



AAH-003-003621 Seat No. _____

B. C. A. (Sem. VI) (CBCS) Examination

March / April - 2016

CS-32 : Data Warehousing & Data Mining

Faculty Code : 003

Subject Code : 003621

Time : $2\frac{1}{2}$ Hours]

[Total Marks : 70

1 Attempt the following MCQs. : 20

- (1) A data warehouse is like a container of all the data needed to carry out _____ operations.
(A) Business Intelligence
(B) Information Intelligence
(C) Artificial Intelligence
(D) Testing Intelligence
- (2) In 1962, Rosenblatt introduced the first machine learning model is called _____.
(A) TRANSFORMATION (B) DECISION
(C) PERCEPTION (D) REGRESSION
- (3) The data cannot be _____ in the data warehouse in real time.
(A) Shifted (B) Edited
(C) Inserted (D) Deleted
- (4) The _____ manage performs all the operation associated with the management of data in warehouse.
(A) data mart manager (B) query manager
(C) load manager (D) warehouse manager
- (5) DSS stand for _____.
(A) decision support system
(B) decide support system
(C) divisive system support
(D) division support system
- (6) Which of the following is a type of data mart structure?
(A) star join (B) snowflake
(C) (A) and (B) (D) none

- (7) _____ use a "Divide and Conquer" technique to split the problem search space into subsets.
- (A) Data Mart (B) Decision tree
(C) Histogram (D) FP –Tree
- (8) Snowflake structure is the _____ of the dimension table in STAR SCHEMA.
- (A) Generalization (B) Centralization
(C) Normalization (D) Decentralization
- (9) _____ Key is used as the primary key in the dimension table.
- (A) Primary (B) Reference
(C) Foreign (D) Surrogate
- (10) _____ is an iterative clustering algorithm in which items are moved among sets of clusters until the desired set is reached.
- (A) Divisive Clustering (B) K– Means Clustering
(C) Nearest Neighbor (D) Apriori
- (11) OLAP performs _____.
- (A) analysis (B) transaction
(C) schema (D) multidimensional
- (12) Rejection of the null hypothesis causes another hypothesis, called the _____ hypothesis.
- (A) additive (B) alternative
(C) collective (D) supportive
- (13) OLAP store _____.
- (A) current data (B) historical data
(C) both (A) and (B) (D) none
- (14) Which of the following part consist by Genetic Algorithm?
- (A) Starting set (B) Mutation algorithm
(C) Fitness Function (D) All of above
- (15) In vertical partitioning mode, HOLAP stores aggregation in _____ and detailed data in _____
- (A) ROLAP, MOLAP (B) OLAP, MOLAP
(C) MOLAP, ROLAP (D) ROLAP, OLAP
- (16) In association analysis the antecedent and consequent are set of items that are _____
- (A) joint (B) disjoint
(C) common (D) associated
- (17) The _____ algorithm is well known association rule algorithm.
- (A) Apriori (B) Sampling
(C) Pincer –Search (D) FP - Tree Growth

- (18) In KDD process, _____ is essential stage where intelligent methods are applied in order to extract data pattern.
- (A) Preprocessing (B) Data Cleaning
(C) Data mining (D) Evaluation
- (19) Observation is important element in _____
- (A) Association Rule
(B) Hypothesis Testing
(C) Clustering Techniques
(D) Machine Learning
- (20) MST stands for _____
- (A) Minimum Spanning Tree
(B) Model Spanning Tree
(C) Master Spanning Tree
(D) Missing Spanning Tree

- 2** (a) Attempt the following : (any **three**) **6**
- (1) Define Meta Data.
(2) What is data Extraction?
(3) List operations performed by Warehouse Manager.
(4) Define: ETL Process.
(5) What is Data Mining?
(6) Define any two from given :
(1) Support (2) Confidence (3) Lift
- (b) Attempt the following : (any **three**) **9**
- (1) Give difference between Operational System and Informational System.
(2) List types of OLAP and Explain any one of them.
(3) Give difference between Data Warehouse and Data Mart.
(4) Write a note on parallel ETL implementation process.
(5) Explain partitioning algorithm for association rule.
(6) Write chi-squared statistic and explain how it is useful in data mining technique.
- (c) Attempt the following : (any **three**) **10**
- (1) Explain Star Schema with example.
(2) Discuss Merits and Demerits of Three-Tiered Architecture Data Warehouse.
(3) Characteristics of DATA WAREHOUSE.
(4) Discuss the different task of Load Manager, Warehouse Manager and Query Manager.
(5) Explain that how decision trees are most popular method of predictive data mining technique

- 3 (a) Attempt the following : (any **three**) 6
- (1) Give full name of following :
 - (1) DBSCAN
 - (2) WEKA
 - (3) ARFF
 - (4) CURE
 - (2) What is Dendrogram ?
 - (3) Define Linkage and its types.
 - (4) Explain Diameter in terms of clustering techniques.
 - (5) What is MSE?
 - (6) Define surrogate key. Where it is used?
- (b) Attempt the following : (any **three**) 9
- (1) Write a Short note: Machine Learning Data Mining Technique.
 - (2) Explain Partition Algorithm.
 - (3) Describe usage and security in a DATA MART
 - (4) Explain architectural components of Data Warehouse.
 - (5) Describe data transformation in detail.
 - (6) Give difference: Supervised V/s Unsupervised Data mining techniques.
- (c) Attempt the following : (any **three**) 10
- (1) Explain K – Means Clustering Technique.
 - (2) In Online Shopping applications, How Data Mining is useful? Highlight the major points of case study.
 - (3) Short Note : KDD Process
 - (4) Explain different phases of data mining process.
 - (5) Attempt following :
 - (a) List the steps of Divisive Clustering.
 - (b) Calculate average distance of given matrix.
 - (c) Clusters it using divisive clustering algorithm.
 - (d) Create two clusters.

$$\begin{array}{c}
 a_1 \quad a_2 \quad a_3 \quad a_4 \\
 \begin{array}{l}
 a_1 \\
 a_2 \\
 a_3 \\
 a_4
 \end{array}
 \begin{bmatrix}
 - & 3 & 5 & 8 \\
 3 & - & 2 & 1 \\
 5 & 2 & - & 4 \\
 8 & 1 & 4 & -
 \end{bmatrix}
 \end{array}$$
