

Reg. No. :

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

Question Paper Code: U4E04

B.E./B.Tech. DEGREE EXAMINATION, APRIL 2024

Fourth Semester

Artificial Intelligence and Data Science

21UAD404 - DATA WAREHOUSING AND DATA MINING

(Regulations 2021)

Duration: Three hours

Maximum: 100 Marks

Answer All Questions

PART A - (10 x 2 = 20 Marks)

1. Compare data warehouse from a database? How are they similar? CO1-U
2. Apply the Galaxy schema for a college library and draw the proper schema for the above. CO2-App
3. Differentiate fact and dimension table. CO1-U
4. What is snow flake schema? CO1-U
5. State why concept hierarchies are useful in data mining. CO2-App
6. What are the process of knowledge discovery in databases (KDD) CO1-U
7. What is meant by Lazy Learner? CO1-U
8. What is MBA in Data mining (candidate generation technique)? CO1-U
9. State the difference between classification and clustering? CO1-U
10. Compare CLARA and CLARANS. CO2-App

PART – B (5 x 16= 80 Marks)

11. (a) Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit. CO2-App (16)
 - (i) Enumerate three classes of schemas that are popularly used for modeling data warehouses. (8 Marks)
 - (ii) Draw a schema diagram for the above data warehouse using one of the schema classes listed in (a) (8 Marks)

Or

- (b) Construct a data warehouse for a University / Hospital / Enterprise using Galaxy schemas with necessary description CO2-App (16)
12. (a) Define in detail about the OLAP Operations in Multi-dimensional Data Model. CO1-U (16)
- Or
- (b) (i) Differentiate Star schema vs Snow flake schema vs Galaxy schema (10 Marks) CO1-U (16)
(ii) With relevant examples discuss the different schema operations. (6 Marks)
13. (a) (i) Explain in detail about Data mining functionalities. (8 Marks) CO1-U (16)
(ii) Explain in detail about Interestingness of patterns in data mining. (8 Marks)
- Or
- (b) Define issues in Data mining and Data preprocessing in data mining. CO1-U (16)
14. (a) Apply the Apriori algorithm for discovering frequent item sets for mining association rules of the following table. Use 3 for the minimum support value and Confidence of 50%. Illustrate each step of the Apriori algorithm. CO2-App (16)

Trans ID	Items Purchased
101	milk, bread, eggs
102	milk, juice
103	juice, butter
104	milk, bread, eggs
105	coffee, eggs
106	coffee
107	coffee, juice
108	milk, bread, cookies, eggs
109	cookies, butter
110	milk, bread

Or

- (b) Suppose we have a dataset of weather conditions and corresponding target variable "Play". So using this dataset we need to decide that whether we should play or not on a particular day according to the weather conditions. CO2-App (16)

Problem: If the weather is sunny, then the Player should play or not?

	Outlook	Temperature	Humidity	Windy	Play Golf
1	Rainy	Hot	High	False	No
2	Rainy	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Sunny	Mild	High	False	Yes
5	Sunny	Cool	Normal	False	Yes
6	Sunny	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Rainy	Mild	High	False	No
9	Rainy	Cool	Normal	False	Yes
10	Sunny	Mild	Normal	False	Yes
11	Rainy	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Sunny	Mild	High	True	No

15. (a) Let's say we want to cluster a group of 20 individuals between the ages of 20 and 40. We have collected data on their ages, which are as follows: CO2-App (16)

25, 22, 28, 36, 32, 23, 27, 30, 31, 29, 33, 24, 26, 34, 37, 38, 21, 35, 39, 40

Our goal is to divide these individuals into two clusters based on their age using the k-means algorithm.

Or

- (b) (i) Select the suitable example to compare and analyze the systematic way of implementing agglomerative and Divisive hierarchical clustering. (10 Marks) CO2-App (16)
- (ii) Compare and contrast the CLARA and CLARANS. (6 Marks)

