

Roll No.

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

Total No. of Pages : 02

Total No. of Questions : 18

B.Tech.(CSE) (Sem.-7,8)

DATA MINING AND DATA WAREHOUSING

Subject Code : BTCS702-18

M.Code : 90488

Date of Examination: 21-12-2023

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

Answer Briefly:

1. What are the differences between data warehouse and database?
2. What is a data cube? What are its characteristics?
3. Let $x_1 = (1, 2, 3)$ and $x_2 = (3, 7, 5)$ represent two points in space. Find the Euclidean distance and Manhattan distance between x_1 and x_2 .
4. Discuss briefly about direct hashing.
5. What are the limitations of decision tree classification method?
6. Write the general characteristics of partitioning based clustering methods.
7. What are the characteristics of search engine?
8. Write four characteristics of web mining.
9. What are the challenges in data mining?
10. What are different types of similarity measures used in data mining?

SECTION-B

11. How the OLAP queries can be efficiently processed?
12. Write a short note on Naive Bayes classification method.
13. 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

14. Discuss the ranking of web pages in detail.
15. How locality and hierarchy help in web data mining? Give examples.

SECTION-C

16. Apply Apriori algorithm for association rule mining on the following data:

Transactional Data for All Electronics

Branch TID	List of item IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

17. Suppose that the data mining task is to cluster points (with (x, y) representing location) into three clusters, where the points are

A1(2,10), A2(2,5), A3(8,4), B1(5,8), B2(7,5), B3(6,4), C1(1,2), C2(4,9).

The distance function is Euclidean distance. Suppose initially we assign A1, B1 and C1 as the center of each cluster, respectively. Use the k-means algorithm to show

- a) The three cluster centers after the first round of execution.
 - b) The final three clusters.
18. What are different types of data used in data mining? Discuss in detail with examples.

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.