T.E (Sem-V) / Nov-2024 / C-Scheme / 07/12/24
COMP/CSE

Time: 3 hours **Total 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.

Q. 1 **Solve any four.** **(05 marks each)**

A Every data structure in the data warehouse contains the time element. Why?

B Explain FP Growth Algorithm.

C Explain different types of attributes.

D Discuss different applications of Web Mining.

E Explain Holdout and Random subsampling method to evaluate the accuracy of classifier.

F Differentiate between Classification and Clustering.

Q.2 **(10 marks each)**

A For a supermarket chain, consider the following dimensions namely product, store, time and promotion. The schema contains a central fact table for sales with three measures unit_sales, dollars_sales and dollar_cost.

1. Draw a star schema.
2. Calculate the maximum number of base fact table records for warehouse with the following values given below:
   - Time period 5 years
   - Store-300 stores reporting daily sales
   - Product-40,000 products in each store (about 4000 sell in each store daily)
   - Promotion- a sold item may be in only one promotion in a store on a given day.

B Explain the different techniques to handle noisy data.

Suppose a group of sales price records has been sorted as follows:

3, 7, 8, 13, 22, 22, 22, 26, 26, 28, 30, 37.

Partition them into three bins by equal-frequency (Equi-depth) partitioning method. Perform data smoothing by bin mean and bin boundary.

Q.3 **(10 marks each)**

A Explain Updates to dimensional table in detail.

B Explain the following data pre-processing methods.

I) Dimensionality reduction II) Data transformation and Discretization

Q.4 **(10 marks each)**

A Given the training data for height classification, classify the tuple,

t=<Rohit, M, 1.95> using Naïve Bayes Classification.

Name	Gender	Height	Output
Kiran	F	1.6m	Short
Jatin	M	2m	Tall
Madhuri	F	1.09m	Medium
Manisha	F	1.88m	Medium
Shilpa	F	1.7m	Short
Bobby	M	1.85m	Medium
Kavita	F	1.6m	Short
Dinesh	M	1.7m	Short
Rahul	M	2.2m	Tall
Shree	M	2.1m	Tall
divya	F	1.8m	Medium
Tushar	M	1.95m	Medium
Kim	F	1.9m	Medium
Aarti	F	1.8m	Medium
Rajashree	F	1.75m	Medium

B Consider four objects with two attribute (X and Y). These four objects are to be grouped together into two clusters using k-means clustering algorithm. Following are the objects with their attribute values.

Object	X	Y
A	1	1
B	2	1
C	4	3
D	5	4

Q. 5 **(10 marks each)**

A Given the following data, apply the Apriori algorithm. Find frequent item set and strong association rules. Given Support threshold=50%, Confidence=60%

Transaction	Items
T1	I1, I2, I3
T2	I2, I3, I4
T3	I4, I5
T4	I1, I2, I4
T5	I1, I2, I3, I5
T6	I1, I2, I3, I4

B What is Web Mining? Differentiate between Web Mining and Data Mining. Explain types of Web Mining.

Q. 6 **Write short note on.** **(5 marks each)**

A Decision Tree Induction Algorithm

B K-medoids clustering Algorithm

C Multilevel and multidimensional association rule mining

D Page Rank Algorithm

****