.

Printed Page:-3 **Subject Code:- AMTCSE0114**

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)
M.Tech
SEM: I - THEORY EXAMINATION (2025 - 2026)
Subject: Data Warehousing & Data Mining

Time: 3 Hours **Max. Marks: 70**

General Instructions:

IMP: *Verify that you have received the question paper with the correct course, code, branch etc.*

1. *This Question paper comprises of* ***three Sections -A, B, & C.*** *It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.*

2. *Maximum marks for each question are indicated on right -hand side of each question.*

3. *Illustrate your answers with neat sketches wherever necessary.*

4. *Assume suitable data if necessary.*

5. *Preferably, write the answers in sequential order.*

6. *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. A star schema has_____ relationship between a dimension & fact table. (CO1, K1) 1

(a) Many to many
(b) One to One
(c) One to Many
(d) All the above

1-b. Source data from the warehouse comes from _____. (CO2, K1) 1

(a) ODS
(b) TDS
(c) MDDB
(d) ORDBMS

1-c. The basic idea of the Apriori algorithm is to generate _____ item sets of a particular size & scans.(CO3, K1) 1

(a) Candidate.
(b) Primary.
(c) Secondary.
(d) Superkey.

1-d. The sigmoid function also knows as ____functions. (CO4, K1) 1

(a) Regression
(b) Logistic
(c) Probability
(d) Neural

.

1-e. _____ clustering technique start with as many clusters as there are records, with each cluster having only one record. (CO5, K1) 1

(a) Agglomerative.
(b) Divisive.
(c) Partition.
(d) Numeric.

2. Attempt all parts:-

2.a. Distinguish between Data Mart & Meta Data. (CO1, K3) 2

2.b. Discuss the bottom up approach of data warehouse design. (CO2, K2) 2

2.c. Briefly describe Smoothing. (CO3, K1) 2

2.d. List any two drawbacks of MM-Max normalization. (CO4, K2) 2

2.e. Explain the meaning of Market Basket analysis. (CO5, K2) 2

SECTION-B 20

3. Attempt all parts:-

3.a. Answer any one of the following:-

3.a.(i) Differentiate between Data warehouse and OLAP. (CO1, K3) 4

3.a.(ii) Enlist some commands of DDL, DML and DCL. (CO1, K2) 4

3.b. Answer any one of the following:-

3.b.(i) Explain the data cleaning process in data pre-processing. (CO2, K2) 4

3.b.(ii) Explain the steps of knowledge discovery in databases. (CO2, K2) 4

3.c. Answer any one of the following:-

3.c.(i) Briefly explain CURE and Chameleon clustering methods. (CO3, K2) 4

3.c.(ii) Discuss K-means algorithm and it's importance in data warehousing and data mining. (CO3, K2) 4

3.d. Answer any one of the following:-

3.d.(i) Explain how the accuracy of classifier can be improved for the class-imbalanced data. (CO4, K2) 4

3.d.(ii) Distinguish between Boosting and Ada boosting. (CO4, K3) 4

3.e. Answer any one of the following:-

3.e.(i) Discuss about Multimedia data mining. (CO5, K2) 4

3.e.(ii) Discuss is the role of text mining in datamining. (CO5, K2) 4

SECTION-C 35

4. Answer any one of the following:-

4-a. Describe the various components of Data warehouse with its architecture. (CO1, K2) 7

4-b. Illustrate with an example which schema is better while designing the data warehouse. (CO1, K3) 7

5. Answer any one of the following:-

5-a. Explain parametric and non-parametric methods of Numerosity reduction with suitable examples. (CO2, K2) 7

.

5-b.	Define data discretization. Explain the various approaches in data discretization. (CO2, K2)	7
6. Answer any one of the following:-		
6-a.	Write about decision tree classification, explain its functionality and also write the steps involve in it. (CO3, K3)	7
6-b.	Explain the generalization error in terms of the SVM. (CO3, K2)	7
7. Answer any one of the following:-		
7-a.	Write about the Hierarchical methods of clustering and how they are classified to bottom-up and top-down approach. (CO4, K2)	7
7-b.	Explain the Neural network approach for clustering . Also define SOM in Neural network approach. (CO4, K2)	7
8. Answer any one of the following:-		
8-a.	Discuss a scenario in present where data mining is done in various sectors and is used frequently. (CO5, K3)	7
8-b.	Explain how will you do pattern mining in multilevel and multidimensional space explain in detail. (CO5, K2)	7