

Seat No. : _____

MA-121

May-2022

M.Sc. (CA & IT), Sem.-VIII

DM & DA Theory

(Data Mining and Data Analytics)

Time : 2 Hours]

[Max. Marks : 50

SECTION – I

Attempt any 3 questions From Q.1 to Q.5 :

1. (A) Explain with example: Star Schema, Snowflake Schema and Fact Constellations. 7
(B) Explain Distributive, Algebraic and Holistic Measures. 7
2. (A) Explain unsupervised learning approach. What is Cluster analysis ? Write and explain the requirements for cluster analysis. 7
(B) Explain any three functions from OLAP operations with example. 7
3. (A) Explain Principal Component Analysis with diagram. 7
(B) Explain the method Stepwise forward selection and stepwise backward selection in Attribute Subset Selection. 7
4. Apply Apriori Algorithm on the following data set and find the final association where minimum support = 50% and Threshold Confidence = 70%. 14

DataSet :

Transaction ID	List of Items
T100	1,2,5
T200	2,4
T300	2,3
T400	1,2,4
T500	1,3
T600	1,2
T700	1,3
T800	1,2,3,5
T900	1,2,3

5. Define the following : (Any 7)

14

- (1) Data Mart
- (2) ETL Process
- (3) Metadata Repository
- (4) Data Cube
- (5) Star Schema
- (6) Enterprise Warehouse
- (7) Facts
- (8) Base Cuboid

SECTION – II

MCQ [Attempt any 8 from this section]

8

- (1) Each dimension in a data cube has a table associated with it. This table is called _____.
(A) Dimension (B) Dimension Table
(C) Facts (D) Metadata
- (2) _____ gathers data from multiple heterogeneous and external sources.
(A) Data cleaning (B) Data Extraction
(C) Transformation (D) Loading
- (3) 0-D cuboid which holds the highest level of summarization is called _____.
(A) Apex Cuboid (B) 2-D Cube
(C) 3-D Cube (D) Base cuboid
- (4) Which of the following is not a type of materialization ?
(A) No materialization (B) Full materialization
(C) Partial materialization (D) Half materialization
- (5) _____ OLAP supports multi-dimensional data views through array-based storage engines.
(A) ROLAP (B) MOLAP
(C) HOLAP (D) None
- (6) In the data cleaning method, Smoothing by Bin means, the value of bin is replaced by _____.
(A) Mode (B) Mean
(C) Median (D) Boundary values
- (7) Which of the following is not caused by discrepancy detection ?
(A) Deliberate errors (B) Human errors
(C) System errors (D) Clean data
- (8) Which of the following is a discrepancy detection tool ?
(A) Data cleaning (B) Data integration
(C) Data scrubbing tools (D) Transformation
- (9) What does a null hypothesis mean ?
(A) Both attributes are independent
(B) Both attributes as dependent
(C) If one attribute increases other decreases
(D) If one attribute decreases other also decreases
- (10) In correlation coefficient $r(a, b) = +1$ means _____.
(A) Perfect +ve correlation (B) No correlation
(C) Perfect – ve correlation (D) None