

Roll No.

--	--	--	--	--	--	--	--	--	--	--	--

Total No. of Pages : 02

Total No. of Questions : 09

B.Tech. (CSE) (Sem.-7,8)
DATA MINING AND DATA WAREHOUSING
Subject Code : BTCS702-18
M.Code : 90488
Date of Examination : 26-11-2024

Time : 3 Hrs.

Max. Marks : 60

INSTRUCTIONS TO CANDIDATES :

1. SECTION-A is COMPULSORY consisting of TEN questions carrying TWO marks each.
2. SECTION-B contains FIVE questions carrying FIVE marks each and students have to attempt any FOUR questions.
3. SECTION-C contains THREE questions carrying TEN marks each and students have to attempt any TWO questions.

SECTION-A

1. Write briefly :

- a) What are the similarities between data warehouse and database?
- b) Describe different types of data cube operations.
- c) Let $x_1 = (4, 8, 2)$ and $x_2 = (1, 3, 8)$ represent two points in space. Find the Euclidean distance and Manhattan distance between x_1 and x_2 .
- d) Compute the information gain for the columns age and income for the Table given in Question 16.
- e) What are the limitations of Apriori algorithm for association rule mining?
- f) Write the general characteristics of hierarchical based clustering methods.
- g) Discuss briefly about the enterprise search engine software.
- h) Which software is used for web data mining?
- i) What are the characteristics of OLAP?
- j) How clustering is done in large databases?

SECTION-B

2. How data cube operations help in data warehousing?
3. Write a short note on Frequent Pattern mining without candidate generation.
4. Given the following confusion matrix for two classes "positive" and "negative", compute the performance measures i.e. accuracy, error rate, sensitivity, specificity and precision.

		Predicted Class	
		Positive	Negative
Actual Class	Positive	6954	46
	Negative	412	2588

5. Discuss the search engine architecture in detail.
6. Discuss and differentiate between web usage mining and web structure mining.

SECTION-C

7. Develop a decision tree for the following data (Use Gini index as the splitting criterion):
Class-Labeled Training Tuples from the *All Electronics* Customer Database

RID	Age	Income	Student	Credit Rating	Class : Buys Computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	Middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	Middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	Middle_aged	medium	no	excellent	yes
13	Middle_aged	high	yes	fair	yes
14	Senior	medium	no	excellent	no

8. Describe in detail and give examples of each of the following approaches to clustering: partitioning methods, hierarchical methods and density-based methods.
9. What is data preprocessing? What are different methods of data preprocessing? Explain in detail.

NOTE : Disclosure of Identity by writing Mobile No. or Making of passing request on any page of Answer Sheet will lead to UMC against the Student.