

1. Which of the following refers to the problem of finding abstracted patterns (or structures) in the unlabelled data?

A. Supervised learning

**B**

**B. Unsupervised learning**

C. Hybrid learning

D. Reinforcement learning

2) Which one of the following refers to querying the unstructured textual data?

A. Information access

**C**

B. Information update

C. Information retrieval

D. Information manipulation

3. Which of the following can be considered as the correct process of Data Mining?

A. Infrastructure, Exploration, Analysis, Interpretation, Exploitation

**A**

B. Exploration, Infrastructure, Analysis, Interpretation, Exploitation

C. Exploration, Infrastructure, Interpretation, Analysis, Exploitation

D. Exploration, Infrastructure, Analysis, Exploitation, Interpretation

4) Which of the following is an essential process in which the intelligent methods are applied to extract data patterns?

A. Warehousing

**B**

B. Data Mining

C. Text Mining

D. Data Selection

5) What is KDD in data mining?

A. Knowledge Discovery Database

**A**

B. Knowledge Discovery Data

C. Knowledge Data definition

D. Knowledge data house

6) The adaptive system management refers to:

A.Science of making machine performs the task that would require intelligence when performed by humans. **C**

B.A computational procedure that takes some values as input and produces some values as the output.

**C.It uses machine learning techniques, in which programs learn from their past experience and adapt themselves to new conditions or situations.**

D.All of the above.

7) For what purpose, the analysis tools pre-compute the summaries of the huge amount of data?

A.In order to maintain consistency

**D**

B.For authentication

C.For data access

**D.To obtain the queries response**

8) What are the functions of Data Mining?

A.Association and correctional analysis classification

**D**

B.Prediction and characterization

C.Cluster analysis and Evolution analysis

**D.All of the above**

9. In the following given diagram, which type of clustering is used?

**A.Hierarchal**

**A**

B.Naive Bayes

C.Partitional

D.None of the above

10) Which of the following statements is incorrect about the hierarchal clustering?

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

**A**

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

11) Which one of the following can be considered as the final output of the hierarchal type of clustering?

A.A tree which displays how the close thing are to each other **A**

B.Assignment of each point to clusters

C.Finalize estimation of cluster centroids

D.None of the above

12) Which one of the following statements about the K-means clustering is incorrect?

A.The goal of the k-means clustering is to partition (n) observation into (k) clusters **C**

B.K-means clustering can be defined as the method of quantization

**C.The nearest neighbor is the same as the K-means**

D.All of the above

13) Which of the following statements about hierarchal clustering is incorrect?

**A.The hierarchal clustering can primarily be used for the aim of exploration** **A**

B.The hierarchal clustering should not be primarily used for the aim of exploration

C.Both A and B

D.None of the above

14) Which one of the clustering technique needs the merging approach?

A.Partitioned **C**

B.Naïve Bayes

C.Hierarchical

D.Both A and C

15) The self-organizing maps can also be considered as the instance of \_\_\_\_\_ type of learning.

A.Supervised learning **B**

B.Unsupervised learning

C.Missing data imputation

D.Both A & C

16) The following given statement can be considered as the examples of\_\_\_\_\_

Suppose one wants to predict the number of newborns according to the size of storks' population by performing supervised learning

- A. Structural equation modelling **C**
- B. Clustering
- C. Regression
- D. Classification

17) In the example predicting the number of newborns, the final number of total newborns can be considered as the \_\_\_\_\_

- A. Features **D**
- B. Observation
- C. Attribute
- D. Outcome

18) Which of the following statement is true about the classification?

- A. It is a measure of accuracy **B**
- B. It is a subdivision of a set
- C. It is the task of assigning a classification
- D. None of the above

19) Which of the following statements is correct about data mining?

- A. It can be referred to as the procedure of mining knowledge from data **D**
- B. Data mining can be defined as the procedure of extracting information from a set of the data
- C. The procedure of data mining also involves several other processes like data cleaning, data transformation, and data integration
- D. All of the above**

20) In data mining, how many categories of functions are included?

A.5

C

B.4

C.2

D.3

21) Which of the following can be considered as the classification or mapping of a set or class with some predefined group or classes?

A.Data set

D

B.Data Characterization

C.Data Sub Structure

D.Data Discrimination

22) The analysis performed to uncover the interesting statistical correlation between associated -attributes value pairs are known as the \_\_\_\_\_.

A.Mining of association

B

B.Mining of correlation

C.Mining of clusters

d.All of the above

23) Which one of the following can be defined as the data object which does not comply with the general behavior (or the model of available data)?

A.Evaluation Analysis

B

**B.Outliner Analysis**

C.Classification

D.Prediction

24) Which one of the following statements is not correct about the data cleaning?

A.It refers to the process of data cleaning

D

B.It refers to the transformation of wrong data into correct data

C.It refers to correcting inconsistent data

**D.All of the above**

25) The classification of the data mining system involves:

A.Database technology

D

B.Information Science

C.Machine learning

**D.All of the above**

26) In order to integrate heterogeneous databases, how many types of approaches are there in the data warehousing?

- A.3
- B.4
- C.5
- D.2**

**D**

27) The issues like efficiency, scalability of data mining algorithms comes under\_\_\_\_\_

- A.Performance issues**
- B.Diverse data type issues
- C.Mining methodology and user interaction
- D.All of the above

**A**

28) Which of the following is the correct advantage of the Update-Driven Approach?

- A.This approach provides high performance.**
- B.The data can be copied, processed, integrated, annotated, summarized and restructured in the semantic data store in advance.
- C.Both A and B
- D.None of the above

**A**

**29.The term “data” came from the Latin root term-**

- A.Detem
- B.Datam
- C.Datum**
- D.Data

**C**

**30. Two types of research data include**

- A.Recognised and unrecognised data.
- B.Structured and unstructured data.
- C.Qualitative and quantitative data.**
- D.Organised and processed data.

**C**

**31. What is it called when the data is sourced from the place of origin?**

A.Secondary

B

**B.Primary**

C.Secondary and primary

D.All of the above

**32. What is it called when the data source is gathered and compiled with others?**

A.Primary Data

B

**B.Quantitative data**

C.Secondary data

D.None of the above

**33. Information of research is called –**

A.Qualitative

C

B.Quantitative

**C.Qualitative and Quantitative both**

D.None of the above

**34. Structured observation is a part of**

**A.Descriptive study**

A

B.Experimental study

C.Fundamental study

D.All of the above

**35. The primary data is gathered through**

A.Survey

C

B.Experiment

**C.Both options a and b**

D.None of the options a and b

**36. What are the different types of attributes?**

A.Nominal

☐ B. Ordinal

D

☒ C. Spacial

☐ D. All of the Above

**37. Examples of Nominal can be:**

- ☐ A. **ID Numbers, eye color, zip codes** **A**
- ☐ B. Rankings, taste of potato chips, grades, height
- ☐ C. Calendar dates, temperatures in celsius or Fahrenheit, phone numbers
- ☐ D. The temperature in Kelvin, length, time, counts

**38. Examples of Ordinal can be:**

- ☐ A. ID Numbers, eye color, zip codes
- B**
- ☐ B. **Rankings, taste of potato chips, grades, height**
- ☐ C. Calendar dates, temperatures in Celsius or Fahrenheit, phone numbers
- ☐ D. Temperature in Kelvin, length, time, counts

**39. The type of an Interval attribute depends on which of the following properties:**

- ☐ A. Distinctness & order **D**
- ☐ B. Distinctness, order & addition
- ☐ C. Distinctness
- ☐ D. **All 4 properties**

**40. Important Characteristics of Structured Data are:**

- ☐ A. Generality **D**
- ☐ B. Dimensionality
- ☐ C. Resolution
- ☐ D. All of the Above

**41. Which of these is correct about data mining?**

- a. It is a procedure in which knowledge is mined from data. **D**
- b. It involves processes like Data Transformation, Data Integration, Data Cleaning.
- c. It is a procedure using which one can extract information out of huge sets of data.
- d. **All of the above**

42. What is the analysis conducted for uncovering some interesting statistical correlations between various associated-attribute-value pairs called?

- a. Mining of Clusters
- b. Mining of Correlations**
- c. Mining of Association
- d. None of the above

**B**

43. In Data Warehousing, which of these is the correct advantage of the Update-Driven Approach?

- a. It provides high performance.
- b. It can be processed, copied, annotated, integrated, restructured and summarised in advance in the semantic data store.
- c. Both of the above**
- d. None of the above

**C**

44. Which of the following are methods under measures of dispersion?

- a. Standard deviation
- b. Mean deviation
- c. Range
- d. All of the above**

**D**

45. The coefficient of variation is a percentage expression for \_\_\_\_\_.

- a. Standard deviation**

**A**

b. Quartile deviation

c. Mean deviation

d. None of the above

46. Which of the following correctly refers the data selection?

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

**C**

B. The actual discovery phase of a knowledge discovery process

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

D. All of the above

47. Which one of the following correctly refers to the task of the classification?

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

**B. The task of assigning a classification to a set of examples**

**C**

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

D. None of the above

48. Which of the following correctly defines the term "Hybrid"?

A.Approach to the design of learning algorithms that is structured along the lines of the theory of evolution.

B.Decision support systems that contain an information base filled with the knowledge of an expert formulated in terms of if-then rules.

**C.Combining different types of method or information** C

D.None of these

49. Which of the following correctly defines the term "Discovery"?

A.It is hidden within a database and can only be recovered if one is given certain clues (an example IS encrypted information).

b.An extremely complex molecule that occurs in human chromosomes and that carries genetic information in the form of genes.

**c.It is a kind of process of executing implicit, previously unknown and potentially useful information from data**

D.None of the above C

50. Which sensor can detect nearby target?

a.Photoconductive sensor

b.Touch sensor

**c.Proximity sensor** C

D.Piezoelectric sensor

46.

---

techniques can be used  
to reduce the number of  
values for a given

continuous attribute, by dividing the range of the attribute into intervals.

a) **Discretization**

b) Transformation

c) Smoothing

d) Generalization



