Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute)

Affiliated to Dr. A.P.J. Abdul Kalam Technical University, Uttar Pradesh, Lucknow

M.Tech

SEM: I - CARRY OVER THEORY EXAMINATION - AUGUST 2022

Subject: Data Warehousing & Data Mining

Time: 03:00 Hours   Max. Marks: 70

General Instructions:

1. The question paper comprises three sections, A, B, and C. You are expected to answer them as directed.

2. Section A - Question No- 1 is 1 marker & Question No- 2 carries 2 marks each.

3. Section B - Question No-3 is based on external choice carrying 4 marks each.

4. Section C - Questions No. 4-8 are within unit choice questions carrying 7 marks each.

5. No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.

SECTION A   15

1. Attempt all parts:-

1   A star schema has_____ relationship between a dimension & fact table? (CO1)   1

                (a) Many to many

                (b) One to One

                (c) One to Many

                (d) All the above

1   _____ is the process of finding a model that describes and distinguishes data classes or   1

concepts. (CO2)

                (a) Data Characterization

                (b) Data Classification

                (c) Data Discrimination

                (d) ORDBMS

1   The basic idea of the Apriori algorithm is to generate _____ item sets of a particular   1

size & scans.(CO3)

                (a) Candidate.

                (b) Primary.

(c) Secondary.

(d) Superkey.

1   _____ clustering technique start with as many clusters as there are records, with each   1
cluster having only one record? (CO4)

       (a) Agglomerative.

       (b) Divisive.

       (c) Partition.

       (d) Numeric.

1   Dimensionality reduction reduces the data set size by removing _____? (CO5)   1

       (a) composite attributes

       (b) derived attributes

       (c) relevant attributes

       (d) irrelevant attributes

2. Attempt all parts:-

2   What is a Data Warehouse? (CO1)   2

2   What is the difference between operational data store and data warehouse? (CO2)   2

2   Compare Roll Up and Drill Down operation? (CO3)   2

2   Define normalization. How do you normalize data using min-max normalization? (CO4)   2

2   What do you mean by Market Basket analysis and how it can help in a supermarket? (CO5)   2

<center>SECTION B   20</center>

3. Answer any <u>five</u> of the following:-

3   Differentiate between Meta Data and Data Dictionary? (CO1)   4

3   Enlist some commands of DDL, DML and DCL? (CO1)   4

3   Explain the data cleaning process in data pre-processing? (CO2)   4

3   Explain the steps of knowledge discovery in databases? (CO2)   4

3   Brief CURE and Chameleon clustering methods? (CO3)   4

3   How can the accuracy of classifier be improved for the class-imbalanced data? (CO4)   4

3   What is meant by text mining ? (CO5)   4

<center>SECTION C   35</center>

4. Answer any <u>one</u> of the following:-

4   Describe the various components of Data warehouse with its architecture? (CO1)   7

4	Discuss about Data Extraction, Cleanup and Transformation Tools? (CO1)	7

5. Answer any <u>one</u> of the following:-

5	Define data discretization. Explain the various approaches in data discretization? (CO2)	7
5	Briefly explain the various Data Reduction techniques, which one is having minimum loss of information content? (CO2)	7

6. Answer any <u>one</u> of the following:-

6	What is decision tree classification, explain its functionality and also write the steps involve in it? (CO3)	7
6	Explain classification by Back Propogation in detail ? (CO3)	7

7. Answer any <u>one</u> of the following:-

7	What are Grid-Based methods. Explain STING in detail? (CO4)	7
7	What is the role of Outlier Analysis in data mining, Explain in detail with example? (CO4)	7

8. Answer any <u>one</u> of the following:-

8	Explain, how MOLAP,ROLAP,HOLAP are related or different from each other? (CO5)	7
8	Differentiate between bottom up tier and middle tier of data warehouse architecture? (CO5)	7