

No. of Printed Pages : 5

MCS-221

**MASTER OF COMPUTER
APPLICATIONS (MCA-NEW)**

Term-End Examination

June, 2026

**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) With the help of a block diagram,
explain the three-tier data warehouse

architecture. Also mention the best practices for designing the data warehouse architecture. 10

(b) What are Aggregate Tables in dimensional modelling ? What is the need for building aggregate fact tables and derived dimension tables ? Explain with an example for each. 10

(c) What is data-cleaning process ? How are the following treated during the data cleaning process : 10

(i) Missing values

(ii) Noisy data

- (d) Write and explain Naive-Bayesian classification. Also, list the advantages and disadvantages of Naive Bayes classifier. 10
2. (a) Write and explain Rule-based classification with the help of an example. 10
- (b) What are Receiver Operating Curves (ROC) ? How are ROC's constructed/ plotted ? Explain various regions of an ROC graph. Distinguish between a perfect classifier, reasonable classifier and a poor classifier by looking into the ROC graphs. 10

3. (a) Write and explain k-medoids or PAM (Partitioning Around Medoids) algorithm of partitioning method of clustering. 10
- (b) What is web-mining ? List its essential features. What kind of tasks can be performed using web-mining ? Also mention its types. 10
4. (a) List and explain the applications of the following : 10
- (i) Data warehousing
- (ii) Data mining
- (b) Discuss the following latest trends in data warehousing : 3+3+4
- (i) Data Lakes
- (ii) Data Swamp
- (iii) Cloud Data Warehouses

5. Write short notes on the following : 4×5=20

- (a) Data Warehouse Automation
- (b) Multidimensional Operations (Roll-up, Drill-down Slice and Dice, Pivot)
- (c) Decision tree approach
- (d) DBSCAN cluster method

× × × × ×