BACHELOR OF TECHNOLOGY (CBCS) (2021-COURSE)

B. Tech. Sem - IV Computer Science & Engineering AI & ML : WINTER : 2023

SUBJECT : ITC-II: DATA WAREHOUSING & MINING

Day : Thursday

Date : 23-11-2023

Time : 02:30 PM-05:30 PM

W-23949-2023

Max. Marks : 60

N.B.

1) All questions are **COMPULSORY**.
2) Figures to the **RIGHT** indicate **FULL** marks.
3) Assume suitable data **WHEREVER** necessary.
4) Draw neat labeled diagrams **WHEREVER** necessary.

Q.1 Explain various OLAP operations in details with diagram. **(10)**

OR

Q.1 Describe the process of Extraction, Transformation and Loading with neat and labeled diagram. **(10)**

Q.2 With neat and clean diagram explain schemas for multidimensional data modal in details. **(10)**

OR

Q.2 What is cube materialization? Explain in details. **(10)**

Q.3 What is Bitmap Indexing? Explain its use and implementation in details. **(10)**

OR

Q.3 Differentiate between ROLAP, MOLAP and HOLAP. **(10)**

Q.4 What is Discretization in data mining? Explain techniques of Data Discretization. **(10)**

OR

Q.4 What is CRISP-Data Mining? Explain its phases in details. **(10)**

Q.5 Car theft problem : Attributes are color, type, origin and the subject stolen can be either Yes or No. solving using Naïve Bayes Classification algorithm. **(10)**

Dataset

Car no.	Color	Type	Origin	stolen
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

OR

Q.5 A dataset has five transactions. Let min-support = 60% and min-confidence=80%. Find frequent item sets by using Apriori Algorithm. T–ID is the transaction ID. **(10)**

T_ID	Items bought
T-1000	M,O,N,K,E,Y
T-1001	D,O,N,K,E,Y
T-1002	M,A,K,E
T-1003	M,U,C,K,Y
T-1004	C,O,O,K,E,

PTO

Q.6 Differentiate between simple linkage, average linkage and complete linkage algorithms. Use complete linkage algorithm to find the clusters from the following dataset. **(10)**

X	4	8	15	24	24
Y	4	4	8	4	12

OR

Q.6 Enlist and explain major clustering methods in details. **(10)**

*****