BACHELOR OF TECHNOLOGY (CBCS - 2023)
B. Tech. Sem-IV Computer Science & Engineering AI & ML : SUMMER : 2025
SUBJECT: DATA WAREHOUSING & MINING

Day : Wednesday
Date : 04/06/2025

S-29282-2025

Time : 10:00 AM-01:00 PM
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 diagrams WHEREVER necessary.

Q.1 Compare OLTP and OLAP system. (10)

OR

Q.1 Explain multitier architecture of Datawarehouse. (10)

Q.2 Illustrate with various operations and examples of OLAP cube. (10)

OR

Q.2 A social media platform wants to analyze user engagement data to improve content recommendations and user experience. The INTERACTIONS fact table contains information about user interactions, including interaction details, user information, content details, and time periods. The dimension tables provide additional context about users, content, categories, and time periods. (10)
Design a star schema and snowflake schema for the same.

Q.3 Write a short note on a) Efficient data cube computation b) Partial materialization. (10)

OR

Q.3 Describe the basic architectures of an OLAP server. (10)

Q.4 Explain following terms i) data mining ii) data preparation iii) data cleaning. (10)

OR

Q.4 What is data mining? Discuss any 5 applications of data mining. (10)

Q.5 Explain how Naïve Bayes classification makes predictions and discuss the "naïve" assumption in Naïve Bayes. Provide an example to illustrate the application of Naïve Bayes in a real-world scenario. (10)

OR

P.T.O.

Q.5 A company wants to predict whether a customer will subscribe to a premium membership based on their demographic and browsing behavior data. The dataset contains information about customers, including age, gender, income, browsing time, and subscription status. **(10)**

Age	Gender	Income	Browsing Time	Subscription
20-30	Male	High	10am-12pm	Yes
20-30	Female	Medium	2pm-4pm	Yes
30-40	Male	Low	8am-10am	No
30-40	Female	High	4pm-6pm	Yes
>40	Male	medium	6pm-8pm	Yes
>40	Female	Medium	8am-10am	No
>40	Male	High	6pm-8pm	Yes
20-30	Female	Low	10am-12pm	No
20-30	Male	Medium	2pm-4pm	Yes
30-40	Female	High	8am-10am	Yes

Use ID3 to build the decision tree and predict the following example:

Age	Gender	Income	Browsing Time
20-30	Male	medium	10am-12pm

Q.6 Following table gives fat and proteins content of items. Apply single linkage clustering and construct dendrogram. **(10)**

Food Item	Protein	Fat
1	1.1	60
2	8.2	20
3	4.2	35
4	1.5	21
5	7.6	15
6	2.0	55
7	3.9	39

OR

Q.6 Explain K-means clustering algorithm. Discuss its advantages and disadvantages. Apply K-means for Following dataset with 3 clusters. **(10)**

2,3,6,8,9,12,15,18,22

*****