Reg. No. : | | | | | | | | | | |

Question Paper Code: U4E04

B.E./B.Tech. DEGREE EXAMINATION, NOV 2025

Fourth 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.	State why data preprocessing is an important issue for data warehousing and data mining?	CO1-U
2.	List down the major tasks involved in data cleaning process	CO1-U
3.	Write down any one of the multi-dimensional data tools for an engineering problem with proper description	CO2-App
4.	Differentiate between OLTP vs. OLAP.	CO1-U
5.	State why concept hierarchies are useful in data mining.	CO2-App
6	What are the essential steps in the process of knowledge discovery in databases (KDD)	CO1-U
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)	Construct a data warehouse for a University / Hospital / Enterprise using Galaxy schemas with necessary description.	CO2-App	(16)

Or

(b) Suppose your task as a software engineer at Big-University is to design a data mining system to examine their university course database, which contains the following information: the name, address, and status (e.g., undergraduate or graduate) of each student, the courses taken, and their cumulative grade point average (GPA). Describe the architecture you would choose. What is the purpose of each component of this architecture? 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)
(ii) With relevant examples discuss the different schema operations. (6) CO1-U (16)

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%

Or

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

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 centroid based techniques (k-means) and object based techniques (k-mediods) with relevant example. CO1-U (16)

Or

(b) With relevant example discuss constraint based cluster analysis. CO1-U (16)