

LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034



M.Sc. DEGREE EXAMINATION – STATISTICS

FOURTH SEMESTER – APRIL 2023

PST 4502 – DATA WAREHOUSING AND DATA MINING

Date: 03-05-2023

Dept. No.

Max. : 100 Marks

Time: 09:00 AM - 12:00 NOON

Part A

Answer all the questions

(10 X 2 =20)

1. What is meant by metadata in data warehouse?
2. State different types of storage in data warehouse.
3. Define staging area.
4. List out the name of indexes.
5. Define high data cardinality.
6. What is pruning in data mining?
7. State any two issues in data mining.
8. Define bagging in data mining.
9. How do you measure accuracy rate from confusion matrix?
10. What are the uses of EM algorithm?

Part B

Answer any five questions

(5 X 8 =40)

11. Elucidate different types of Horizontal partition in data warehousing
12. What are the roles of system administrator in data warehousing? Explain.
13. Describe Star schema in data warehousing.
14. What are the three important types of data model? Explain.
15. Explain Indexes in data warehousing?
16. Explain the following in data mining:
 - a. Splitting
 - b. Stopping
 - c. Pruning
17. How does Apriori algorithm work in rule mining?
18. Find out the first most important variable from the following dataset (dependent variable is play) using Classification tree algorithm.

Outlook	humidity	windy	play
rain	moderate	TRUE	TRUE
rain	moderate	TRUE	FALSE
rain	moderate	TRUE	FALSE
sunny	moderate	TRUE	FALSE
rain	moderate	FALSE	TRUE
rain	moderate	FALSE	TRUE
sunny	moderate	FALSE	FALSE
rain	moderate	TRUE	TRUE
rain	moderate	TRUE	TRUE
rain	moderate	FALSE	TRUE
sunny	high	TRUE	FALSE
sunny	high	TRUE	FALSE
rain	high	TRUE	FALSE
rain	high	TRUE	TRUE

rain	high	TRUE	TRUE
sunny	high	TRUE	TRUE
rain	high	TRUE	FALSE

Part C

Answer any TWO questions

(2 X 20 =40)

19. a) What are the kinds of data used in data mining?

b) Explain the various issues in data mining.

(10+10)

20. Explain CRISP Data Mining Architecture.

21. Describe the classification tree algorithm and its advantages and disadvantages.

22. How does Naïve Bayesian classifier algorithm works in data mining? Explain advantages and disadvantages of Naïve Bayesian classifier.

(14+6)

\$\$\$\$\$\$