

Seat No. : _____

AB-107

April-2025

BCA, Sem.-VI

CC-308 : Introduction to Data Mining and Data Warehousing

Time : 2:30 Hours]

[Max. Marks : 70

1. (A) What is data mining ? Explain the steps involved in knowledge discovery process. 7
- (B) Explain the different types of data that can be mined. 7

OR

1. (A) Discuss : Major Issues in Data Mining 7
- (B) Explain the role of statistics and importance of machine learning in data mining. 7
2. (A) Explain in detail data warehouse design process. 7
- (B) Explain the concept of data warehouse and discuss its key characteristics. 7

OR

2. (A) What is a data cube and how does it represent multidimensional data ? 7
- (B) Explain data warehousing as a multi-tiered architecture. 7
3. (A) What is meant by Noisy data ? Explain the data smoothing techniques. 7
- (B) What is data reduction and why is it essential in data mining ? Explain data reduction techniques : histogram analysis and data cube aggregation. 7

OR

3. (A) Explain the significance of data preprocessing in data mining. 7
- (B) Discuss : Apriori algorithm 7
4. (A) Discuss : Decision Tree Induction 7
- (B) What is cluster analysis ? Discuss different types of clustering methods in data mining. 7

OR

4. (A) What is classification in data mining ? Explain the general approach to classification. 7
- (B) Explain application of data mining for recommender system. 7

5. Attempt any 7 :

14

- (1) Which of the following techniques is used for predictive analysis in data mining ?
 - (a) Clustering
 - (b) Outlier Analysis
 - (c) Classification and Regression
 - (d) Data cleaning
 - (2) What is information retrieval mainly used for in data mining ?
 - (a) Extracting useful information from unstructured data like text documents
 - (b) Predicting future trends using past data
 - (c) Storing large amounts of structured data
 - (d) Generating frequent itemsets
 - (3) Which of the following is an example of OLAP operations ?
 - (a) Drill-down
 - (b) Roll-up
 - (c) Pivot
 - (d) All of these
 - (4) What type of tables does the data warehouse view include ?
 - (a) Only dimension table
 - (b) Only fact table
 - (c) Both fact and dimension table
 - (d) Raw transactional data
 - (5) Which of the following is not a common technique used in data cleaning ?
 - (a) Filling in missing values
 - (b) Removing duplicate records
 - (c) Adding noise to data
 - (d) None of these
 - (6) Which reduction techniques replace the original data volume by alternative, smaller forms of data representation ?
 - (a) Numerosity reduction
 - (b) Dimensionality reduction
 - (c) Data compression
 - (d) All of these
 - (7) The k-means algorithm defines the _____ of a cluster as the mean value of the points within the cluster.
 - (a) member
 - (b) attribute
 - (c) centroid
 - (d) None of these
 - (8) Which of the following sectors heavily rely on data mining applications ?
 - (a) Retail
 - (b) Finance
 - (c) Computer Science
 - (d) All of these
 - (9) The main purpose of cluster analysis is to normalize data. [TRUE/FALSE]
 - (10) Association rule mining is used finding missing values in data set.
[TRUE/FALSE]
 - (11) The cuboid that holds lowest level of summarization is called the Base Cuboid.
[TRUE/FALSE]
 - (12) Sampling is used to select a subset of data that represents the entire dataset.
[TRUE/FAESE]
-