Chandigarh College of Engineering and Technology (Degree Wing), Chandigarh
Department of Computer Science and Engineering
Sessional Test -1 (May-2021)

Subject (Code):Data Mining and Analysis (CS605C)   Semester: VIth
Name of Teacher: Dr. Ankit Gupta   Date (Session): 07.05.2021(12-1:30 PM)
Time: 1:30 Hrs   Maximum Marks 30
Note:

1. Attempt all questions.

2. All question carry equal marks.

3. Your answers should not be copied from Google/Internet.

4. Make sure that you upload the answer script well in time before 2:00 PM

5. Ensure to upload clear, readable answer scripts. Late submission will not be considered as you are already given 30 minutes time to upload.

Q.No.	Questions	CO
1.	**a.** Suppose that the minimum and maximum values for the some attribute "income" are $30,000 and $195,000, respectively. We would like to map income to the range [2.0, 12.0]. Calculate the transformed value of $100,000 by min-max normalization. **b.** Explain the concept of data smoothing. How it helps in Data mining process. Consider following data: {0,4,12,16,16,18,24,26,28,32,36,40,44,48,52,60}. Demonstrate the use of Equal frequency binning method on this data set and then apply various data smoothing techniques on it.	CO1
2.	Design a hypothetical Student and staff management system and its database for our university with sufficient number of attributes. Discuss how Attribute Oriented Induction can be helpful on this data base for retrieving information. Assume suitable and missing data yourself.	CO1
3.	**a.** What are the characteristics features of Data Warehouse.? How it affects business decisions? **b.** Explain star net query model with the help of suitable example? (Ensure that the example should not be from textbook)	CO1
4.	**a.** Calculate the number of cuboids in an 10-dimensional cube with 4 levels? **(2)** **b.** Suppose that a dataware house has following 4 dimensions {**time, item, location, supplier**}. Draw the cube corresponding cuboids. Mention the base and apex cuboid. **(3)**	CO1
5.	Compare OLTP and OLAP. Which one of these two can be considered for business decisions and why?	CO1
6.	**a.** Discuss interestingness measure and how it affects overall mining output. **b.** Analyze the Data Generalization concept with the help of suitable example.	CO1