

Seat No. : 00049

3

AC-120

April-2024

BCA., Sem.-VI

CC-308 : Introduction to Data Mining and Data Warehousing

Time : 2½ Hours]

[Max. Marks : 70

1. (A) What is KDD ? Explain data mining as a step in knowledge discovery. 7
(B) What are the challenges of data mining regarding *data mining methodology* and *user interaction issue*. 7

OR

1. (A) What kind of data can be mined ? 7
(B) Explain the following data mining functionalities : 7
• classification and Regression for predictive analysis
• cluster analysis

2. (A) What is data warehouse ? Differentiate between OLAP and OLTP. 7
(B) Explain the different approaches and steps involved in data warehouse. 7

OR

2. (A) With necessary diagrams and examples of data cube explain various OLAP operations. 7
(B) Explain about the three-tier data warehouse architecture with a neat diagram. 7

3. (A) Explain the Apriori algorithm for finding frequent item sets with an example. 7
(B) What is noise ? What are the techniques used for removing noise from data ? 7

OR

3. (A) What is data reduction ? Also explain different data reduction strategies. 7
(B) "Data quality can be assessed in terms of accuracy, completeness and consistency." Give suitable arguments in support of this statement. Also propose two other factors affecting data quality. 7

- (A) Discuss : Data mining for intrusion detection and prevention. 7
(B) Explain the concept of Decision Tree Induction. 7

OR

- (A) What is classification ? What are the steps involved in preparing the data for classification ? 7
(B) Discuss different clustering methods. 7

5. Attempt any 7 :

14

- (1) Which of the listed below helps to infer a model from labelled data ?
(a) Hybrid Learning (b) Supervised learning
(c) Unsupervised learning (d) Reinforcement learning
- (2) Among which of the following can query the unstructured textual data ?
(a) Information Retrieval (b) Information Access
(c) Information Manipulation (d) Information update
- (3) Missing values can be handled by _____.
(a) Ignoring tuples (b) Filling missing value manually
(c) Filling with global constant (d) All of these
- (4) A _____ says that each value of the given attribute must be different from all other values for that attribute.
(a) Null rule (b) Unique rule
(c) Consecutive rule (d) None of these
- (5) The cube created at the highest abstraction level is called _____.
(a) Base cuboid (b) Apex cuboid
(c) Lattice of cuboid (d) None of these
- (6) Which of the following are required in construction of data warehouse ?
(a) Data cleaning (b) Data consolidation
(c) Data integration (d) All of these
- (7) In the K-means algorithm for partitioning, each cluster is represented by the _____ of objects in the cluster.
(a) mode (b) median
(c) mean (d) member
- (8) In which histogram, the width of each bucket range is uniform ?
(a) Equal-frequency (b) equal-width
(c) multidimensional (d) None of these
- (9) Give full form of SRSWR.
- (10) Recommender system provides personalization for customers of e-commerce. [TRUE / FALSE]
- (11) Outlier analysis can uncover fraudulent usage of credit cards. [TRUE / FALSE]
- (12) Decision trees cannot be used for generating classification. [TRUE / FALSE]