1. _____ is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of

management decisions.

A. Data Mining.

B. Data Warehousing.

C. Web Mining.

D. Text Mining.

Answer: b

2. _____describes the data contained in the data warehouse.

A. Relational data.

B. Operational data.

C. Metadata.

D. Informational data.

Answer: c

3. _____predicts future trends & behaviors, allowing business managers to make proactive,

knowledge-driven decisions.

A. Data warehouse.

B. Data mining.

C. Datamarts.

D. Metadata.

Answer: B

4. OLAP stands for

a) Online analytical processing

b) Online analysis processing

c) Online transaction processing

d) Online aggregate processing

Answer: a

Explanation: OLAP is the manipulation of information to support decision making.

5. Data that can be modeled as dimension attributes and measure attributes are called _____ data.

a) Multidimensional

b) Singledimensional

c) Measured

d) Dimensional

Answer: a

Explanation: Given a relation used for data analysis, we can identify some of its attributes as measure attributes, since they measure some value, and can be aggregated upon.Dimension attribute define the dimensions on which measure attributes, and summaries of measure attributes, are viewed

6. The generalization of cross-tab which is represented visually is _____ which is also called as data cube.

a) Two dimensional cube

b) Multidimensional cube

c) N-dimensional cube

d) Cuboid

Answer: a

Explanation: Each cell in the cube is identified for the values for the three dimensional attributes.

7. The process of viewing the cross-tab (Single dimensional) with a fixed value of one attribute is

a) Slicing

b) Dicing

c) Pivoting

d) Both Slicing and Dicing

Answer: a

Explanation: The slice operation selects one particular dimension from a given cube and provides a new sub-cube. Dice selects two or more dimensions from a given cube and provides a new sub-cube.

8. The operation of moving from finer-granularity data to a coarser granularity (by means of aggregation) is called a _____

a) Rollup

b) Drill down

c) Dicing

d) Pivoting

Answer: a

Explanation: The opposite operation—that of moving fromcoarser-granularity data to finer-granularity data—is called a drill down.

9. In SQL the cross-tabs are created using

a) Slice

b) Dice

c) Pivot

d) All of the mentioned

Answer: a

Explanation: Pivot (sum(quantity) for color in ('dark','pastel','white')).

10. { (item name, color, clothes size), (item name, color), (item name, clothes size), (color, clothes size), (item name), (color), (clothes size), () }

This can be achieved by using which of the following ?

a) group by rollup

b) group by cubic

c) group by

d) none of the mentioned

Answer: d

Explanation: 'Group by cube' is used .

11. What do data warehouses support?

a) OLAP

b) OLTP

c) OLAP and OLTP

d) Operational databases

Answer: a

12. SELECT item name, color, clothes SIZE, SUM(quantity)

FROM sales

GROUP BY rollup(item name, color, clothes SIZE);

How many grouping is possible in this rollup?

a) 8

b) 4

c) 2

d) 1

Answer: b

Explanation: { (item name, color, clothes size), (item name, color), (item name), () }.

13. Which one of the following is the right syntax for DECODE?

a) DECODE (search, expression, result [, search, result]… [, default])

b) DECODE (expression, result [, search, result]… [, default], search)

c) DECODE (search, result [, search, result]… [, default], expression)

d) DECODE (expression, search, result [, search, result]… [, default])

Answer: d

14. Point out the correct statement.

a) The choice of an appropriate metric will influence the shape of the clusters

b) Hierarchical clustering is also called HCA

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

d) All of the mentioned

Answer: d

Explanation: Some elements may be close to one another according to one distance and farther away according to another.

15. 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

Answer: b

Explanation: Hierarchical clustering is an agglomerative approach.

16. Which of the following is required by K-means clustering?

a) defined distance metric

b) number of clusters

c) initial guess as to cluster centroids

d) all of the mentioned

Answer: d

Explanation: K-means clustering follows partitioning approach.

17. Point out the wrong statement.

a) k-means clustering is a method of vector quantization

b) k-means clustering aims to partition n observations into k clusters

c) k-nearest neighbor is same as k-means

d) none of the mentioned

Answer: c

Explanation: k-nearest neighbor has nothing to do with k-means.

18. Which of the following combination is incorrect?

a) Continuous – euclidean distance

b) Continuous – correlation similarity

c) Binary – manhattan distance

d) None of the mentioned

Answer: d

Explanation: You should choose a distance/similarity that makes sense for your problem.

19. Hierarchical clustering should be primarily used for exploration.

a) True

b) False

Answer: a

Explanation: Hierarchical clustering is deterministic.

20.Which of the following clustering requires merging approach?

a) Partitional

b) Hierarchical

c) Naive Bayes

d) None of the mentioned

Answer: b

Explanation: Hierarchical clustering requires a defined distance as well.

21. Predicting with trees evaluate _____ within each group of data.

a) equality

b) homogeneity

c) heterogeneity

d) all of the mentioned

Answer: b

Explanation: Predicting with trees is easy to interpret.

22. Point out the wrong statement.

a) Training and testing data must be processed in different way

b) Test transformation would mostly be imperfect

c) The first goal is statistical and second is data compression in PCA

d) All of the mentioned

Answer: a

Explanation: Training and testing data must be processed in same way.

23. Point out the correct statement.

a) Prediction with regression is easy to implement

b) Prediction with regression is easy to interpret

c) Prediction with regression performs well when linear model is correct

d) All of the mentioned

Answer: d

Explanation: Prediction with regression gives poor performance in non linear settings.

24. _____ is a system where operations like data extraction, transformation and loading operations are executed.

a) Data staging

b) Data integration

c) ETL

d) None of the mentioned

Answer: a

Explanation: In computing, Extract, Transform and Load (ETL) refers to a process in database usage and especially in data warehousing

25. Which of the following gave rise to need of graphs in data analysis?

a) Data visualization

b) Communicating results

c) Decision making

d) All of the mentioned

Answer: d

Explanation: A picture can tell better story than data.

26. Which of the following is characteristic of exploratory graph?

a) Made slowly

b) Axes are not cleaned up

c) Color is used for personal information

d) All of the mentioned

Answer: c

Explanation: A large number of exploratory graphs are made.

27. Point out the correct statement.

a) Data has only qualitative value

b) Data has only quantitative value

c) Data has both qualitative and quantitative value

d) None of the mentioned

Answer: a

Explanation: Data belongs to the set of items.

28. Data that summarize all observations in a category are called _____ data.

a) frequency

b) summarized

c) raw

d) none of the mentioned

Answer: b

Explanation: The summary could be the sum of the observations, the number of occurrences, their mean value, and so on.

29. Processing data includes subsetting, formatting and merging only.

a) True

b) False

Answer: b

Explanation: There are many other techniques applied to raw data.

30. Which of the following process involves structuring datasets to facilitate analysis?

a) Data tidying

b) Data mining

c) Data booting

d) All of the mentioned

Answer: a

Explanation: The principles of tidy data provide a standard way to organize data values within a dataset.

31. _____ is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of

management decisions.

A. Data Mining.

B. Data Warehousing.

C. Web Mining.

D. Text Mining.

Answer: b

32. Which of the following is a common error measure?

a) Sensitivity

b) Median absolute deviation

c) Specificity

d) All of the mentioned

Answer: d

Explanation: Sensitivity and specificity are statistical measures of the performance of a binary classification test, also known in statistics as classification function.

33. For k cross-validation, larger k value implies more bias.

a) True

b) False

Answer: b

Explanation: For k cross-validation, larger k value implies less bias.

34._____ is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of

management decisions.

A. Data Mining.

B. Data Warehousing.

C. Web Mining.

D. Text Mining.

Answer: b

35. The data Warehouse is_____.?

A. read only.

B. write only.

C. read write only.

D. none.

Answer: a

36. Expansion for DSS in DW is_____.

A. Decision Support system.

B. Decision Single System.

C. Data Storable System.

D. Data Support System.

Answer: a

37. The important aspect of the data warehouse environment is that data found within the data warehouse

is_____.

A. subject-oriented.

B. time-variant.

C. integrated.

D. All of the above.

Answer: d

38. The time horizon in Data warehouse is usually _____.

A. 1-2 years.

B. 3-4years.

C. 5-6 years.

D. 5-10 years.

Answer: d

39. The data is stored, retrieved & updated in _____.

A. OLAP.

B. OLTP.

C. SMTP.

D. FTP.

Answer: b

40. _____describes the data contained in the data warehouse.

A. Relational data.

B. Operational data.

C. Metadata.

D. Informational data.

Answer: c

41. _____predicts future trends & behaviors, allowing business managers to make proactive,

knowledge-driven decisions.

A. Data warehouse.

B. Data mining.

C. Datamarts.

D. Metadata.

Answer: b

42. _____ is the heart of the warehouse.

A. Data mining database servers.

B. Data warehouse database servers.

C. Data mart database servers.

D. Relational data base servers.

Answer: b

43. _____maps the core warehouse metadata to business concepts, familiar and useful to end users.

A. Application level metadata.

B. User level metadata.

C. Enduser level metadata.

D. Core level metadata.

Answer: a

44. _____consists of information in the enterprise that is not in classical form.

A. Mushy metadata.

B. Differential metadata.

C. Data warehouse.

D. Data mining.

Answer: a

45. The star schema is composed of _____ fact table.

A. one.

B. two.

C. three.

D. four.

Answer: a

46. The time horizon in operational environment is _____.

A. 30-60 days.

B. 60-90 days.

C. 90-120 days.

D. 120-150 days.

Answer: b

47. The key used in operational environment may not have an element of_____.

A. time.

B. cost.

C. frequency.

D. quality.

Answer: a

48. Data can be updated in _____environment.

A. data warehouse.

B. data mining.

C. operational.

D. informational.

Answer: c

49. The source of all data warehouse data is the_____.

A. operational environment.

B. informal environment.

C. formal environment.

D. technology environment.

Answer: a

50. Data warehouse contains_____data that is never found in the operational environment.

A. normalized.

B. informational.

C. summary.

D. denormalized.

Answer:c