Paper Code : 3855

M Sc (CA & I T) Sem – VII
JUL/SEP – 2021
705 : DATA MINING

કુલ સમય : ૬૦ મીનીટ **કુલ પ્રશ્નો : ૫૦** **વિદ્યાર્થીએ ટીક કરવાના પ્રશ્નો : ૩૫**

નોંધ : વિદ્યાર્થીએ માત્ર ૩૫ પ્રશ્નો જ ટીક કરવાના રહેશે. જો વધારે ટીક કરેલ હશે તો પ્રથમ ૩૫ પ્રશ્નો જ ગણતરીમાં લેવામાં આવશે.

---

1. Which of the following is not the attribute of data warehouse?
   A. Subject oriented B. Integrated C. Non-volatile D. Time-invariant
2. What is the full form of KDD?
   A. Knowledge Discovery from Data B. Knowledge Distraction from Data
   C. Know Disconnected Data D. Knowledge Discovery from Documents
3. A …….. is interesting if it is valid on test data with some degree of certainty, novel, potentially useful and easily understood by humans.
   A. Data B. Missing value C. Pattern D. Outlier
4. Which of the following is not the challenge in data mining research?
   A. Mining methodology B. User interaction C. Efficiency and scalability D. Information retrieval
5. Outliers are often discarded as…….
   A. Sound B. Noise C. Boundary D. Garbase
6. Collection of data is called ……...
   A. Data mining B. DBMS C. Database D. RDBMS
7. What is IQR?
   A. Interquartile ratio B. Interquartile range C. Innerquartile range D. Innerquadratic ratio
8. Which of the following is not a part of Five number summary?
   A. Minimum B. Maximum C. Boxplot D. Median
9. What is range?
   A. Max-min B. (max-min)/2 C. (max+min)/2 D. Max-min-1
10. Which of the following is not a type of binary attribute?
    A. Asymmetric B. Symmetric C. Similar D. None of the above
11. Which of the following distance measure is popular?
    A. Manhattan B. Minkowski C. Supremum D. Euclidean
12. Cosine similarity is used to measure similarity of ……..
    A. Documents B. Numbers C. Nominal D. Equations
13. Which are the data smoothing techniques?
    A. Binning B. Regression C. Outlier analysis D. All of above
14. Correlation analysis is used to identify redundancy in _____.
    A. Nominal data B. Ordinal data C. Numeric data D. All of above
15. ETL?
    A. Extraction/Transform/Loading B. Extraction/Transparent/Loading
    C. Entity/Tuple/Loading D. None of above
16. Which of the following is the operational processing?
    A. OLAP B. OLPP C. OLTP D. OOAP
17. How many OLAP operations are available?
    A. 3 B. 4 C. 5 D. 6
18. Climbing up concept hierarchy is called _____.
    A. Rollup B. Drilldown C. Pivot D. Slice
19. ……is a visualization operation that rotates the data axes in view to provide an alternative data representation.
    A. Roll up B. Drilldown C. Pivot D. Slice
20. If there were no hierarchy associated with any dimension, how many cuboids are there in 3-dimensional data cube?
    A. 3 B. 2^3 C. 3^2 D. 9
21. MOLAP?
    A. Multinational OLAP B. Multilingual OLAP C. Multidimensional OLAP D. Multiattribute OLAP
22. Support(A→B)=_____
    A. P(AUB) B. P(B/A) C. P(AUB)/2 D. P(AUB)/P(A)
23. "*All nonempty subsets of a frequent itemset must also be frequent.*" is the property of _____.
    A. Vertical data format (Eclat) Algorithm B. Apriori Algoritm C. Both A and B D. none of the above
24. Which of the following is not a pattern evaluation measure?
    A. Cosine B. Kulc C. All_conf D. Variance
25. ……..algorithm requires sorting of support count in initial phase.
    A. Eclat B. Apriori C. FP-Tree D. All of the above
26. …….. is the major issue in most popular Apriori algorithm.
    A. Frequent itemset B. Database scan on every iteration C. Minimum support D. All of the above
27. Which algorithm does not require candidate generation?
    A. Apriori B. Eclat C. FP-tree D. All of the above

28. Lift is a simple correlation measure that is given as follows.
    A. P(AUB)/P(A)*P(B) B. P(A)/P(AUB) C. P(AUB) D. P(B)/P(A)
29. Classification is a ……step process.
    A. One B. Two C. Three D. Four
30. In …….._learning, class labels are provided in advance.
    A. Basic B. Complex C. Unsupervised D. Supervised
31. A decision tree is a flow chart like structure where each internal node denotes …….and branch represents _____
    A. Test on attribute, outcome of the test B. Test on attribute, class label
    C. Class label, test on attributes D. All of the above
32. Splitting attribute is selected whose information gain is …….
    A. Lowest B. Equal C. Highest D. Negative
33. An attribute which has a large number of values also has highest ……..
    A. Gain ratio B. Gini index C. Voting D. Information gain
34. What is accuracy of classifier?
    A. FP+FN/(P+N) B. TP+TN/(P+N) C. TP/P D. TN/N
35. Which of the following is used to generate binary tree in decision tree induction?
    A. Information gain B. Gain ratio C. Gini index D. All of above
36. Cross validation method is used for which of the following?
    A. Accuracy B. Error rate C. Training set D. None of the above
37. Cluster analysis is …….
    A. Supervised learning B. Semi supervised learning C. Expert learning D. Unsupervised learning
38. …….._is used to calculate centroid of the cluster.
    A. Mean value of the points within cluster B. Mean value of the points of all clusters
    C. Mean values of the selected points within cluster D. Mean value of the selected points of all clusters
39. What is the drawback of the K-means algorithm?
    A. Missing value B. Noisy data/outliers C. Median D. Mode
40. …….._uses top-down strategy.
    A. Agglomerative hierarchical clustering method B. Divisive hierarchical clustering method
    C. K-means clustering method D. K-medoids clustering method
41. Sequence data consists of long sequences of numeric data, recorded at equal time interval is called ……..
    A. Web series data B. Text series data C. Time series data D. None of the above
42. Which of the following uses patterns from hyperlinks?
    A. Web content mining B. Web structure mining C. Web usage mining D. None of the above
43. Which is data mining application?
    A. Financial data analysis B. Retail and telecommunication industries
    C. Science and Engineering D. All of the above
44. ……..is used for mining audio signals
    A. Visual data mining B. Ubiquitous data mining C. Audio data mining D. Audio structure mining
45. Classification is ……
    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
46. Binary attribute are
    A. This takes only two values. In general, these values will be 0 and 1 and .they can be coded as one bit
    B. The natural environment of a certain species
    C. Systems that can be used without knowledge of internal operations D. None of these
47. Cluster is
    A. Symbolic representation of facts or ideas from which information can potentially be extracted
    B. Operations on a database to transform or simplify data in order to prepare it for a machine-learning algorithm
    C. Group of similar objects that differ significantly from other objects D. None of these
48. Euclidean distance measure is
    A. A stage of the KDD process in which new data is added to the existing selection.
    B. The process of finding a solution for a problem simply by enumerating all possible solutions according to some pre-defined order and then testing them
    C. The distance between two points as calculated using the Pythagoras theorem D. None of these
49. Heterogeneous databases referred to
    A. A set of databases from different b vendors, possibly using different database paradigms
    B. An approach to a problem that is not guaranteed to work but performs well in most cases.
    C. Information that is hidden in a database and that cannot be recovered by a simple SQL query.
    D. None of these
50. Operational database is
    A. A measure of the desired maximal complexity of data mining algorithms
    B. A database containing volatile data used for the daily operation of an organization
    C. Relational database management system D. None of these

---