)

Answer any TEN questions.

1. Define the term Analytics.

2. What is Stratified Sampling?

3. What are Filters? List out the filters that are used for variable selection with respect to continuous and Categorical Variables.

4. What is a decision tree?

5. What are Ensemble Methods? What is its advantages?

6. Define Clustering.

7. What is Survival Analysis? List out any two examples where survival Analysis is useful.

8. Define a Social Network. Give some examples.

9. What are bigraphs? Give some applications of Bi-graphs.

10. Explain the term Benchmarking.

11. List out the privacy issues with respect to data collected for data Analytics.

12. Mention any four open source data analytics tools.

PART B — ($5 \times 6 = 30$ marks)

Answer any FIVE questions.

13. Discuss briefly on the various sources of data.

14. Write short notes on Random forests.

15. Discuss on the various methods for outlier detection.

16. Explain the measurements in Survival Analysis.

17. Discuss on the various phases of data quality Management program.

18. What is churn prediction? Explain churn prediction process with a neat diagram.

19. What is social Network Learning? Explain the various components of a Social Network Learner.

PART C — ($3 \times 10 = 30$ marks)

Answer any THREE questions.

20. Discuss briefly on the requirements of an Analytical Model.

21. Write Short notes on the following.

(a) Self Organizing Maps.

(b) Egonets

22. Discuss on the various data quality dimensions.

23. Explain the various activities involved in the development of a Back testing framework.

24. Write short notes on Recommender Systems.

———————