7. Write short note on :

(a) Back end tools and utilities.

(b) Indexing OLAP data.

(c) Data Mining Query Language. (5×3=15)

Roll No. ..... Total Pages : 3

602304

December 2022

MCA-III SEMESTER

Data Warehousing and Data Mining (MCA-20-207-4)

Time : 3 Hours] [Max. Marks : 75

Instructions :

1. *It is compulsory to answer all the questions (1.5 marks each) of Part-A in short*
2. *Answer any four questions from Part-B in detail.*
3. *Different sub-parts of a question are to be attempted adjacent to each other.*

PART-A

1. (a) How is a data warehouse different from a database? (1.5)

(b) Suppose that the minimum and maximum values for the attribute income are $12,000 and $98,000, respectively. We would like to map income to the range [0.0, 1.0]. Transform a value of $73,600 for income, by min-max normalization.

(c) Differentiate between ROLAP and HOLAP. (1.5)

(d) Discuss an example for Multilevel Association rule. (1.5)

(e) Define frequent itemset, support and confidence. (1.5)

(f) How would you measure the quality of clusters? (1.5)

(g) Write a short note on market basket analysis. (1.5)

(h) Define classifier accuracy. (1.5)

(i) What are the major issues in Data Mining? Explain briefly. (1.5)

(j) Describe Confusion Matrix. (1.5)

PART-B

2. (a) Define each of the following data mining functionalities: characterization, association, classification, and clustering and outlier analysis. Give examples of each data mining functionality, using a real-life database that you are familiar with. (10)

(b) In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem. (5)

3. A data warehouse can be modelled by either a star schema or a snowflake schema. Briefly describe the similarities and the differences of the two models using suitable examples, and then analyze their advantages and disadvantages with regard to one another. (15)

4. (a) Explain data mining as a step in knowledge discovery process. (10)

(b) With illustrative examples explain various OLAP operations. (5)

5. What is the main objectives of clustering? Give the cateogrization of clustering approaches. Brielfy discuss them. (15)

6. (15)

TID	List of Item IDS
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Table Shows Transactional data for All Electronics Branch.

(a) Find all frequent itemsets using Apriori algorithm.

(b) List all the strong association rules (with supports and confidence.

(c) matching the following metarule, where X is a variable representing customers, and item i denotes variables representing items (e.g., "A," "B,"): x ε transaction, buys(X,item 1) buys(X,item 2) buys(X,item 3) [s,c].