No. of Printed Pages : 6 **MCS–221**

MASTER OF COMPUTER APPLICATIONS

(MCA–NEW)

Term-End Examination

June, 2025

MCS-221 : DATA WAREHOUSING AND DATA MINING

Time : 3 Hours *Maximum Marks : 100*

Weightage : 70%

Note **:** *(i) Question No.* ***1*** *is compulsory.*

(ii) Answer any ***three*** *questions from the rest.*

1. (a) "Resampling method is a common approach that can be used for model selection in supervised learning." With

reference to resampling method, explain the following approaches : 10

(i) K-fold cross validation

(ii) Leave-one-out method

(iii) Random split

(iv) Time-based split

(b) What is frequent pattern mining ? Briefly explain the following classifications of frequent pattern mining along with an example for each : 10

(i) Based on the levels of abstraction involved in the rule-set.

(ii) Based on the types of values handled in the rule.

(iii) Based on the kinds of rules to be mined.

(c) What is Online Analytical Processing (OLAP) ? Explain briefly the following OLAP data cube operations with the help of an example for each : 10

(i) Roll-up

(ii) Drill-down

(iii) Slice and Dice

(d) What is a star-schema dimensional modeling ? What are its characteristics ? Draw a star schema for four dimensions–Time, Item, Branch, Location with 2 measures namely 'Units-sold' and 'Amount-sold' and 'sales' is the Fact Table. Also, explain the star-schema diagram. Assumptions can be made wherever necessary. 10

2. (a) With the help of a suitable diagram for each, explain Top-down Approach (proposed by Bill Inmon) and Bottom-up Approach (proposed by Kimball) of Data Warehouse design. 10

   (b) What are Datamarts ? How are they different from a Data Warehouse ? Write and explain all the steps involved in a datamart design. 10

3. (a) Write and explain DBSCAN algorithm a density based method in clustering. Also, mention any *two* advantages and disadvantages. 10

(b) What are outliers in data mining ? How are they handled ? Explain briefly any *two* outlier detection techniques. 10

4. (a) What is a data lake ? Draw a diagram and explain all the four layers in data lake architecture. 10

(b) With reference to HADOOP architecture, explain the following : 10

(i) Name Node

(ii) Data Nodes

(iii) File System Namespace

(iv) Data Replication

5. Write short notes on any ***four*** of the following : 4×5=20

(a) Data Warehouse key challenges

(b) Cloud Data Warehousing

(c) ELT and its benefits

(d) Data Mining Life Cycle

(e) Data Integration issues in Data Preprocessing

× × × × ×