NATIONAL INSTITUTE OF TECHNOLOGY TIRUCHIRAPPALLI
END SEMESTER EXAMINATION - JAN. 2021 SESSION

DEPARTMENT : Computer Science and Engineering

DATE & TIME OF EXAM : 13-05-2021 & 11.00 a.m. – 01.00 p.m.

SUB CODE : CS PE14 DURATION: 2 hours

SUB NAME : Data Warehousing and Data Mining

FACULTY NAME : Dr. E. Sivasankar

Answer all Questions **5 x 10 =50 marks**

1. Suppose that a data warehouse for Big University consists of the four dimensions *student*, *course*, *semester*, and *instructor*, and two measures *count* and *avg grade*. At the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the *avg grade* measure stores the actual course grade of the student. At higher conceptual levels, *avg grade* stores the average grade for the given combination.
   (a) Design a data warehouse schema using galaxy schema to facilitate effective online analytical processing in multidimensional space.
   (b) Write the DMQL corresponding to the schema design

2. Suppose your task as a software engineer at a Bank & Financial Institution is to design a data mining system. Describe the architecture you would choose. What is the purpose of each component of this architecture?

3. Construct a Bayesian classifier for the following data. Predict the class label play Tennis for Outlook= Rain, Temperature=Mild, Humidity=Normal and Wind=Strong.

Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

4. A database has ten transactions. Let min support = 30%

Transaction ID	Items Bought
1	{Laptop, Printer, Tablet, Headset}
2	{Printer, Monitor, Tablet}
3	{Laptop, Printer, Tablet, Headset}
4	{Laptop, Monitor, Tablet, Headset}
5	{Printer, Monitor, Tablet, Headset}
6	{Printer, Tablet, Headset}
7	{Monitor, Tablet}
8	{Laptop, Printer, Monitor}
9	{Laptop, Tablet, Headset}
10	{Printer, Tablet}

Find all frequent item sets using Apriori method

. 5. Explain the techniques used for similarity-based retrieval and mining association rules in multimedia databases