Seat No.:

MARCH - 2022 (Summer session) Examination

Subject Code: 66423
(विध्यार्थ्यांनी हा विषय कोड OMR वर लिहावा / Student should fill this code on OMR sheet)
Subject Name: Bachelor of Computer Application_66423_66423 - Data Mining and Data Warehousing_26.07.2022_10.00 AM (Part - III, Sem - VI)

Date: 26-07-2022   Time: 10:00:00 to 11:00:00
QP Code: 8978QP
Total Marks : 50 Each Question 2 Marks, Total 25 Ques, Duration 1 Hr

---

1. Adaptive system management is-

a. It uses machine-learning techniques. Here program can learn from past experience and adapt themselves to new situations

b. Computational procedure that takes some value as input and produces some value as output.

c. cience of making machines performs tasks that would require intelligence when performed by humans

d. None of these

2. Bayesian classifiers is-

a. A class of learning algorithm that tries to find an optimum classification of a set of examples using the probabilistic theory.

b. Any mechanism employed by a learning system to constrain the search space of a hypothesis.

c. An approach to the design of learning algorithms that is inspired by the fact that when people encounter new situations, they often explain them by reference to familiar experiences, adapting the explanations to fit the new situation.

d. None of these

3. Cluster is-

a. Group of similar objects that differ significantly from other objects

b. Operations on a database to transform or simplify data in order to prepare it for a machine-learning algorithm

c. Symbolic representation of facts or ideas from which information can potentially be extracted

d. None of these

4. Data selection is-

a. The actual discovery phase of a knowledge discovery process

b. The stage of selecting the right data for a KDD process

c. A subject-oriented integrated time variant non-volatile collection of data in support of management

d. None of these

5. Classification task referred to-

a. A subdivision of a set of examples into a number of classes

b. A measure of the accuracy, of the classification of a concept that is given by a certain theory

c. The task of assigning a classification to a set of examples   d. None of these

6. An essential process used for applying intelligent methods to extract the data patterns is-
a. Data Mining   b. Data Analysis
c. Data Implementation   d. Data Computation

7. Classification and regression are the properties of...
a. Data manipulation   b. Data analysis
c. Data mining   d. None of these

8. Strategic value of data mining is ....
a. Cost sensitive   b. Technical sensitive
c. Time sensitive   d. Design sensitive

9. Among the followings which is not a characteristic of Data Warehouse?
a. Integrated   b. Volatile
c. Time -Varient   d. Subject-Oriented

10. Which is the the incorrect answer regarding OLAP?
a. OLAP is used for multidimensional analysis   b. OLAP tools do Not provide a DSS capability
c. OLAP for the warehouse provides analytical power   d. All of the above

11. Information Visualization techniques is/ are ....
a. Pie Chart   b. Line Chart
c. Flow Chart   d. All of the Above

12. Star schema is a ---------- in the middle connected to a number of dimension tables.
a. Without a object   b. Single object
c. Two objects   d. Multiple objects

13. A logical subset of a complete data warehouse is called as .....
a. Data Dictionary   b. Data mart
c. Snowflake schema   d. Both a and b

14. R functionality is divided into a number of _____
a. Packages   b. Functions
c. Domains   d. Classes

15. What are the benefits of data visualization?
a. Better analysis   b. Identifying patterns
c. Exploring business insights   d. All of the Above

16. What is the meaning of the following R function? x <- c(4, 5, 1, 2, 3, 3, 4, 4, 5, 6) x <-as.factor(x)

a. x becomes a factor

c. x does not exist

b. x is a factor

d. x is not a vector

17. Which of the following statements is incorrect about the hierarchal clustering?

a. The hierarchal type of clustering is also known as the HCA

b. The choice of an appropriate metric can influence the shape of the cluster

c. In general, the splits and merges both are determined in a greedy manner

d. All of the above

18. Multiple Regression means

a. Data are modeled using a straight line

b. Data are modeled using a curve line

c. Extension of linear regression involving only one predicator value

d. Extension of linear regression involving more than one predicator value

19. R has how many atomic classes of objects?

a. 1

c. 3

b. 2

d. 5

20.   Which statement is true about the K-Means algorithm?

a.   All attribute values must be categorical.

b. The output attribute must be cateogrical.

c. Attribute values may be either categorical or numeric.

d. All attributes must be numeric.

21. Which of the following is finally produced by Hierarchical Clustering?

a. Final estimate of cluster centroids

b. Tree showing how close things are to each other

c. Assignment of each point to clusters

d. All of the mentioned

22. .... approach is best when we are interested in finding all possible interactions among a set of attributes.

a. Decision tree

c. K-Means algorithm

b. Association rules

d. Genetic learning

23. Which of the following is the collection of data objects that are similar to one another within the same group?
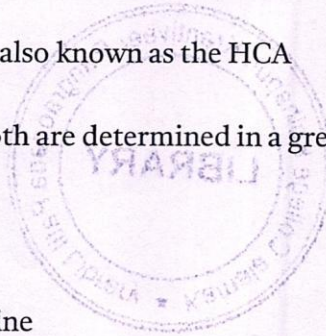
a. Partitioning

c. Cluster

b. Grid

d. Table

24. A Business Intelligence system requires data from:

a. Data warehouse

c. Database servers.

b. Operational systems

d. All of the above

25. Which of the process includes data cleaning, data integration, data selection, data transformation, data mining, pattern evolution and knowledge presentation?

a. KDD process

b. ETL process

c. KTL process

d. None of these