

Seat No. : _____

JB-101

July-2021

BCA., Sem.-VI

**CC-308 : Data Mining and Data Warehouse
(New Course)**

Time : 2 Hours]

[Max. Marks : 50

- Instructions :**
- (1) All Questions of **Section – I** carry equal marks.
 - (2) Attempt any **two** Questions in **Section – I**.
 - (3) Question – **5** in **Section – II** is **COMPULSORY**, Attempt any **Five**.

SECTION – I

1. (A) What is Data Mining ? List the KDD steps. **10**
(B) Which types of technologies are used in mining of data ? **10**
2. (A) (1) Explain ETL (*Extraction, Transformation and Loading*) of data warehouse. **10**
(2) Explain data warehouse definition by W. H. Inmon. **10**
(B) Explain OLAP operations in brief. **10**
3. (A) Explain Apriori Algorithm with proper example. **10**
(B) Explain any two Data Reduction Strategies. **10**
4. (A) Define decision tree induction and its pruning with proper example. **10**
(B) Define Selection Measures (*Information gain* and *Gain ratio*) with proper example (training data) **10**

SECTION – II

5. Attempt any **Five** :

10

- (1) Data warehouse is the subset of the data mining (True/False)
 - (2) Classification and regression is based on :
 - (a) Training data
 - (b) Test data
 - (c) New data
 - (d) Meta-data
 - (3) Clustering is also called classification of data set (True/False)
 - (4) OLAP stand for _____.
 - (5) Data are arranged in subject or department wise, is called
 - (a) Data Cube
 - (b) Meta-data
 - (c) Data Mart
 - (d) Data repository
 - (6) Data cube is the part of the multi-dimensional data (True/False)
 - (7) Data cleaning is process of
 - (a) Remove noisy data
 - (b) Delete data
 - (c) Update data
 - (d) Sharp data
 - (8) Mined data result may be based on
 - (a) tabular
 - (b) graphical
 - (c) Diagrammatical
 - (d) all of these
 - (9) Decision tree are normalised by
 - (a) Pruning
 - (b) indexing
 - (c) classification
 - (d) training data
 - (10) Clustering is called supervised learning (true/false)
-