No. of Printed Pages : 5

MCS–221

MASTER OF COMPUTER APPLICATIONS

(MCA–NEW)

Term-End Examination

December, 2024

MCS-221 : DATA WAREHOUSING AND DATA MINING

Time : 3 Hours *Maximum Marks : 100*

Weightage : 70%

Note : *Question No. 1 is compulsory. Answer any* ***three*** *questions from the rest.*

---

1. (a) Draw snowflake schema diagram for hotel occupancy, where there are four dimensions namely hotel details, room

details, time and customer with two measures namely occupied rooms and vacant rooms. Also, explain the snowflake schema diagram.

Assumptions can be made wherever necessary. 10

(b) What is ETL in Data Warehousing ? Why do you need ETL ? Describe the layered implementation of ETL in a data warehouse. 10

(c) What is Noisy Data ? How binning method solves the problem of noisy data ? Illustrate with an example. 10

(d) Write and explain K-nearest neighbour's algorithm. Write its advantages and disadvantages. 10

2. (a) What is text mining ? With reference to text mining, explain the following techniques :

   (i) Information Extraction

   (ii) Text Summarization

   (iii) Text Categorization

   (iv) Text Clustering

   Also, mention any ***four*** applications of text mining. 10

   (b) Explain the following clustering methods briefly : 10

   (i) Partitioning method

   (ii) Density-based method

3. (a) With reference to data warehousing, explain the following terms : 10

   (i) Metadata and Data warehousing

(ii) Data Granularity

(iii) Operational data store

(iv) Data Mart

(b) Briefly explain the single-tier and three-tier data warehouse architectures with the help of a suitable diagram for each. 10

4. (a) What is Data Mining ? Briefly explain the following data mining techniques with the help of an example for each : 10

(i) Association Rule Mining

(ii) Outlier Detection

(iii) Regression Analysis

(b) Write and explain Apriori algorithm to identify the most frequently occurring elements and meaningful associations in a dataset. 10

5. Write short notes on any ***four*** of the following :

4×5=20

(a) Mining multilevel association rules

(b) Data reduction (with reference to data pre-processing)

(c) Mining multimedia data on the web

(d) K-Medoids algorithm of clustering

(e) Rule-based classification

× × × × × × ×