



DA-003-003621

Seat No. _____

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

April / May – 2015

CS-32 : Data Warehousing and Data Mining

Faculty Code : 003

Subject Code : 003621

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

[Total Marks : 70

1 Attempt following MCQs : 20

(1) Data Mart is _____ database.

- (A) Rational (B) Irrational
(C) Thematic (D) Logical

(2) A data warehouse is collection of which ?

- (A) Databases (B) Tables
(C) Records (D) Data marts

(3) Which is most well-known descriptive data mining method?

- (A) Cluster analysis
(B) Regression analysis
(C) Multidimensional scalling
(D) Frequency distribution

(4) In neural network, which layer performs analysis ?

- (A) Input layer (B) Output layer
(C) Hidden layer (D) Internal layer

- (5) Kohonen networks can be considered as a _____ method of cluster analysis.
- (A) Nearest-neighbourhood
 - (B) Uni-dimensional
 - (C) Bi-dimensional
 - (D) Non-hierarchical
- (6) To detect frauds, which data mining methods is most suitable?
- (A) Association rules
 - (B) Nearest-neighbourhood
 - (C) Kohonen network
 - (D) Euclidian distance
- (7) For market basket analysis which data mining method is most suitable ?
- (A) Association rules
 - (B) Nearest-neighbourhood
 - (C) Kohonen network
 - (D) Euclidian distance
- (8) Typically which kind of data is used for data mining ?
- (A) Statistical
 - (B) Observational
 - (C) Depended
 - (D) Independent
- (9) The data in the data warehouse is _____
- (A) Volatile
 - (B) Non-volatile
 - (C) Non-integrated
 - (D) Coupled

- (10) Historical data is available in _____
- (A) OLTP (B) OLAP
(C) Data mart (D) All
- (11) ETL software is the implementation of _____
- (A) Data processing (B) Business processing
(C) BI processing (D) Parallel processing
- (12) CART and CHAID are example of _____ algorithm.
- (A) Genetic (B) Euclidian distance
(C) Decision tree (D) Cluster
- (13) In which algorithm iteratively merges clusters until all item belong in one cluster ?
- (A) K-means (B) Pincer-Search
(C) Agglomerative (D) FP-Tree Growth
- (14) In which algorithm items are moved among sets of clusters until the desired set is reached ?
- (A) K-means (B) Pincer-Search
(C) Agglomerative (D) FP-Tree Growth
- (15) Clustering is _____ classification.
- (A) Supervised (B) Unsupervised
(C) Linear (D) Non-linear

- (16) WEKA is supported by which platform ?
- (A) .NET (B) Java
(C) RDBMS (D) CORBA
- (17) Which one deals with detecting and removing errors and inconsistencies from data in order to improve the quality of data ?
- (A) Data Parsing (B) Data Filtering
(C) Data Cleansing (D) Data Reduction
- (18) From the following, which one is a temporary location where data from source system is collected, cleaned, integrated and detailed data in normalized form exists ?
- (A) Presentation Area (B) Process Area
(C) Staging Area (D) Source System Area
- (19) Which one describes concepts or task-relevant data sets in concise, summarative, informative form and presents interesting properties of data ?
- (A) Data Characterization
(B) Data Summarization
(C) Descriptive Data Mining
(D) Comparison of Data
- (20) _____ locates and identifies individual data elements in the source file and then isolates these data elements in the target file.
- (A) Staging (B) Parsing
(C) Standardizing (D) Consolidating

- 2 (a) Attempt following questions : (any **three**) **6**
- (1) Define association rule.
 - (2) What is data cleansing ?
 - (3) Describe Data Mining process in brief.
 - (4) Explain : Point Estimation.
 - (5) What is pattern evaluation ?
 - (6) Define Bayes theorem.
- (b) Attempt following questions : (any **three**) **9**
- (1) What is Outlier Discovery ? Explain.
 - (2) Distinguish : Database v/s Data Warehouse.
 - (3) Distinguish : OLAP v/s OLTP.
 - (4) Write a note on Machine Learning.
 - (5) Write differences between fact data and dimension data.
 - (6) Explain ETL with suitable example.
- (c) Attempt following questions : (any **two**) **10**
- (1) Write and explain steps in Data Mining process.
 - (2) Explain single-dimensional association rule for large database in detail.
 - (3) List various Genetics algorithms. Explain any one.
 - (4) What is clustering ? List various clustering algorithms. Explain any one.
 - (5) State and discuss prime steps of Apriori algorithm.

- 3** (a) Attempt following questions : (any **three**) **6**
- (1) What are the uses of Statistics in Data Mining.
 - (2) List various popular Data Mining tools. Explain any one with its functionality.
 - (3) Differentiate : Operational system and informational system.
 - (4) Describe Spatial mining in brief.
 - (5) Write a brief note on correlation analysis.
 - (6) Explain : Data Mart.
- (b) Attempt following questions : (any **three**) **9**
- (1) Write a note on architectural components of Data Warehouse.
 - (2) List different types of OLAP systems. Explain any one.
 - (3) How Decision tree works ? Explain.
 - (4) Describe the advantages of open source software available for Machine Learning.
 - (5) Write a note on Meta data architecture.
 - (6) What is Grid-based clustering ? Explain.
- (c) Attempt following questions : (any **two**) **10**
- (1) With a case study write role of Data Mining in market basket analysis.
 - (2) In Insurance applications how Data Mining is useful ? Highlight the major points of case study.

- (3) Describe the case study in Education field.
 - (4) Write comparison and contradiction of various ETL tools.
 - (5) Describe strength and weakness of K-means method with suitable example.
-