

AH-106

April-2022

B.C.A., Sem.-VI

CC- 308 : Introduction to Data Mining and Data Warehousing**[Max. Marks : 50]****Time : 2 Hours]**

- Instructions :** (1) All questions in **Section – I** carry equal marks.
 (2) Attempt any **two** questions in **Section – I**.
 (3) Question **V** in **Section – II** is compulsory.

SECTION – I

- I. (A) What is Data Mining ? Explain the technologies used in data mining. **10**
 (B) Discuss major issues of data mining in detail. **10**
- II. (A) What is Data Warehouse ? Explain its key features in detail. **10**
 (B) Explain OLAP operations with appropriate example. **10**
- III. (A) What is data cleaning ? Discuss the methods of data cleaning. **10**
 (B) What is data reduction ? Explain Histogram and Sampling data reduction strategies in detail. **10**
- IV. (A) What is cluster analysis ? Explain various clustering methods in detail. **10**
 (B) Explain any two data mining applications in detail. **10**

SECTION – II

- V. MCQs (any five) **10**
- (1) _____ is an essential process where intelligent methods are applied to extract data patterns.
- (a) Data warehousing (b) Data mining
 (c) Text mining (d) Data selection

- (2) _____ is an intermediate storage area used for data processing during the extract, transform and load process of data warehousing.
- (a) Buffer (b) Virtual memory
(c) Staging area (d) Inter-Storage area
- (3) Which one is not clustering method ?
- (a) Partitioning (b) Relational
(c) Hierarchical (d) Density based
- (4) What is the use of data cleaning ?
- (a) To remove the noisy data
(b) Correct the inconsistencies in data
(c) Transformations to correct the wrong data
(d) All of these
- (5) The full name of OLAP is _____.
(a) Online Analytical Processing (b) Online Advanced Processing
(c) Online Advanced Preparation (d) Online Analytical Performance
- (6) Which of the following is not a kind of data warehouse application ?
- (a) Information processing (b) Analytical processing
(c) Data mining (d) Transaction processing
- (7) _____ is a top-down splitting method based on a specified number of bins.
- (a) Normalization (b) Binning
(c) Clustering (d) Classification
- (8) _____ technique is used to detect relationships or associations between specific values of categorical variables in large data sets.
- (a) Data characterization (b) Data classification
(c) Association rule mining (d) Data discrimination
- (9) Decision tree is the most powerful for _____.
(a) classification (b) prediction
(c) both (a) and (b) (d) None of these
- (10) The learning which is used for inferring a model from labeled training data is called
- (a) Supervised learning (b) Predictive learning
(c) Hybrid learning (d) Unsupervised learning