MCS-221

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

Term-End Examination

December, 2022

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) Define Dimensional Modelling. With reference to this, define the terms Facts, Fact Table, Dimensions and Dimensional Table. Give an example use-case and derive fact table and dimension tables. *10*

(b) With the help of a Data Warehouse Architecture diagram, explain the following components and their significance : *10*

(i) ETL

(ii) Metadata

(iii) Data Warehouse Access Tools

(iv) Data Warehouse Reporting Layer

(c) Define a Decision Tree. With the help of an example, explain the construction and representation of decision tree. Also, mention its strengths and weaknesses. *10*

(d) Discuss the following categories of Data Mining Issues : *10*

(i) Mining Methodology and User Iteration Issues

(ii) Performance-based Issues

(iii) Diverse Data Types Issues

2. (a) Define Data Cleaning. Explain the ways and means of handling the Missing Values and Noisy Data while data preprocessing. *10*

(b) Write and explain the K-means algorithm for clustering. How does it work ? *10*

3. (a) Why does dimensionality reduction of text need to be done ? Explain Tokenization process and Vector from text approach, with the help of a suitable example for each. *10*

(b) Enumerate the best practices for Data Warehouse Architecture. *5*

(c) Describe the following types of data marts : *5*

(i) Dependent data marts

(ii) Independent data marts

4. (a) With the help of an example use-case, explain the Snowflake schema. List its advantages and disadvantages. *10*

(b) Define Web Mining. What kind of tasks can be performed using it ? Discuss its features and also a few applications. *10*

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

(a) Data Transformation (with reference to data preprocessing)

(b) Cloud Data Warehousing

(c) Data Lake and its Architecture

(d) Data Warehouse Automation