Hall Ticket Number:

Y	1	8	A	C	C	5	7	5	

18CSD-12

III/IV B.Tech (Regular) DEGREE EXAMINATION

February, 2021

Fifth Semester

Time: Three Hours

Computer Science and Engineering

Data ware Housing and Data Mining

Maximum: 50 Marks

Answer ALL Questions from PART-A (1X10 = 10 Marks)

Answer ANY FOUR questions from PART-B (4X10=40 Marks)

Part - A

1. Answer all questions (1X10=10 Marks)
   a) Define Data Warehouse.
   b) How to deal with missing values in an attribute?
   c) What is the importance of a fact table?
   d) What is Data Mining?
   e) How to compute confidence measure for an association rule?
   f) What is Data Mart?
   g) Differentiate qualitative and quantitative attributes.
   h) Define cluster analysis?
   i) What are properties of good Clustering?
   j) Define outlier?

Part - B

2. a) What are Data Mining functionalities? Explain briefly. 5M
   b) What is Data Cleaning? Explain various data cleaning tasks. 5M

3. a) Discuss various issues in Data Mining. 5M
   b) Illustrate the Data Transformation by Normalization. 5M

4. a) Explain Data Warehouse architecture. 5M
   b) What is the difference between operational DBMS and Data Warehouse 5M

5. a) Discuss the star and snowflake schema in detail with suitable example. 5M
   b) Explain various OLAP operations. 5M

6. a) A database has six transactions. Let min-sup = 50% and min-conf = 75%. 5M

TID	List of items
001	Pencil, sharpener, eraser, color papers
002	Color papers, charts, glue sticks
003	Pencil, glue stick, eraser, pen
004	Oil pastels, poster colours, correction tape
005	Whitener, pen , pencil, charts, glue stick
006	Colour pencils, crayons, eraser, pen

   Find all frequent item sets using Apriori algorithm. List all the strong association rules.
   b) Write the advantages and disadvantages of Apriori and FP-growth Algorithm 5M

7. a) What are different methods to improve Apriori algorithm's efficiency? 5M
   b) Explain constraint based rule mining. 5M

8. a) What is the goal of clustering? How does partitioning around medoids algorithm achieve this goal? 5M
   b) Write K-means clustering algorithm. 5M

9. Explain Hierarchical Clustering algorithm. 10M