Printed Page:- Subject Code:- AMTCSE0114

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 - THEORY EXAMINATION (2021 - 2022)

Subject: Data Warehousing & Data Mining

Time: 03:00 Hours Max. Marks: 70

General Instructions:

1. All questions are compulsory. It comprises of three Sections A, B and C.

- Section A - Question No- 1 is objective type question carrying 1 mark each & Question No- 2 is very short type questions carrying 2 marks each.
- Section B - Question No- 3 is Long answer type - I questions carrying 4 marks each.
- Section C - Question No- 4 to 8 are Long answer type - II questions carrying 7 marks each.
- 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. The data is stored, retrieved and updated in _____ (CO1) 1

1. OLAP
2. OLTP
3. SMTP
4. FTP

1-b. _____ is the process of finding a model that describes and distinguishes data classes or concepts. (CO2) 1

1. Data Characterization
2. Data Classification
3. Data Discrimination
4. ORDBMS

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

1. Candidate.
2. Primary.
3. Secondary.
4. Superkey.

1-d. Which of the following is a predictive model? (CO4) 1

1. Clustering
2. Regression
3. Summarization
4. Association rules

1-e. _____ is a complex chemical process in neural networks. (CO5) 1

1. Receiving process.
2. Sending process.
3. Transmission process.
4. Switching process.

2. Attempt all parts:-

2-a.	Distinguish between Data Mart & Meta Data? (CO1)	2
2-b.	Shortly described Data Transformations? (CO2)	2
2-c.	How can we handle missing values? (CO3)	2
2-d.	Define normalization. How do you normalize data using min-max normalization? (CO4)	2
2-e.	Discuss about k-nearest neighbor classification algorithm with an example? (CO5)	2

SECTION B 20

3. Answer any five of the following:-

3-a.	What is Shared disk architecture? (CO1)	4
3-b.	What are the main features of OLAP? (CO1)	4
3-c.	Explain the data cleaning process in data pre-processing? (CO2)	4
3-d.	What do you understand by data aggregation and data generalization? (CO2)	4
3-e.	What is SVM. What do you mean by kernel in SVM ? (CO3)	4
3-f.	Distinguish between Boosting and Ada boosting? (CO4)	4
3-g.	Discuss about Spatial Data Mining? (CO5)	4

SECTION C 35

4. Answer any one of the following:-

4-a.	Explain Normalization in detail and What are the different types of Normalization? (CO1)	7
4-b.	Explain how OLAP is used for analytical processing and also explain how it is related to Business Intelligence? (CO1)	7

5. Answer any one of the following:-

5-a.	Explain data integration, transformation and loading? (CO2)	7
5-b.	What is Apriori Algorithm explain in detail and also write the steps involved using a sample transaction data? (CO2)	7

6. Answer any one of the following:-

6-a.	Explain the concept of classification in brief and also explain CURE and Chameleon clustering methods? (CO3)	7
6-b.	What is decision tree classification, explain its functionality and also write the steps involve in it? (CO3)	7

7. Answer any one of the following:-

7-a.	What are the Hierarchical methods of clustering and how they are classified to bottom-up and top-down approach? (CO4)	7
7-b.	What are Grid-Based methods. Explain STING in detail? (CO4)	7

8. Answer any one of the following:-

8-a.	How will you do pattern mining in multilevel and multidimensional space explain in detail? (CO5)	7
8-b.	Explain about the three-tier Data Warehouse architecture with neat diagram and what is a metadata repository? (CO5)	7