Seat No. : _____

AB-104

April-2023

BCA, Sem.-VI

CC-308 : Introduction to Data Mining & Data Warehouse

Time : 2:30 Hours] **[Max. Marks : 70**

Instructions : (i) **All** questions in Section – **I** carry equal marks.
(ii) Question – **5** in Section – **II** is compulsory.

Section – I

1. (A) Which are the major challenges that stimulates further improvement in Data Mining ? **7**
   (B) Explain different Data mining technologies that are incorporated in various domains. **7**

OR

   (A) Explain how Frequent Patterns leads to Associations and Correlations within data. **7**
   (B) What is Data ? Using examples, explain its various types. **7**

2. (A) Which are the major features of a Data warehouse as defined by William H. Inmon ? **7**
   (B) Explain typical operations carried out in OLAP as multidimensional data model. **7**

OR

   (A) Draw a proper diagram of a three tier Data warehousing architecture and explain. **7**
   (B) "Data cubes are n-dimensional." Explain it with 2-D, 3-D and 4-D data cubes. **7**

3. (A) Which are the methods used to fill-in missing values for attributes in data cleaning ? **7**
   (B) Using an example, explain the steps of Apriori algorithm for mining frequent item-sets. **7**

OR

   (A) List out and explain Binning methods applied on sorted data values to remove noise. **7**
   (B) Explain in detail how singleton buckets / range buckets are used in Histogram technique. **7**

4. (A) Draw a decision tree and explain with an example how it is used for classification. **7**
   (B) Explain the use of Data mining applications in Financial Data Analysis area. **7**

OR

   (A) Explain a centroid based set-partitioning technique called K-Means in detail. **7**
   (B) "Science and Engineering Area make the most use of data mining techniques" – Explain. **7**

Section – II

5. Answer the following MCQ's. : (Any **Seven**) **14**

(i) _____ process is helpful in retrieving data from database for analysis.

(A) Data cleaning (B) Data integration
(C) Data selection (D) Data transformation

(ii) _____ labels for the group of data is the result of Clustering.

(A) Class (B) Object
(C) Instance (D) Dataset

(iii) _____ can be sub-categorized as Active learning and supervised learning.

(A) Passive learning (B) Machine learning
(C) Unsupervised Learning (D) Semi-supervised Learning

(iv) Data cube has _____ table, associated to each dimension of it.

(A) Dimension table (B) Fact table
(C) Association table (D) Cuboid table

(v) _____ type of Data warehouse provides Information from a historical perspective.

(A) Non-volatile (B) Volatile
(C) Time-variant (D) History-variant

(vi) In Data cube presentation, _____ gives highest level of summarization.

(A) Base cuboid (B) Side cuboid
(C) Top cuboid (D) Apex cuboid

(vii) Data quality is concerned with ____.

(A) Accuracy (B) Completeness
(C) Timeliness (D) All of the above

(viii) _____ rule of association technique reflects certainty of the rule.

(A) Confidence (B) Support
(C) Minimum-confidence (D) Minimum-support

(ix) To remove the non-frequent subset of items from item list, _____ operation is performed.

(A) Prune (B) Delete
(C) Remove (D) Outcast

(x) Which of the following predict categorical class labels ?

(A) Test sets (B) Classifiers
(C) Tuple sets (D) Overfit sets

(xi) _____ step of classification process generates new data tuples as an output.

(A) Learning (B) Supervised Learning
(C) Classification (D) Clustering

(xii) Loan payment prediction can be obtained through data mining of _____.

(A) Science and Engineering (B) Retail and Telecommunication
(C) Intrusion Detection (D) Financial Data Analysis