Paper / Subject Code: 31924 / Data Warehousing & Mining

Time: 3 hours **Max. Marks: 80**

==

Note: 1. Question no.1 is compulsory.
2. Attempt any three out of remaining five.
3. Assumptions made should be clearly indicated.
4. Figures to the right indicates full marks.
5.Assume suitable data whenever necessary.

Question 1 **Write a short note on the following. Solve any four.**

(5 marks each)

A Write a note on web usage mining. Also state its any two applications.

B Describe any five issues in data mining.

C Explain how Naive Bayes classification makes predictions and discuss the "naive" assumption in Naive Bayes. Provide an example to illustrate the application of Naive Bayes in a real-world scenario.

D Suppose the data for clustering is {6,14,18,22,1,40,50,11,25} consider k=2, cluster the given data using k means algorithm.

E Explain the concept of market basket analysis with example.

F Differentiate between ER modeling vs Dimensional modeling.

Question 2 **10 marks each**

A Describe in detail about how to evaluate accuracy of the classifier.

B Illustrate major steps in ETL process.

Question 3 **10 marks each**

A Explain KDD process with neat diagram. Also state any five applications of data mining.

B For the table given perform Apriori algorithm and show frequent item set and strong association rules. Assume Minimum Support of 30% and Minimum confidence of 70%.

TID	Items
1	1,4,6,8
2	2,5,3
3	7,1,3,8
4	9,10
5	1,5

Question 4 **10 marks each**

A 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. Design a star schema and snowflake schema for the same.

B Explain Multilevel Association Rules Mining and Multidimensional Association Rules Mining with examples.

Question 5 **10 marks each**

A 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.

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	12pm-2pm	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

B Illustrate page rank algorithm with example.

Question 6 **10 marks each**

A Following table gives fat and proteins content of items. Apply single linkage clustering and construct dendrogram.

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*

B Explain in brief what is data discretization and concept hierarchy generation.