Reg. No. :

Question Paper Code: U4E04

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

Second Semester

Artificial Intelligence & Data Science

21UAD404 - DATA WAREHOUSING AND DATA MINING

(Regulations 2021)

Duration: Three hours   Maximum: 100 Marks

PART A - (10 x 2 = 20 Marks)

1. Illustrate a data warehouse from a database? How are they similar?   CO1-U

2. Define the benefits of data warehousing.   CO1-U

3. Apply the HOLAP process for an engineering problem with proper description. CO2-App

4. What is fact constellation?   CO1-U

5. List down the major tasks involved in data cleaning process   CO2-App

6   What are the essential steps in the process of knowledge discovery in databases CO1-U
    (KDD)

7   Give few techniques to improve the efficiency of Apriori algorithm.   CO1-U

8   When the association rules are interesting?   CO1-U

9   List the types of data used in cluster analysis.   CO1-U

10 Compare CLARA and CLARANS?   CO1-U

PART – B (5 x 16= 80 Marks)

11. (a) Identify and apply appropriate data warehousing schema techniques   CO2-App   (16)
         for the result analysis of our SIT Exam cell.

Or

(b) Design a data warehouse for a regional weather bureau. The weather bureau has about 1,000 probes, which are scattered throughout various land and ocean locations in the region to collect basic weather data, including air pressure, temperature, and precipitation at each hour. All data are sent to the central station, which has collected such data for over 10 years. Your design should facilitate efficient querying and on-line Analytical processing, and derive general weather patterns in multidimensional space.　　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)　　CO1-U　(16)

(ii) With relevant examples discuss the different schema operations. (6)

13. (a) Define and Discuss about the Integration of a Data Mining system with a Data Warehouse.　　CO3-Ana　(16)

Or

(b) Describe about the Classification in DataMining Systems and explain the various types of classification algorithm　　CO3-Ana　(16)

14. (a) Make use of Apriori algorithm to find the support and confidence from the following transaction table　　CO2-App　(16)

TID	ITEMSETS
T1	A, B
T2	B, D
T3	B, C
T4	A, B, D
T5	A, C
T6	B, C
T7	A, C
T8	A, B, C, E
T9	A, B, C

Given: Minimum Support= 2, Minimum Confidence= 50%

U4E04

Or

(b) Apply Frequency pattern growth for discovering frequent item sets   CO2-App (16)
for mining association rules of the following table.

Trans ID	Items Purchased
101	milk, bread,eggs
102	milk, juice
103	juice,b u t t e r
104	milk,bread,eggs
105	coffee,eggs
106	coffee
107	coffee , juice
108	milk, read,cookies,eggs
109	cookies, butter
110	milk , bread

15. (a) Explain in detail about the method involved in hierarchical clustering   CO1-U (16)
and Write the difference between the partitioning method and
Hierarchical method

Or

(b) Explain with an example about the partitioning based methods with   CO1-U (16)
neat diagram.

U4E04