# LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

**M.Sc.** DEGREE EXAMINATION – **STATISTICS** 

## FOURTH SEMESTER - APRIL 2022

## PST 4502 – DATA WAREHOUSING AND DATA MINING

Dept. No. Date: 17-06-2022 Max.: 100 Marks Time: 01:00 PM - 04:00 PM

#### Part A

- 1. What are the three strategies to develop valuable insights that provides greater business value?
- 2. Define Data Mart in data warehousing.
- 3. State the definition of Schema in data warehousing.
- 4. List out the name of three types of data models.
- 5. How do you calculate Classification error impurity in decision tree?
- 6. Define Decision Stump in gradient boosting.
- 7. What is pruning in data mining?
- 8. State the formula for net input and net output in ANN classifier.
- 9. What is boosting in data mining?
- 10. Define Lift in Apriori association rule mining.

### Part B

### Answer any FIVE questions

Answer all the questions

- 11. What is schema in data warehousing? Also explain any two schemas.
- 12. What are the factors used to determine which index technique should be built on a column? Explain.
- 13. Explain any four Administrator role in data warehousing.
- 14. Describe the applications of data mining tools in any five situations
- 15. What are the advantages and disadvantages of Classification tree model?
- 16. Explain splitting, stopping and pruning in regression tree method.
- 17. What are the common strategies used in decision tree models.
- 18. How does Gradient boosting algorithm work?

### Part C

### Answer any TWO questions

- 19. a) Explain Horizontal and vertical partitions in data warehousing. (16)
  - b) What are online analytical processing operations? (4)
- 20. Explain CRISP Data mining Architecture. (20)
- 21. a) Why is data mining popular now? Explain. (10) b) How does Regression tree algorithm work? (10)
- 22. Explain the Naïve Bayesian classifier for classification problem. (15)
  - b) What are the merits and demerits of Naïve Bayesian Classifier? (5)

 $2 \times 20 = 40$ 

5 X 8 = 40

10 X 2 = 20