MCQ
SUBJECT: DATA MINING AND WAREHOUSING
UNIT-I

1. _____ is as finding hidden information in a database.
    a) Data mining
    b) Database access
    c) DBMS
    d) Data warehouse.
2. KDD means _____ discovery in databases.
    a) King
    b) Kite
    c) Knowledge
    d) Kind
3. _____ model makes a prediction about values of data using known results found from different data.
    a) Descriptive
    b) Preference
    c) Predictive
    d) Algorithm
4. _____ maps data into predefined grouped or classes.
    a) Classification
    b) Regression
    c) Prediction
    d) Summarization
5. _____ model identifies patterns or relationships in data.
    a) Predictive
    b) Non-predictive
    c) Descriptive
    d) Unpredictable
6. _____ is he use of algorithm to extract the information and patterns derived by the KDD process.
    a) Data mining
    b) Data base
    c) Data access
    d) Data processing
7. _____ is he process of finding useful information and paterns in data.
    a) Data mining
    b) KDD
    c) Data warehouse
    d) Data processing
8. _____ is a type of classification where an input pattern is classified into one of several classes based on predefined classes
    a) Pattern recognition
    b) TSA
    c) Clustering

d) Prediction

9. _____is used to map data item into real valued prediction variable.
    a) Clustering
    b) Classification
    c) Regression
    d) TSA
10. _____ is used to visualize the time series.
    a) Time series plot
    b) Watch dog
    c) Time series analysis
    d) Grouping
11. Clustering is also called as _____.
    a) Grouping
    b) Segmentation
    c) Unsupervised learning
    d) All the above
12. Summarization is also called as_____.
    a) Characterization
    b) Generalization
    c) Simple description
    d) All the above
13. _____ maps data into sunsets with associated simple description .
    a) Summarization
    b) Association Rules
    c) Classification
    d) Clustering
14. _____ refers to the DM task of uncovering relationships among data.
    a) Link analysis
    b) Clustering
    c) TSA
    d) Summarization
15. _____ is a model that identifies specific types of data association.
    a) TSA
    b) Sequence discovery
    c) Clustering
    d) Association Rules
16. _____ is used to determine sequential patterns in data.
    a) TSA
    b) Sequence discovery
    c) Clustering
    d) Association rules

17. The definition of KDD includes the keyword _____.
    a) Useful
    b) This
    c) DM
    d) All the above

18. In transformation _____ is used to reduce the number of possible data values being considered.
    a) Data reduction
    b) Data interchange
    c) Errorneous of data
    d) Clearence of data
19. _____ techniques are used to make the data easier to mine and more useful and to provide meaningful results.
    a) Preprocessing
    b) Selection
    c) Transformation
    d) Interpretation
20. _____ refers to the visual representation of data.
    a) GUI
    b) Interpretation
    c) Visualization
    d) Hybrid
21. _____ techniques include the box plot and scatter diagram.
    a) Graphical
    b) Geometric
    c) Icon-based
    d) Pixel-based
22. _____ is used to proceed from specific knowledge to more general information.
    a) Compression
    b) Induction
    c) Hybrid
    d) Pruning
23. _____ occurs when the model does not fit future states.
    a) Overfitting
    b) Human interaction
    c) Outliers
    d) Integration
24. There are many data entries that do not fit nicely into derived model.
    a) Overfitting
    b) Human interaction
    c) Outliers
    d) Integration
25. IR stands for_____.
    a) Information reduction
    b) Information retrieval
    c) Information results
    d) Information relation
26. _____ is a software that is used to access the database.
    a) DBMS
    b) OLTP
    c) SQL
    d) CFMS

27. _____ data is said to be invalid or incorrect.
    a) Missing data
    b) Irrelevant data
    c) Noisy data
    d) Changing data
28. ROI stands for _____.
    a) Return on investment
    b) Return on instruction
    c) Return on information
    d) Return on invalid data
29. The use of other attributes that increase the complexity and decrease in algorithm is called _____.
    a) Dimensionality Curse
    b) Dimensionality reduction
    c) Dimensionality attribute
    d) Dimensionality
30. _____ techniques are targeted to such application as fraud detection, criminal suspects, prediction of terrorist.
    a) DM
    b) DB
    c) DBMS
    d) OLTP
31. _____ access a database using a well defined query stated in language such as SQL.
    a) DBMS
    b) DBS
    c) KDD
    d) Database queries
32. A database is partitioned into disjoint grouping of similar tuples called _____.
    a) Clustering
    b) Classification
    c) Segmentation
    d) Generlization
33. _____ finds occurrences of a predefined pattern in data.
    a) Patterning
    b) Pattern recognition
    c) Patterning of data
    d) Pattern analysis
34. In KDD, the input to the process is known as _____ and the Output is _____.
    a) Informtion , data
    b) Field,record
    c) Record,field
    d) Data ,information
35. In KDD,obtaining the data from various DB, files,and other sources is called _____.
    a) Preprocessing
    b) Selection
    c) Tranformation
    d) Evaluation

36. Link analysis is otherwise called as_____.
    a) Association
    b) Association rule
    c) Affinity analysis
    d) All the above
37. Prediction application include _____.
    a) Flooding
    b) Speech recognition
    c) Machine learning
    d) All the above
38. In regression, some type of error analysis is used to determine which function is____.
    a) Good
    b) Best
    c) Excellent
    d) Bad
39. Data mining is otherwise called as_____.
    a) Data analysis
    b) Data discovery
    c) Deductive learning
    d) All the above
40. The rise of DBMS tool is_____.
    a) 1960
    b) 1970
    c) 1980
    d) 1990
41. The metrics used include the traditional metrics of space and time based on _____.
    a) Complexity analysis
    b) Effectiveness
    c) Usefulness of data
    d) Scalability
42. _____ data are noisy and have many missing attributes values.
    a) Real world
    b) Abstract
    c) Assumption
    d) Authorized
43. The use of _____ data is found in GIS data base .
    a) Missing
    b) Irrelevant
    c) Noisy
    d) Multimedia

44. A large DB can be viewed as using_____ to help uncover hidden information about the data.
    a) Search
    b) Compression

c) Approximation
d) Querying

45. Interfaces between technical experts and domain comes under_____ issues.
    a) Overfitting
    b) Human interaction
    c) Outlier
    d) Application
46. The data Mining process can itself be vies a type of _____ underlying database.
    a) Querying
    b) Induction
    c) Search
    d) Processing
47. ___ requests may be treated as special,unusual or one time needs.
    a) KDD
    b) DM
    c) DBMS
    d) DB
48. _____ and_____ are effective tools to attack scalability problems.
    a) Dimensionality & Parallelization
    b) Sampling &Dimensionality
    c) Effectiveness &Sampling
    d) Sampling & Parallelization
49. Large data set is otherwise called as _____.
    a) Massive datasets
    b) High datasets
    c) Noisy datasets
    d) Irrelevent datasets
50. KDD process consists of _____ steps .
    a) One
    b) Three
    c) Four
    d) Five

UNIT-II

51. _____ models describe the relationship between I/O through algebraic equation.
    a) Parametric
    b) Non-parametric
    c) Static
    d) Dynamic
52. _____ may also be used to estimate error.
    a) Squared error
    b) Root mean error
    c) Mean Root square
    d) Mean squared error
53. _____ assumes that a linear relationship exists between the input data and the output data.
    a) Bivariate regression
    b) Correlation
    c) Multiple regression
    d) Linear regression
54. The _____ algorithm solves the estimation problem with incomplete data.
    a) Expectation maximization
    b) Expectation minimization
    c) Summarization-maximization
    d) Summarization minimization
55. Decision tree uses a _____ techniques.
    a) Greedy
    b) Divide & Conquer
    c) Shortest Path
    d) BFS
56. Null hypothesis and _____ hypothesis are two complementary hypothesis.
    a) Classical
    b) Testing
    c) Alternative
    d) None of the above
57. The BIAS of an estimator is the difference between _____ & _____values.
    a) Expected,actual
    b) Actual ,Expected
    c) Maximal,Minimal
    d) Minimal,Maximal
58. An _____ estimator is one whose BIAS is 0.
    a) Unbiased
    b) Rule biased
    c) Mean Root square
    d) Mean squared error

59. _____ is defined as the expected value of the squared difference between the estimate and the actual value.
    a) MSE
    b) RMS
    c) EM
    d) MLE
60. The_____ may also be used to estimate error or another statistic to describe a distribution.
    a) RMS
    b) MLE
    c) EM
    d) MSE
61. _____ is a technique to estimate the likelihood of a property given the set of data as evidence or input.
    a) Point Estimation
    b) Models based on summarization
    c) Bayes theorem
    d) Hypothesis testing

62. In Box plot the Total range of the data value is divided into _____.
    a) Regions
    b) Quartiles
    c) Divisions
    d) Partitions
63. _____ measure is used instead of similarity measures.
    a) Distance
    b) Dissimilarity
    c) Both a,b
    d) None of the above
64. _____ relates the overlap to the average size of the two sets together.
    a) Dice
    b) Jaccard
    c) Cosine
    d) Overlap

65. _____ is used to measure the overlap of two sets as related to the whole set caused by their union.
    a) Dice
    b) Jaccard
    c) Cosine
    d) Overlap
66. _____ coefficient relates the overlap to the geometric average of the two sets.
    a) Dice
    b) Jaccard
    c) Cosine
    d) Overlap

67. The_____ metrics determines the degree to which the two sets overlap.
    a) Dice
    b) Jaccard
    c) Cosine
    d) Overlap

68. _____ is a predictive modeling technique used in classification ,clustering,etc.
    a) Neural networks
    b) Decision tree
    c) Genetic algorithm
    d) All the above
69. The neural networks can be viewed as a directed graph with _____ nodes.
    a) Two
    b) Three
    c) Four
    d) One
70. Internal nodes are also called as _____.
    a) Input
    b) Output
    c) Hidden
    d) Sink
71. In neural networks _____ activation function produces a linear output value based on the input.
    a) Threshold
    b) Step
    c) Linear
    d) Sigmoid
72. _____ is a bell shaped curve with output values in the range [0,1].
    a) Linear
    b) Guassian
    c) Hyperbolic
    d) Sigmoid
73. In neural network , _____ is an S shaped curve with output values -1,1
    a) Sigmoid
    b) Linear
    c) Step
    d) Hyperbolic
74. The crossover technique generates new individual called_____.
    a) Offspring
    b) Children
    c) Both a, b
    d) None of the above
75. _____ is used to determine the best individuals in a population.
    a) Crossover
    b) Mutation
    c) Fitness function
    d) All the above

76. The_____ operation randomly changes character in the offspring.
    a) Crossover
    b) Mutation
    c) Fitness function
    d) Both a,b
77. _____ is defined by precise algorithms that indicate how to combine the given set of individual to produce new ones.
    a) Production
    b) Reproduction
    c) Genetic algorithms
    d) Crossover
78. The activation function is also called as_____.
    a) Processing element function
    b) Squashing function
    c) Firing rule
    d) All the above
79. The subsections of the chromosomes are called_____.
    a) Cross over
    b) Genes
    c) Alleles
    d) Offspring
80. _____ is used to estimate error or to describe a distribution.
    a) RMS
    b) MSE
    c) SE
    d) Jackknife
81. _____ can be defined as a value proportional to actual probability with specific distribution.
    a) Likelihood
    b) Maximum Likelihood
    c) Estimation
    d) None of the above
82. In hypothesis testing O represents _____.
    a) Outliers
    b) Observed data
    c) Output
    d) None of the above
83. One standard formula to measure linear correlation is the _____.
    a) Correlation coefficient
    b) Classification
    c) Clustering
    d) Dissimilarity measures
84. _____ are often used instead o similarity measures.
    a) Distance
    b) Dissimilarity measure
    c) Both a,b
    d) None of the above

85. A variation of sigmoid function is called_____.
    a) Gaussian
    b) Hyperbolic
    c) Linear
    d) Threshold
86. Gaussian function is a _____ shaped curve.
    a) S
    b) V
    c) Bell
    d) C
87. _____ is used to determine the best individuals in a population.
    a) Mutation
    b) Fitness function
    c) Crossover
    d) Starting set
88. One of the most important components of a genetic algorithm is_____.
    a) How to select individual
    b) How to select offspring
    c) How to select crossover
    d) How to select fitness
89. _____ coefficient is used to measure the overlap of two sets as related to whole set caused by their union.
    a) Dice
    b) Jaccard
    c) Cosine
    d) Overlap
90. _____ coefficient is used to relates the overlap to the average size of two sets together.
    a) Dice
    b) Jaccard
    c) Cosine
    d) Overlap
91. _____ coefficient relates the overlap to the geometric average of the two sets.
    a) Dice
    b) Jaccard
    c) Cosine
    d) Overlap
92. The_____ metric determines the degree to which the two sets overlap.
    a) Overlap
    b) Dice
    c) Cosine
    d) Jaccard
93. Rejection of null hypothesis causes another hypothesis called_____ hypothesis.
    a) Alternative
    b) Similarity measure
    c) Correlation
    d) Mutation

94. The input nodes exist in _____ layer.
    a) Output
    b) Input
    c) Hidden
    d) All the above
95. Internal nodes is called _____ nodes.
    a) Input
    b) Output
    c) Hidden
    d) All the above
96. Artificial NNs can be classified based on the type of_____.
    a) Connectivity
    b) Learning
    c) Both a, b
    d) None of the above
97. _____ occurs when the NNs is trained to fit one set to data.
    a) Outlier
    b) Noisy data
    c) Missing data
    d) Overfitting
98. To avoid overfitting _____ NNs are advisable.
    a) Larger
    b) Smaller
    c) Medium
    d) All the above
99. In sigmoid , c is a _____.
    a) Change
    b) Constant
    c) Crossover
    d) Children
100. ____ is defined as the excepted value of the squared difference between the estimate and the actual value.
    a) MSE
    b) RMSE
    c) BIAS
    d) RMS

UNIT-III

101. Estimation and prediction may be viewed as types of _____.
   a) Clustering
   b) Classification
   c) Regression
   d) Time Series

102. Classification performed by dividing the input space of potential database tuples into _____.
   a) Regions
   b) Class
   c) Space
   d) Sector

103. _____ values cause during both training and the classification process itself.
   a) Data
   b) Class
   c) Predicate
   d) Missing data

104. The performance of classification usually examined by evaluating the ___of the classification.
   a) Accuracy
   b) Contribution
   c) Special value
   d) Missing values

105. Classification true positives and false positives are calculated by the following curve.
   a)MOC
   b) NOC
   c)ROC
   d) COC

106. The _____ matrix illustrates the accuracy of the solution to a classification problem.
   a) Confusion
   b) Mutation
   c) Crossover
   d) Gaussian

107. _____ problems deal with estimation of an output value based on input values.
   a) Prediction
   b) Classification
   c) Clustering
   d) Regression

108. _____ is erroneous data.
   a) OC
   b) Regression
   c) Noise
   d) Linear model

109. Which are data values that are exceptions to the usual and expected data?
   a) Outliers
   b) Noise
   c) Regression
   d) Poor fit

110. The _____ classification can be viewed as both a descriptive and a predictive type of algorithm.
   a) Naive
   b) Bayes
   c) Naive bayes
   d) Prediction

111. The similarity (or) distance measures may be used to identify the _____ of different items in the database.
   a) Likeness
   b) Alikeness
   c) Outliers
   d) Centroid

112. A straightforward distance based approach assuming the each class Ci is represented by__.
   a) Centroid
   b) Outlier
   c) Medoid
   d) KNN

113. Expand : KNN
   a) K Normal Neighbors
   b) K Nearest Neighbor
   c) K Normal Nextvalue
   d) K Nearest Nest

114. The decision tree approach to classification is to divide search space into _____ regions.
   a) Square
   b) Triangular
   c) Circular
   d) Rectangular

115. In DT ,each internal node is labled with an _____.
    a) Class
    b) Attribute
    c) Arc
    d) Database
116. In DT, each leaf node labled with ____.
    a) Class
    b) Attribute
    c) Arc
    d) Link
117. The _____ technique to building a DT is based on information theory and attempts to minimize the expected number of comparisons.
    a) CART
    b) ID3
    c) C.4.5
    d) ROC
118. Neural networks are more robust than DTs because of the _____.
    a) Arcs
    b) Links
    c) Weights
    d) Classes
119. In NN, the normal approach used for processing is called_____.
    a) Activation function
    b) Interconnections
    c) Training data
    d) Propagation
120. The NN starting state is modified based on feedback of its performance is referred to as__.
    a) Supervised
    b) Unsupervised
    c) Both (a) and (b)
    d) None of these
121. _____ learning can also be performed if the output is not known.
    a) Supervised
    b) Unsupervised
    c) Neither (a) or (b)
    d) Oral

122. The Mean Squared Error (MSE) is found by _____.
    a) (yi-di)2/2
    b) (yi+di)2/2
    c) (di-yi)2/2
    d) (di+yi)2/2
123. The ____ can be used to find a total error over all nodes in the network.
    a) RDF
    b) ROC
    c) MSE
    d) CMC
124. Which learning technique that adjusts weights in the NN by propagating weight changes backward from the sink to the source nodes?
    a) Propagation
    b) perceptrons
    c) MSE
    d) Back propagation
125. In radial basis function (RBF) central point value is _____.
    a) 0
    b) 1
    c) +1
    d) -1
126. The simplest Neural Network is called a _____.
    a) Neuron
    b) Gene
    c) Perceptron
    d) Single neuron
127. In rule-based algorithms, _____rules that cover all cases.
    a) if-else
    b) if-then
    c) switch-case
    d) nested if
128. The _____is used to predict a future classification value.
    a) Genetic algorithm
    b) Decision Tree
    c) Rule-based Algorithm
    d) Neural Network

129. Multiple independent approaches can be applied to a classification problem is referred to as ___.

a) CMC
b) RBF
c) ROC
d) DCS

130. In which technique the classifier that has the best accuracy in database sample?

a) CMC
b) RBF
c) DCS
d) ROC

131. OC stands for_____.

a) Operating characteristics
b) Operating curve
c) Operating classifications
d) None of the above

132. Rule based classification algorithms generate _____ rules to perform the classifications.

a) If
b) Then
c) If-then
d) If - else

133. In OC curve , the horizontal axis has the percentage of _____Positives for a sample DB.

a) False
b) True
c) Either a, b
d) None of the above

134. In OC curve , the vertical axis has the percentage of _____Positives for a sample DB.

a) False
b) True
c) Either a, b
d) None of the above

135. The_____ approach is most useful in classification problem.

a) Incremental rule
b) Cluster
c) NN
d) Decision tree

136. _____ techniques use labeling of the items to assist In the classification process.

a) Intrinsic
b) Extrinsic
c) Overlapping
d) Numerical

137. A _____ curve shows the relationship between false positives and true positives.

a) BOC
b) ROS
c) ROC
d) BOS

138. Task of CART is_____.

a) Only regression
b) Only classification
c) Both a,b
d) None of the above

139. A variation of the complete link algorithm is called _____ algorithm.

a) Nearest
b) Neighbour
c) Farthest Neighbour
d) All the above

140. K nearest neighbor is a classification scheme based on the use of_____.

a) Distance Measure
b) Similarity
c) Complete link
d) Average

141. A perceptron is a _____ neuron with multiple inputs and one output.

a) single
b) Multiple
c) Double
d) None of the above

142. The classes that exist for a classification problem are indeed _____.

a) Equivalence classes
b) Variance classes
c) Mean classes
d) Median

143. The formula for straight line is_____.

a) Y=mx+b
b) y=mx
c) Y=M+b
d) Y=m

144. _____ are data values that are exception to the usual and expected data.

a) Outliers
b) Noise
c) Error
d) Overfit

145. _____ is an errorneous data.

a) Overfit
b) Outlier
c) Noise
d) Missing

146. _____ problems deal with the estimation of output value based on input value.

a) Baysian classification
b) K nearest Neighbour
c) Regression
d) All the above

147. ____ problem can be thought of as estimating the formula for a straight line.

a) Regression
b) Linear regression
c) Bayesian classification
d) K nearest neighbour

148. Logistic regression uses _____ technique.

a) Box plot
b) Logistic curve
c) Straight line
d) Logistic line

149. Decision tree is otherwise called as_____.

a) Classification tree
b) Regression tree
c) K nearest neighbor
d) Clustering tree

150. Data objects are described by a number of _____that capture the basic characteristics of an object.

a) Data sets
b) Elements
c) Record
d) Attribute

UNIT-IV

151. _____is similar to classification in that data are grouped.

a) Classification
b) Regression
c) Clustering
d) DT

152. One of the first domain in which clustering was used as _____taxonomy.

a) Biological
b) Zoological
c) Mathematical
d) Scientific

153. Cluster results are_____.

a) Static
b) Realistic
c) Acoustic
d) Dynamic

154. _____ clustering , the algorithm creates only one set of clusters.

a) Dynamic
b) Hierarchical
c) Partitional
d) Static

155. With _____ clustering, a nested set of clusters to be created.

a) Partitional
b) Hierarchical
c) Dynamic
d) Static

156. In similarity measures, metric attributes satisfy the _____ inequality.

a) Rectangular
b) Triangular
c) Square
d) Circle

157. The ____ is the "middle" of the cluster it need not be actual point in the cluster.

a) Radius
b) Diameter
c) Centroid
d) Metoid

158. The cluster is represented by one centrally located object in the cluster called a_____.

a) Centroid
b) Medoid
c) Radius
d) Diameter

159. The ____is the square root of the average mean squared distance from any point in the cluster to centroid.

a) Radius
b) Medoid
c) Diameter
d) Centroid

160. The _____is the square root of the average mean squared distance between all pairs of points in the cluster.

a) Radius
b) Medoid
c) Diameter
d) Centroid

161. Largest distance between an element in one cluster and an element in the other is_____.

a) Single Link
b) Complete Link
c) Average Link
d) Centroid

162. Smallest distance between an element in the cluster and an element in the other is____.

a) Centroid
b) Medoid
c) Complete link
d) Single link

163. _____ are sample points with values much different from those of the remaining set of data.

a) Centroid
b) Medoid
c) Outliers
d) Compression

164. In hierarchical clustering , a tree data structure is called _____.

a) Connected component
b) Dendrogram
c) Minimum spanning tree
d) Bond energy

165. The root in the dendrogram tree contains ____ clusters ,where all elements aretogether.
   a) Four
   b) Three
   c) Two
   d) One
166. The space complexity for hierarchical algorithms is_____.
   a) O(n)
   b) O(N+2)
   c) O(n2)
   d) O(2N)
167. A _____ component is a graph in which there exists a path between any two vertices.
   a) Connected
   b) Unconnected
   c) Nested
   d) Stylish
168. A _____ is a maximal graph in which there is an edge between vertices.
   a) Connected graph
   b) Clique
   c) Candidates
   d) Dendrogram
169. The ____are sample points with values much different from those of the remaining set of data.
   a) Clusters
   b) Outliers
   c) Candidates
   d) Mining
170. _____is the process of identifying outliers in a set of data.
   a) Outlier detection
   b) Outlier avoidance
   c) Outlier collision
   d) Outlier prediction
171. The outliers can be detected by well-known tests such as _____-.
   a) Chi-square test
   b) Random test
   c) Discordancy test
   d) Unit test

172. Clustering applications include plant and _____ classifications.
   a) Medical
   b) Biological
   c) Zoological
   d) Animal
173. ____ clustering , all items are initially placed in one cluster and clusters are repeat.
   a) Random
   b) Divisive
   c) Nearest neighbour
   d) Partitional
174. BEA stands for__.
   a) Band Echo Algorithm
   b) Bond Echo Algorithm
   c) Balance Energy Algorithm
   d) Bond Energy Algorithm
175. _____is an iterative clustering algorithm.
   a) K-means
   b) LARGE DB
   c) KDD
   d) BEA
176. The nearest neighbor algorithm uses _____technique.
   a) Single link
   b) Complete link
   c) Average link
   d) Centroid
177. The PAM algorithm also called _____algorithm.
   a) K-means
   b) K-medoids
   c) K-centroid
   d) K-radius
178. The time complexity of nearest neighbor algorithm is_____.
   a) O(n)
   b) O(N+2)
   c) O(n2)
   d) O(2N)

179. In a distributed database, each resulting cluster is called a _____.
   a) Horizontal Fragment
   b) Vertical Fragment
   c) Both(a) & (b)
   d) None
180. In neural network, the number of input nodes is the same as the number of___.
   a) Levels
   b) Clusters
   c) Points
   d) Attributes
181. The goal of _____ is to discover both the dense and sparse regions of a data set.
   a) Association rule
   b) Classification
   c) Clustering
   d) Genetic Algorithm
182. _____ clustering techniques starts with all records in one cluster and then try to split that cluster into small pieces.
   a) Agglomerative
   b) Divisive
   c) Partition
   d) Numeric
183. _____ seeks to find groups of closely related observations so that observations that belong the same cluster are more similar to each other.
   a) Association
   b) Anomaly detection
   c) Clustering
   d) None
184. In web mining, _____ is used to find natural groupings of users, pages, etc.
   a) Clustering
   b) Associations
   c) Sequential analysis
   d) Classification
185. In _____ algorithm each cluster is represented by the center of gravity of the cluster.
   a) k-medoid
   b) k-means
   c) STIRR
   d) ROCK

186. In _____ each cluster is represented by one of the objects of the cluster located near the center.
   a) k-medoid
   b) k-means
   c) STIRR
   d) ROCK
187. Pick out a k-medoid algoithm.
   a) DBSCAN
   b) BIRCH
   c) PAM
   d) CURE
188. Pick out a hierarchical clustering algorithm.
   a) DBSCAN
   b) BIRCH
   c) PAM
   d) CURE
189. _____ is the process of identifying outliers in a set of data.
   a) Outlier
   b) Outlier detection
   c) Segmentation
   d) Processing
190. The space complexity of adjacency matrix is_____.
   a) O(n)
   b) O(kn)
   c) O(n2)
   d) None o the above
191. A variation of complete link algorithm is called the _____.
   a) Farthest nearest neighbor
   b) Nearest neighbor
   c) Average
   d) Single
192. A tree data structure called_____ is used to illustrate the hierarchical clustering technique.
   a) Dendogramming
   b) Dendo
   c) Dendogram
   d) Dendograms

193. The term _____ indicates the ability of these NN to organize the nodes into clusters based on the similarity between them.

a) Competitive
b) Non-competitive
c) Self organizing
d) None of the above

194. CF stands for_____

a) Clustering Features
b) Clustering future
c) Classification Features
d) Classification Future

195. The space complexity for K-means is_____.

a) O(n)
b) O(kn)
c) n
d) O(n2)

196. The squared error algorithm has _____ type.

a) Hierarchical
b) Partitional
c) Mixed
d) Agglomeative.

197. The time complexity for single link algorithm is_____.

a) O(kn2)
b) O(n)
c) O(kn)
d) O(n2)

198. The squared error clustering algorithm minimizes_____ .

a) Error
b) Squared error
c) Square
d) All the above

199. With _____ clustering the algorithm creates only one set of clusters.

a) Partitional
b) Hierarchical
c) Agglomerative
d) None of the above

200. _____ techniques use labeling of the items to assist in the classification process.

a) Intrinsic

b) Extrinsic

c) Both a,b

d) All the above

UNIT-V

201. The purchasing of one product when another product is purchased represents an_____.

a) Decision Tree

b) Association Rule

c) Classification

d) Clustering

202. The _____of an item is the percentage of transactions in which that item occurs.

a) Confidence

b) Support

c) Association rule

d) Itemset

203. The _____is called the number of scans of the database.

a) Support

b) Confidence

c) Strength

d) Both (b) & (c)

204. Potentially large item sets are called ____.

a) Support

b) Confidence

c) Candidates

d) Itemset

205. In association rule algorithm, the notation "P" indicates.

a) Confidence

b) Candidates

c) Partitions

d) Transactions

206. Any subset of a large itemset must be _____.

a) Small

b) Medium

c) Average

d) Large

207. The large itemsets are also said to be _____closure.
   a) Upward
   b) Middleware
   c) Downward
   d) None
208. Additional candidates are determined by applying the _____ border function.
   a) Positive
   b) Negative
   c) Average
   d) Medium
209. The Apriori algorithm shows the sample is performed using a support called __.
   a) High
   b) Low
   c) Average
   d) Smalls
210. The basic _____ reduces the number of database scans to two.
   a) Divisive algorithm
   b) Parallel algorithm
   c) Partition algorithm
   d) Sampling algorithm
211. The candidates are partitioned and counted separately at each processor is called____.
   a) Data parallelism
   b) Task parallelism
   c) Candidates
   d) Data reduction
212. One data parallelism algorithm is the _____.
   a) MSE
   b) FIS
   c) DDA
   d) CDA
213. One task parallelism algorithm is called _____.
   a) CDA
   b) MSE
   c) DDA
   d) BCD

214. An algorithm all rules that satisfy a given support and confidence level is called____.
   a) Target
   b) Type
   c) Data type
   d) Data source
215. The most common data structure used to store the candidates itemsets and their counts is a_____.
   a) Binary tree
   b) B-tree
   c) Balanced tree
   d) Hash tree
216. Which technique is used to improve on the performance of an algorithm given distribution Or amount of main memory?
   a) Architecture
   b) Optimization
   c) Parallelism
   d) Itemset
217. A leaf node in the hash tree contains_____.
   a) Attributes
   b) Itemset
   c) Candidates
   d) Data
218. One incremental approach,_____is based on the Apriori algorithm.
   a) CDA
   b) DDA
   c) fast update
   d) slow update
219. A variation of generalized rules are _____ association rules.
   a) Multiple-level
   b) Hierarchical-level
   c) Multi-level
   d) Hybrid-level
220. A _____ association rule is one that involves categorical and quantitative data.
   a) Categorical
   b) Qualitative
   c) Quantitative
   d) Spanning

221. MIS stands for _____.
   a) Medium item support
   b) Maximum item support
   c) Minimum item support
   d) Medium item scale
222. A __rule is defined as a set of itemsets that are correlated.
   a) Correlation
   b) Co-efficient
   c) MIS
   d) Modification
223. Correlation(A=>B)= _____?
   a) P(A,B) / P(A)P(B)
   b) (b)P(A) / (P(A) P(B)
   c) P(B) / P(A) P(B)
   d) P(A) P(B) / P(A) – P(B)
224. Conviction has a value of ____ if A and B are not related.
   a) 0
   b) 1
   c) 2
   d) ∞
225. Which one is not an association rule algorithm?
   a) Apriori
   b) CDA
   c) DDA
   d) PAM
226. _____ algorithms may be able to adapt better to limited main memory.
   a) Divisive
   b) Sampling
   c) Partitioning
   d) Distributed
227. During the _____ scan, additional candidates are generated and counted.
   a) First
   b) Second
   c) Third
   d) Fourth

228. Chi-squared statistic is denoted by the _____symbol.
   a) X2
   b) E[X]
   c) 2X
   d) X3
229. ___ are used to show the relationships between data items.
   a) Clustering
   b) Regression
   c) Association rules
   d) Classification
230. The most two important property of an association rules are _____.
   a) Support, confidence
   b) Itemset, data
   c) Neuron, gene
   d) Lift, interest
231. A _____ is defined as a set of itemsets that are correlated.
   a) Correlation rule
   b) Association rule
   c) Conviction
   d) Probability of correlation
232. Confidence or strength are indicated by _____.
   a) ©
   b) ®
   c) €
   d) α
233. In association rule l stands for_____.
   a) Large item sets in L
   b) Set of large item set
   c) Both a,b
   d) None of the above
234. _____is the most well known association rule algorithm and is used in most commercial products.
   a) Apriori algorithm
   b) Partition algorithm
   c) Distributed algorithm
   d) Pincer-search algorithm

235. The basic idea of the apriori algorithm is to generate_____ item sets of a particular size & scans the database.

a) Candidate
b) Primary
c) Secondary
d) Superkey

236. The number of iterations in a priori _____.

a) Increases with the size of the maximum frequent set.
b) Decreases with increase in size of the maximum frequent set.
c) Increases with the size of the data.
d) Decreases with the increase in size of the data.

237. After the pruning of a priori algorithm, _____ will remain.

a) Only candidate set
b) No candidate set
c) Only border set
d) No border set

238. The a priori frequent itemset discovery algorithm moves _____ in the lattice.

a) Upward
b) Downward
c) Breadthwise
d) Both upward and downward

239. The _____ step eliminates the extensions of (k-1)-itemsets which are not found to be frequent, from being considered for counting support.

a) Candidate generation
b) Pruning
c) Partitioning
d) Itemset eliminations

240. The second phaase of A Priori algorithm is _____.

a) Candidate generation
b) Itemset generation
c) Pruning
d) Partitioning

241. The first phase of A Priori algorithm is _____.

a) Candidate generation
b) Itemset generation
c) Pruning
d) Partitioning

242. The A Priori algorithm is a _____.

a) top-down search

b) breadth first search

c) depth first search

d) bottom-up search

243. A priori algorithm is otherwise called as _____.

a) width-wise algorithm

b) level-wise algorithm

c) pincer-search algorithm

d) FP growth algorithm

244. The right hand side of an association rule is called _____.

a) Consequent

b) Onset

c) Antecedent

d) Precedent

245. The left hand side of an association rule is called _____.

a) Consequent

b) Onset

c) Antecedent

d) Precedent

246. The value that says that transactions in D that support X also support Y is called _____.

a) Confidence

b) Support

c) Support count

d) None Of the above

247. The absolute number of transactions supporting X in T is called _____.

a) Confidence

b) Support

c) Support count

d) None Of the above

248. _____ are effective tools to attack the scalability problem.

a) Sampling

b) Parallelization

c) Both A & B

d) None of the above

249. Discovery of cross-sales opportunities is called _____.
   a) Segmentation
   b) Visualization
   c) Correction
   d) Association

250. In web mining, _____ is used to know which URLs tend to be requested together.
   a) Clustering
   b) Associations
   c) Sequential analysis
   d) Classification

ANSWER KEY

UNIT-I

1	A	2	C	3	C	4	A	5	C	6	A	7	B	8	A	9	C	10	A
11	D	12	D	13	A	14	A	15	D	16	B	17	A	18	A	19	C	20	C
21	B	22	B	23	A	24	C	25	B	26	A	27	C	28	A	29	A	30	A
31	D	32	C	33	B	34	D	35	B	36	D	37	D	38	B	39	D	40	B
41	A	42	A	43	D	44	C	45	B	46	A	47	A	48	D	49	A	50	D

UNIT-II

51	**A**	**52**	**B**	**53**	**D**	**54**	**A**	**55**	**B**	**56**	**C**	**57**	**A**	**58**	**A**	**59**	**A**	**60**	**A**
61	C	62	B	63	C	64	A	65	B	66	C	67	D	68	B	69	B	70	C
71	C	72	B	73	A	74	C	75	C	76	B	77	B	78	C	79	B	80	A
81	A	82	B	83	A	84	C	85	B	86	C	87	B	88	A	89	B	90	A
91	C	92	A	93	A	94	B	95	C	96	C	97	D	98	B	99	B	100	A

UNIT-III

101	**B**	**102**	**A**	**103**	**D**	**104**	**A**	**105**	**C**	**106**	**A**	**107**	**D**	**108**	**C**	**109**	**A**	**110**	**C**
111	B	112	A	113	B	114	D	115	B	116	C	117	B	118	C	119	D	120	A
121	B	122	A	123	C	124	D	125	A	126	C	127	B	128	D	129	A	130	C
131	A	132	C	133	A	134	C	135	B	136	B	137	C	138	C	139	C	140	A
141	A	142	A	143	A	144	A	145	C	146	C	147	B	148	B	149	A	150	D

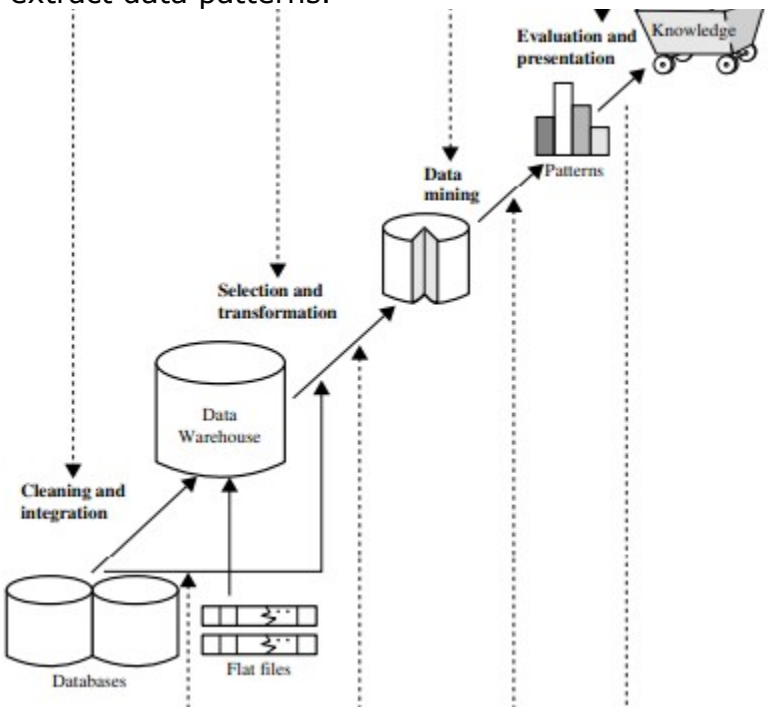
UNIT-IV

151	**C**	**152**	**A**	**153**	**D**	**154**	**C**	**155**	**B**	**156**	**B**	**157**	**C**	**158**	**B**	**159**	**A**	**160**	**C**
161	B	162	D	163	C	164	B	165	D	166	C	167	A	168	B	169	B	170	A
171	C	172	D	173	B	174	D	175	A	176	A	177	B	178	C	179	B	180	D
181	C	182	B	183	C	184	A	185	B	186	A	187	C	188	A	189	B	190	C
191	A	192	C	193	C	194	A	195	A	196	B	197	A	198	B	199	A	200	B

UNIT-V

201	B	202	B	203	D	204	C	205	C	206	D	207	C	208	B	209	D	210	C
211	B	212	D	213	C	214	A	215	D	216	B	217	C	218	C	219	A	220	C
221	C	222	A	223	A	224	B	225	D	226	A	227	B	228	A	229	C	230	A
231	A	232	D	233	A	234	A	235	A	236	A	237	B	238	A	239	B	240	C
241	A	242	B	243	A	244	C	245	A	246	A	247	C	248	C	249	D	250	B

1. ..... is an essential process where intelligent methods are applied to extract data patterns.



A) Data warehousing

B) Data mining

C) Text mining

D) Data selection

2. Data mining can also applied to other forms such as .....

i) Data streams

ii) Sequence data

iii) Networked data

iv) Text data

v) Spatial data

A) i, ii, iii and v only

B) ii, iii, iv and v only

C) i, iii, iv and v only

D) All i, ii, iii, iv and v

3. Which of the following is not a data mining functionality?

A) Characterization and Discrimination

B) Classification and regression

C) Selection and interpretation

D) Clustering and Analysis

4. ..... is a summarization of the general characteristics or features of a target class of data.

A) Data Characterization

B) Data Classification

C) Data discrimination

D) Data selection

5. ..... is a comparison of the general features of the target class data objects against the general features of objects from one or multiple contrasting classes.

A) Data Characterization

B) Data Classification

C) Data discrimination

D) Data selection

6. Strategic value of data mining is .....

A) cost-sensitive

B) work-sensitive

C) time-sensitive

D) technical-sensitive

7. ..... is the process of finding a model that describes and distinguishes data classes or concepts.

A) Data Characterization

B) Data Classification

C) Data discrimination

D) Data selection

8. The various aspects of data mining methodologies is/are .....

i) Mining various and new kinds of knowledge

ii) Mining knowledge in multidimensional space

iii) Pattern evaluation and pattern or constraint-guided mining.

iv) Handling uncertainty, noise, or incompleteness of data

A) i, ii and iv only

B) ii, iii and iv only

C) i, ii and iii only

D) All i, ii, iii and iv

9. The full form of KDD is .....

A) Knowledge Database

B) Knowledge Discovery Database

C) Knowledge Data House

D) Knowledge Data Definition

10. The out put of KDD is .....

A) Data

B) Information

C) Query

D) Useful information

Data Warhouse & Data Mining 700 - MCQ's

TOPIC ONE – INTRODUCTION TO DATA MINING

EASY QUESTIONS

1.Data mining is an integral part of _____.
A. SE.
B. DBMS.
C. KDD.
D. OS.
ANSWER: C

2. _____ is a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management decisions.
A. Data Mining.
B. Data Warehousing.
C. Web Mining.
D. Text Mining.
ANSWER: B

3. KDD describes the _____.
A. whole process of extraction of knowledge from data
B. extraction of data
C. extraction of information
D. extraction of rules
ANSWER: A

4. The data Warehouse is_____.
A. read only.
B. write only.
C. read write only.
D. none.
ANSWER: A

5. Expansion for DSS in DW is_____.
A. Decision Support system.
B. Decision Single System.
C. Data Storable System.
D. Data Support System.
ANSWER: A

6. 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

7. The data is stored, retrieved & updated in _____.
A. OLAP.
B. OLTP.
C. SMTP.
D. FTP.
ANSWER: B

8. _____describes the data contained in the data warehouse.
A. Relational data.
B. Operational data.
C. Metadata.
D. Informational data.
ANSWER: C

9. _____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

10. _____ 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

11. _____defines the structure of the data held in operational databases and used byoperational applications.
A. User-level metadata.
B. Data warehouse metadata.
C. Operational metadata.
D. Data mining metadata.
ANSWER: C

12. _____ is held in the catalog of the warehouse database system.
A. Application level metadata.
B. Algorithmic level metadata.
C. Departmental level metadata.
D. Core warehouse metadata.
ANSWER: B

13. _____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

14. Data can be updated in _____environment.
A. data warehouse.
B. data mining.
C. operational.
D. informational.
ANSWER: C

15. Record cannot be updated in _____.
A. OLTP
B. files
C. RDBMS
D. data warehouse
ANSWER: D

16. Detail data in single fact table is otherwise known as_____.
A. monoatomic data.
B. diatomic data.
C. atomic data.
D. multiatomic data.
ANSWER: C

17. A data warehouse is _____.
A. updated by end users.
B. contains numerous naming conventions and formats
C. organized around important subject areas.
D. contains only current data.
ANSWER: C

18. _____ is data about data.
A. Metadata.
B. Microdata.
C. Minidata.
D. Multidata.
ANSWER: A

19. _____ is an important functional component of the metadata.
A. Digital directory.
B. Repository.
C. Information directory.
D. Data dictionary.
ANSWER: C

20. The term that is not associated with data cleaning process is _____.
A. domain consistency.
B. deduplication.
C. disambiguation.
D. segmentation.
ANSWER: D

21. Capability of data mining is to build _____ models.
A. retrospective.
B. interrogative.
C. predictive.
D. imperative.

ANSWER: C

22. _____ is a process of determining the preference of customer's majority.
A. Association.
B. Preferencing.
C. Segmentation.
D. Classification.
ANSWER: B

23. Exceptional reporting in data warehousing is otherwise called as _____.
A. exception.
B. alerts.
C. errors.
D. bugs.
ANSWER: B

24. The full form of KDD is _____.
A. Knowledge database.
B. Knowledge discovery in database.
C. Knowledge data house.
D. Knowledge data definition.
ANSWER: B

25. Removing duplicate records is a process called _____.
A. recovery.
B. data cleaning.
C. data cleansing.
D. data pruning.
ANSWER: B

26. _____ helps to integrate, maintain and view the contents of the data warehousing system.
A. Business directory.
B. Information directory.
C. Data dictionary.
D. Database.
ANSWER: B

27. Discovery of cross-sales opportunities is called _____.
A. segmentation.
B. visualization.
C. correction.
D. association.
ANSWER: D

28. Data marts that incorporate data mining tools to extract sets of data are called _____.
A. independent data mart.
B. dependent data marts.
C. intra-entry data mart.
D. inter-entry data mart.
ANSWER: B

29. A directory to help the DSS analyst locate the contents of the data warehouse is seen in _____.
A. Current detail data.
B. Lightly summarized data.
C. Metadata.
D. Older detail data.
ANSWER: C

30. Which of the following is not an old detail storage medium?
A. Phot Optical Storage.
B. RAID.
C. Microfinche.
D. Pen drive.
ANSWER: D

31. The dimension tables describe the _____.
A. entities.
B. facts.
C. keys.
D. units of measures.
ANSWER: B

32. Which of the following is not the other name of Data mining?
A. Exploratory data analysis.
B. Data driven discovery.
C. Deductive learning.
D. Data integration.
ANSWER: D

33. Which of the following is a predictive model?
A. Clustering.
B. Regression.
C. Summarization.
D. Association rules.
ANSWER: B

34. Which of the following is a descriptive model?
A. Classification.
B. Regression.
C. Sequence discovery.
D. Association rules.
ANSWER: C

35. A _____ model identifies patterns or relationships.
A. Descriptive.
B. Predictive.
C. Regression.
D. Time series analysis.
ANSWER: A

36. A predictive model makes use of _____.
A. current data.
B. historical data.
C. both current and historical data.
D. assumptions.

ANSWER: B

37. _____ maps data into predefined groups.
A. Regression.
B. Time series analysis
C. Prediction.
D. Classification.
ANSWER: D

38. _____ is used to map a data item to a real valued prediction variable.
A. Regression.
B. Time series analysis.
C. Prediction.
D. Classification.
ANSWER: B

39. In _____, the value of an attribute is examined as it varies over time.
A. Regression.
B. Time series analysis.
C. Sequence discovery.
D. Prediction.
ANSWER: B

40. In _____ the groups are not predefined.
A. Association rules.
B. Summarization.
C. Clustering.
D. Prediction.
ANSWER: C

41. _____ is the input to KDD.
A. Data.
B. Information.
C. Query.
D. Process.
ANSWER: A

42. The output of KDD is _____.
A. Data.
B. Information.
C. Query.
D. Useful information.
ANSWER: D

43. The KDD process consists of _____ steps.
A. three.
B. four.
C. five.
D. six.
ANSWER: C

44. Treating incorrect or missing data is called as _____.
A. selection.
B. preprocessing.
C. transformation.
D. interpretation.
ANSWER: B

45. Converting data from different sources into a common format for processing is called as _____.
A. selection.
B. preprocessing.
C. transformation.
D. interpretation.
ANSWER: C

46. Various visualization techniques are used in _____ step of KDD.
A. selection.
B. transformaion.
C. data mining.
D. interpretation.
ANSWER: D

47. Extreme values that occur infrequently are called as _____.
A. outliers.
B. rare values.
C. dimensionality reduction.
D. Inliers
ANSWER: A

48. Box plot and scatter diagram techniques are _____.
A. Graphical.
B. Geometric.
C. Icon-based.
D. Pixel-based.
ANSWER: B

49. _____ is used to proceed from very specific knowledge to more general information.
A. Induction.
B. Compression.
C. Approximation.
D. Substitution.
ANSWER: A

50. Describing some characteristics of a set of data by a general model is viewed as
A. Induction.
B. Compression.
C. Approximation.
D. Summarization.
ANSWER: B

51. _____ helps to uncover hidden information about the data.
A. Induction.
B. Compression.
C. Approximation.

D. Summarization.
ANSWER: C

52. Incorrect or invalid data is known as _____.
A. changing data.
B. noisy data.
C. outliers.
D. missing data.
ANSWER: B

53. The _____ of data could result in the disclosure of information that is deemed to be confidential.
A. authorized use.
B. unauthorized use.
C. authenticated use.
D. unauthenticated use.
ANSWER: B

54. _____ data are noisy and have many missing attribute values.
A. Preprocessed.
B. Cleaned.
C. Real-world.
D. Transformed.
ANSWER: C

55. _____ describes the discovery of useful information from the web contents.
A. Web content mining.
B. Web structure mining.
C. Web usage mining.
D. Web development.
ANSWER: A

56. _____ is concerned with discovering the model underlying the link structures of the web.
A. Web content mining.
B. Web structure mining.
C. Web usage mining.
D. Web development.
ANSWER: B

57. A _____ algorithm takes all the data at once and tries to create a hypothesis based on this data.
A. supervised.
B. batch learning.
C. unsupervised.
D. incremental learning.
ANSWER: B

58. A _____ algorithm takes a new piece of information at each learning cycle and tries to revise the theory using
new data.
A. supervised.
B. batch learning.
C. unsupervised.
D. incremental learning.
ANSWER: B

59. _____ is used to find the vaguely known data.
A. SQL.
B. KDD.
C. Data mining.
D. Sybase.
ANSWER: C

60. The easiest way to gain access to the data and facilitate effective decision making is to set up a _____.
A. database.
B. data mart.
C. data warehouse.
D. operational.
ANSWER: C

61. Smaller local data warehouse is called as ____.
A. data mart.
B. database.
C. data model.
D. meta data.
ANSWER: B

62. The _____ data are stored in data warehouse.
A. operational.
B. historical.
C. transactional.
D. optimized.
ANSWER: B

63. A decision support system is a system that _____.
A. can constantly change over time.
B. cannot change.
C. copies the data.
D. supports the system.
ANSWER: A

64. Metadata is used by the end users for _____.
A. managing database.
B. structuring database.
C. querying purposes.
D. making decisions.
ANSWER: C

65. The _____ techniques are used to load information from operational database to data warehouse.
A. reengineering.
B. reverse.
C. transfer.
D. replication.
ANSWER: D

66. In machine learning _____ phase try to find the patterns from observations.
A. observation
B. theory
C. analysis
D. prediction
ANSWER: C

67. Information content is closely related to _____ and transparency.
A. algorithm.
B. search space.
C. learning.
D. statistical significance.
ANSWER: D

68. The _____ is used to express the hypothesis describing the concept.
A. computer language.
B. algorithm.
C. definition.
D. theory
ANSWER: A

69. A definition of a concept is complete if it recognizes _____.
A. all the information.
B. all the instances of a concept.
C. only positive examples.
D. negative examples.
ANSWER: B

70. The results of machine learning algorithms are always have to be checked for their _____.
A. observations.
B. calculations
C. programs.
D. statistical relevance.
ANSWER: D

71. A _____ is necessary condition for KDDs effective implement.
A. data set.
B. database.
C. data warehouse.
D. data.
ANSWER: C

72. KDD is a _____.
A. new technology that is use to store data.
B. multidisciplinary field of research.
C. database technology.
D. expert system.
ANSWER: B

INTERMEDIATE QUESTIONS

73. The generic two-level data warehouse architecture includes _____.
A. at least one data mart.
B. data that can extracted from numerous internal and external sources.
C. near real-time updates.
D. far real-time updates.
ANSWER: C

74. Reconciled data is _____.
A. data stored in the various operational systems throughout the organization.
B. current data intended to be the single source for all decision support systems.
C. data stored in one operational system in the organization.
D. data that has been selected and formatted for end-user support applications.
ANSWER: B

75. Transient data is _____.
A. data in which changes to existing records cause the previous version of the records to be eliminated.
B. data in which changes to existing records do not cause the previous version of the records to be eliminated.
C. data that are never altered or deleted once they have been added.
D. data that are never deleted once they have been added.
ANSWER: A

76. The extract process is _____.
A. capturing all of the data contained in various operational systems.
B. capturing a subset of the data contained in various operational systems.
C. capturing all of the data contained in various decision support systems.
D. capturing a subset of the data contained in various decision support systems.
ANSWER: B

77. Data transformation includes _____.
A. a process to change data from a detailed level to a summary level.
B. a process to change data from a summary level to a detailed level.
C. joining data from one source into various sources of data.
D. separating data from one source into various sources of data.
ANSWER: A

78. _____ is the goal of data mining.
A. To explain some observed event or condition.
B. To confirm that data exists.
C. To analyze data for expected relationships.
D. To create a new data warehouse.
ANSWER: A

79. Business Intelligence and data warehousing is not used for _____.
A. Forecasting.
B. Data Mining.
C. Analysis of large volumes of product sales data.
D. Discarding data.
ANSWER: D

80. Classification rules are extracted from _____.
A. root node.
B. decision tree.
C. siblings.
D. branches.
ANSWER: B

81. Reducing the number of attributes to solve the high dimensionality problem is called as _____.
A. dimensionality curse.
B. dimensionality reduction.
C. cleaning.
D. Overfitting.
ANSWER: B

82. Data that are not of interest to the data mining task is called as _____.
A. missing data.
B. changing data.
C. irrelevant data.
D. noisy data.
ANSWER: C

83. Data mining helps in _____.
A. inventory finalisation.
B. sales.
C. marketing products.
D. Debt collection.
ANSWER: A

84. Which of the following is not a desirable feature of any efficient algorithm?
A. to reduce number of input operations.
B. to reduce number of output operations.
C. to be efficient in computing.
D. to have maximal code length.
ANSWER: D

85. All set of items whose support is greater than the user-specified minimum support are called as
A. border set.
B. frequent set.
C. maximal frequent set.
D. lattice.
ANSWER: B

86. Metadata describes _____.
A. contents of database.
B. structure of contents of database.
C. structure of database.
D. database itself.
ANSWER: B

87. The partition of overall data warehouse is _____.
A. database.
B. data cube.
C. data mart.

D. operational data.
ANSWER: C

88. The information on two attributes is displayed in _____ in scatter diagram.
A. visualization space.
B. scatter space.
C. cartesian space.
D. interactive space.
ANSWER: C

89. OLAP is used to explore the _____ knowledge.
A. shallow.
B. deep.
C. multidimensional.
D. hidden.
ANSWER: C

90. Hidden knowledge can be found by using _____.
A. searching algorithm.
B. pattern recognition algorithm.
C. searching algorithm.
D. clues.
ANSWER: B

91. The next stage to data selection in KDD process _____.
A. enrichment.
B. coding.
C. cleaning.
D. reporting.
ANSWER: C

92. Enrichment means ____.
A. adding external data.
B. deleting data.
C. cleaning data.
D. selecting the data.
ANSWER: A

93. The decision support system is used only for _____.
A. cleaning.
B. coding.
C. selecting.
D. queries.
ANSWER: D

94. Which of the following is closely related to statistical significance and transparency?
A. Classification Accuracy.
B. Transparency.
C. Statistical significance.
D. Search Complexity.
ANSWER: B

95. _____ is the technique which is used for discovering patterns in dataset at the beginning of data mining process.
A. Kohenon map.
B. Visualization.
C. OLAP.
D. SQL.
ANSWER: B

96. _____ is the heart of knowledge discovery in database process.
A. Selection.
B. Data ware house.
C. Data mining.
D. Creative coding.
ANSWER: D

97. In KDD and data mining, noise is referred to as _____.
A. repeated data.
B. complex data.
C. meta data.
D. random errors in database.
ANSWER: D

98. The technique of learning by generalizing from examples is _____.
A. incremental learning.
B. inductive learning.
C. hybrid learning.
D. generalized learning.
ANSWER: B

99. The _____ plays an important role in artificial intelligence.
A. programming skill.
B. scheduling.
C. planning.
D. learning capabilities.
ANSWER: D

100. Data mining is used to refer _____ stage in knowledge discovery in database.
A. selection.
B. retrieving.
C. discovery.
D. coding.
ANSWER: C

101. _____ could generate rule automatically.
A. KDD.
B. machine learning.
C. artificial intelligence.
D. expert system.
ANSWER: B

102. A good introduction to machine learning is the idea of _____.
A. concept learning.
B. content learning.
C. theory of falsification.

D. Poppers law.
ANSWER: A

103. The algorithms that are controlled by human during their execution is _____ algorithm.
A. unsupervised.
B. supervised.
C. batch learning.
D. incremental.
ANSWER: B

104. Background knowledge depends on the form of _____.
A. theoretical knowledge.
B. hypothesis.
C. formulae.
D. knowledge representation.
ANSWER: D

ADVANCED QUESTIONS

105. Dimensionality reduction reduces the data set size by removing _____.
A. relevant attributes.
B. irrelevant attributes.
C. derived attributes.
D. composite attributes.
ANSWER: B

106. The main organizational justification for implementing a data warehouse is to provide _____.
A. cheaper ways of handling transportation.
B. decision support.
C. storing large volume of data.
D. access to data.
ANSWER: C

107. Multidimensional database is otherwise known as_____.
A. RDBMS
B. DBMS
C. EXTENDED RDBMS
D. EXTENDED DBMS
ANSWER: B

108. _____ are designed to overcome any limitations placed on the warehouse by the nature of therelational data model.
A. Operational database.
B. Relational database.
C. Multidimensional database.
D. Data repository.
ANSWER: C

109. If a set is a frequent set and no superset of this set is a frequent set, then it is called _____.
A. maximal frequent set.
B. border set.
C. lattice.
D. infrequent sets.

ANSWER: A

110. The goal of _____ is to discover both the dense and sparse regions of a data set.
A. Association rule.
B. Classification.
C. Clustering.
D. Genetic Algorithm.
ANSWER: C

111. Rule based classification algorithms generate _____ rule to perform the classification.
A. if-then.
B. while.
C. do while.
D. switch.
ANSWER: A

112. _____ training may be used when a clear link between input data sets and target output valuesdoes not exist.
A. Competitive.
B. Perception.
C. Supervised.
D. Unsupervised.
ANSWER: D

113. Web content mining describes the discovery of useful information from the _____contents.
A. text.
B. web.
C. page.
D. level.
ANSWER: B

114. Research on mining multi-types of data is termed as _____ data.
A. graphics.
B. multimedia.
C. meta.
D. digital.
ANSWER: B

115. _____ is the way of studying the web link structure.
A. Computer network.
B. Physical network.
C. Social network.
D. Logical network.
ANSWER: C

116. In web mining, _____ is used to find natural groupings of users, pages, etc.
A. clustering.
B. associations.
C. sequential analysis.
D. classification.
ANSWER: A

117. In web mining, _____ is used to know which URLs tend to be requested together.
A. clustering.
B. associations.
C. sequential analysis.
D. classification.
ANSWER: B

118. The _____ engine for a data warehouse supports query-triggered usage of data
A. NNTP
B. SMTP
C. OLAP
D. POP
ANSWER: C

119. _____ displays of data such as maps, charts and other graphical representation allow data to be presented compactly to the users.
A. Hidden
B. Visual
C. Obscured
D. Concealed
ANSWER: B

120. Which of the following are the important qualities of good learning algorithm.
A. Consistent, Complete.
B. Information content, Complex.
C. Complete, Complex.
D. Transparent, Complex.
ANSWER: A

TOPIC TWO – GETTING TO KNOW YOUR DATA

EASY QUESTIONS

121. The _____ is a symbolic representation of facts or ideas from which information can potentially be extracted.
A. knowledge.
B. data.
C. algorithm.
D. program.
ANSWER: B

122. A collection of interesting and useful patterns in database is called _____.
A. knowledge.
B. information.
C. data.
D. algorithm.
ANSWER: A

123. The main organizational justification for implementing a data warehouse is to provide _____.
A. cheaper ways of handling transportation.
B. decision support.
C. storing large volume of data.
D. access to data.
ANSWER: C

124. The process of finding the right formal representing of a certain body of knowledge in order to represent it inknowledge based system is_____.
A. re-engineering.
B. replication.
C. knowledge engineering.
D. reverse engineering.
ANSWER: C

125. OR methods deals with _____type of data.
A. quantitative.
B. qualitative.
C. standard.
D. predict.
ANSWER: A

126. _____analysis divides data into groups that are meaningful, useful, or both.
A. Cluster.
B. Association.
C. Classifiction.
D. Relation.
ANSWER: A

127. A representation of data objects as columns and attributes as rows is called_____.
A. matrix.
B. data matrix.
C. table.
D. file.
ANSWER: B

128. Which of the following is not a data mining attribute?
A. nominal.
B. ordinal.
C. interval.
D. multiple.
ANSWER: D

129. Patterns of machine-language program are_____.
A. definitive theories.
B. hypothesis.
C. not-definitive theories.
D. quantitative.
ANSWER: B

130. Nominal and ordinal attributes are collectively referred to as_____ attributes.
A. qualitative.
B. perfect.
C. consistent.
D. optimized.
ANSWER: A

131. A data set can often be viewed as a collection of _____.
A. data mart.
B. data.

C. data object.
D. template.
ANSWER: C

132. An important element in machine learning is _____.
A. flow.
B. knowledge.
C. observation.
D. language.
ANSWER: C

133. _____ is the closeness of repeated measurements to one another.
A. Precision.
B. Bias.
C. Accuracy.
D. non-scientific.
ANSWER: A
ANSWER: B

134. Which of the following is not a data mining attribute?
A. nominal.
B. ordinal.
C. interval.
D. multiple.
ANSWER: D

135. Patterns of machine-language program are_____.
A. definitive theories.
B. hypothesis.
C. not-definitive theories.
D. quantitative.
ANSWER: B

136. Nominal and ordinal attributes are collectively referred to as_____ attributes.
A. qualitative.
B. perfect.
C. consistent.
D. optimized.
ANSWER: A

137. A data set can often be viewed as a collection of _____.
A. data mart.
B. data.
C. data object.
D. template.
ANSWER: C

138. An important element in machine learning is _____.
A. flow.
B. knowledge.
C. observation.
D. language.
ANSWER: C

139. _____ is used for discrete target variable.
A. Nominal.
B. Classification.
C. Clustering.
D. Association.
ANSWER: B

140. A goal of data mining includes which of the following?
A. To explain some observed event or condition
B. To confirm that data exists
C. To analyze data for expected relationships
D. To create a new data warehouse
ANSWER: A

141. is a subject-oriented, integrated, time-variant, nonvolatile collection of data in supportof management decisions.
A. Data Mining.
B. Data Warehousing.
C. Web Mining.
D. Text Mining.
ANSWER: B

142. Collection, analysis, interpretation or explanation of data.
A. Statistics
B. Information retrieval
C. Data mining
D. Cluster analysis
Answer: A

143. Data objects represesents
A. Values
B. Entity
C. Data
D. Attributes
Answer : B

INTERMEDIATE QUESTIONS

144. The term that is not associated with data cleaning process is _____.
A. domain consistance.
B. de-duplication.
C. disambiguation.
D. segmentation.
ANSWER: D

The _____ is a useful method of discovering patterns at the beginning of data mining process.
A. calculating distance.
B. visualization techniques.
C. decision trees.
D. association rules.
ANSWER: B

145. Data mining methodology states that in optimal situation data mining is an _____.
A. standard process.
B. complete process.
C. creative process.
D. ongoing process.
ANSWER: D

146. _____ is a knowledge discovery process.
A. Data cleaning.
B. Data warehousing.
C. Data mining.
D. Data transformation.
ANSWER: A

147. OLAP is used for _____.
A. online application processing.
B. online analytical processing.
C. online aptitude processing.
D. online administration and processing.
ANSWER: B

148. Which of the following is not an issue related to concept learning
A. Supervised learning.
B. Unsupervised learning.
C. Self learning.
D. Concept learning.
ANSWER: D

149. Removing duplicate records is a process called_____.
A. recovery.
B. data cleaning.
C. data cleansing.
D. data pruning.
ANSWER: B

150. Data marts that incorporate data mining tools to extract sets of data is called_____.
A. independent data mart.
B. dependent data marts.
C. intra-entry data mart.
D. inter-entry data mart.
ANSWER: B

151. The problem of finding hidden structure in unlabelled data is called…
A. Supervised learning
B. Unsupervised learning
C. Reinforcement learning
D. Semisupervised learning
ANSWER : B

152. Task of inferring a model from labelled training data is called
A. Supervised learning
B. Unsupervised learning
C. Reinforcement learning

D. Semisupervised learning
ANSWER : B

153. Self-organizing maps are an example of…
A. Supervised learning
B. Unsupervised learning
C. Reinforcement learning
D. Missing data imputation
ANSWER : A

154. 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

155. Classification rules are extracted from
A. root node
B. decision tree.
C. siblings.
D. branches.
ANSWER: B

156. Which one of the following is not a part of empirical cycle in scientific research?
A. Observation
B. Theory.
C. Self learning.
D. Prediction.
ANSWER: C

157. In machine learning _____ phase try to find the patterns from observations.
A. observation
B. theory
C. analysis
D. prediction
ANSWER: C

158. ANSWER: D
Data warehouse architecture is based on _____.
A. DBMS.
B. RDBMS.
C. Sybase.
D. SQL Server.
ANSWER: B

ADVANCED QUESTIONS

159. The ___ algorithm can be applied in cleaning data.
A. search.
B. pattern recognition.
C. learning.
D. clustering.
ANSWER: B

160. _____ is the type of pollution that is difficult to trace.
A. Duplication of records.
B. Ambiguition.
C. Lack of domain consistency.
D. Lack of information.
ANSWER: C

161. The statement that is true about data mining is _____.
A. data mining is not a single technique.
B. it finds the hidden patterns from data set.
C. it is a real discovery process.
D. all forms of pollutions are found during the data mining stage itself.
ANSWER: D

162. The first step in data mining project is _____.
A. rough analysis of data set using traditional query tools.
B. cleaning the data.
C. recognizing the patterns.
D. visualizing the patterns.
ANSWER: A

163. SQL can find _____ type of data.
A. narrow data.
B. multidimensional data.
C. shallow data.
D. hidden data.
ANSWER: C

164. _____ is used to find relationship between multidimensional data.
A. K-nearest neighbor.
B. Decision trees.
C. Association rules.
D. OLAP tools.
ANSWER: D

165. Which one of the following is not true about OLAP?
A. They create no new knowledge.
B. OLAP is powerful that data mining tool.
C. They cannot search for new solution.
D. OLAP tool store their data in special multidimensional format.
ANSWER: B

166. Genetic algorithm is viewed as a kind of_____.
A. meta learning strategy.
B. machine learning.
C. evolution.
D. OLAP tool.
ANSWER: A

167. The _____is a knowledge that can be found by using pattern recognition algorithm.
A. hidden knowledge.
B. deep.
C. shallow.

D. multidimensional.
ANSWER: A

168. Shannons notation of information content of message is_____.
A. Log 1divided by n equals log n.
B. log n equals log 1divided by n.
C. log 1divided by n equals minus log n.
D. log minus n =log 1divided by n.
ANSWER: C

169. Which of the following features usually applies to data in a data warehouse
A. Data are often deleted.
B. Most applications consist of transactions.
C. Data are rarely deleted.
D. Relatively few records are processed by applications.
ANSWER: C

170. Which of the following is true
A. The data warehouse consists of data marts and operational data
B. The Data Warehouse consists of data marts and application data.
C. The Data Warehouse is used as a source for the operational data.
D. The operational data are used as a source for the data warehouse
ANSWER: D

171. How do you better define a data warehouse as
A. Can be updated by end users.
B. Contains numerous naming conventions and formats.
C. Organized around important subject areas.
D. Contains only current data.
ANSWER: C

172. Which of the following is an operational system
A. A system that is used to run the business in real time and is based on historical data
B. A system that is used to run the business in real time and is based on current data.
C. A system that is used to support decision making and is based on current data.
D. A system that is used to support decision making and is based on historical data.
ANSWER: B

173. The generic two-level data warehouse architecture includes _____.
A. at least one data mart.
B. data that can extracted from numerous internal and external sources.
C. near off-time updates.
D. historic data.
ANSWER: B

174. Which of the following is reconciled data
A. Current data intended to be the single source for all decision support systems
B. Data stored in the various operational systems throughout the organization.
C. Data stored in one operational system in the organization.
D. Data that has been selected and formatted for end-user support applications.
ANSWER: A

175. Which of the following is an extract process
A. Capturing all of the data contained in various operational systems.
B. Capturing a subset of the data contained in various operational systems.
C. Capturing all of the data contained in various decision support systems.
D. Capturing a subset of the data contained in various decision support systems.
ANSWER: B

176. Which of the following is the not a types of clustering?
A. K-means.
B. Hiearachical.
C. Partitional.
D. Splitting.
ANSWER: D

177. Data Transformation includes_____.
A. a process to change data from a detailed level to a summary level.
B. a process to change data from a summary level to a detailed level.
C. joining data from one source into various sources of data.
D. separating data from one source into various sources of data.
ANSWER: A

178. The _____ is called a multi field transformation.
A. conversion of data from one field into multiple fields.
B. conversion of data from fields into field.
C. conversion of data from double fields into multiple fields
D. conversion of data from one field to one field.
ANSWER: A

179. Which of the given technology is not well-suited for data mining
A. Expert system technology.
B. Data visualization.
C. Technology limited to specific data types such as numeric data types.
D. Parallel architecture.
ANSWER: C

180. What is true about the multidimensional model?
A. It typically requires less disk storage.
B. It typically requires more disk storage.
C. Typical business queries requiring aggregate functions take more time.
D. Typical business queries requiring aggregate functions take more time.
ANSWER: B

181. Which of the following function involves data cleaning, data standardizing and summarizing
A. Storing data.
B. Transforming data.
C. Data acquisition.
D. Data Access.
ANSWER: B

182. Which of the following problems bog down the development of data mining projects
A. Financial problem.
B. Lack of technical assistance.
C. Lack of long-term vision.

D. Legal and privacy restrictions.
ANSWER: C

183. _____ is the closeness of repeated measurements to one another.
A. Precision.
B. Bias.
C. Accuracy.
D. non-scientific.
ANSWER: A

184. Which of the following matrix consist asymmetric data?
A. Sparse data matrix.
B. Indentity matrix.
C. Confusion matrix.
D. Cross matrix.
ANSWER: A

185. Which of the following matrix consist asymmetric data?
A. Sparse data matrix.
B. Indentity matrix.
C. Confusion matrix.
D. Cross matrix.
ANSWER: A

186. You are given data about seismic activity in Japan, and you want to predict a magnitude of the next earthquake, this is an example of
Supervised learning
Unsupervised learning
Serration
Dimensionality reduction
ANSWER: A

187. Algoritm is
A. It uses machine-learning technique. Here a program can learn from past experience.
B. Computational procedure that takes some values as input and procedure takes some value as output
C. Science of making machines perform tasks that would require intelligence when performed by humans
D. Processing procedure
ANSWER: A

188. The information on two attributes is displayed in _____ in scatter diagram.
A. visualization space.
B. scatter space.
C. cartesian space.
D. interactive space.
ANSWER: C

189. K-nearest neighbor is one of the _____.
A. learning technique.
B. OLAP tool.
C. purest search technique.
D. data warehousing tool.

ANSWER: C

190. In K- nearest neighbor the input is translated to _____.
A. values
B. points in multidimensional space
C. strings of characters
D. nodes
ANSWER: B

191. What is a tag cloud?
A. Is a visualization of statistics of user-preferred order.
B. Collection of data objects.
C. Data analysis
D. Data mining application
Answer: A

192. Analysis of variance is a statistical method of comparing the _____ of several populations.
A. standard deviations
B. variances
C. means
D. proportions
Answer: A

193. _____ is the specialized data warehouse database.
A. Oracle.
B. DBZ.
C. Informix.
D. Redbrick.
ANSWER: D

194. The source of all data warehouse data is the_____.
A. operational environment.
B. informal environment.
C. formal environment.
D. technology environment.
ANSWER: A

195. Which of the following is a descriptive model?
A. Classification.
B. Regression.
C. Sequence discovery.
D. Association rules.
ANSWER: C

196. A _____ model identifies patterns or relationships.
A. Descriptive.
B. Predictive.
C. Regression.
D. Time series analysis.
ANSWER: A

Dr.G.R.Damodaran College of Science

(Autonomous, affiliated to the Bharathiar University, recognized by the UGC)Re-accredited at the 'A' Grade Level by the **NAAC** and ISO 9001:2008 Certified
CRISL rated 'A' (TN) for MBA and MIB Programmes

II M.Sc(IT) [2012-2014]
Semester III
Core: Data Warehousing and Mining - 363U1
Multiple Choice Questions.

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. The data Warehouse is_____.
   A. read only.
   B. write only.
   C. read write only.
   D. none.
  ANSWER: A

3. Expansion for DSS in DW is_____.
   A. Decision Support system.
   B. Decision Single System.
   C. Data Storable System.
   D. Data Support System.
  ANSWER: A

4. 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

5. 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

6. The data is stored, retrieved & updated in _____.
   A. OLAP.
   B. OLTP.
   C. SMTP.

D. FTP.
ANSWER: B

7. _____describes the data contained in the data warehouse.
A. Relational data.
B. Operational data.
C. Metadata.
D. Informational data.
ANSWER: C

8. _____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

9. _____ 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

10. _____ is the specialized data warehouse database.
A. Oracle.
B. DBZ.
C. Informix.
D. Redbrick.
ANSWER: D

11. _____defines the structure of the data held in operational databases and used by operational applications.
A. User-level metadata.
B. Data warehouse metadata.
C. Operational metadata.
D. Data mining metadata.
ANSWER: C

12. _____ is held in the catalog of the warehouse database system.
A. Application level metadata.
B. Algorithmic level metadata.
C. Departmental level metadata.
D. Core warehouse metadata.
ANSWER: B

13. _____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

14. _____consists of formal definitions, such as a COBOL layout or a database schema.
A. Classical metadata.
B. Transformation metadata.
C. Historical metadata.
D. Structural metadata.
ANSWER: A

15. _____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

16. . _____databases are owned by particular departments or business groups.
A. Informational.
B. Operational.
C. Both informational and operational.
D. Flat.
ANSWER: B

17. The star schema is composed of _____ fact table.
A. one.
B. two.
C. three.
D. four.
ANSWER: A

18. 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

19. The key used in operational environment may not have an element of_____.
A. time.
B. cost.
C. frequency.
D. quality.
ANSWER: A

20. Data can be updated in _____environment.
A. data warehouse.
B. data mining.
C. operational.
D. informational.
ANSWER: C

21. Record cannot be updated in _____.
A. OLTP
B. files
C. RDBMS

D. data warehouse
ANSWER: D

22. The source of all data warehouse data is the_____.
A. operational environment.
B. informal environment.
C. formal environment.
D. technology environment.
ANSWER: A

23. Data warehouse contains_____data that is never found in the operational environment.
A. normalized.
B. informational.
C. summary.
D. denormalized.
ANSWER: C

24. Data redundancy between the environments results in less than _____percent.
A. one.
B. two.
C. three.
D. four.
ANSWER: A

25. Bill Inmon has estimated_____of the time required to build a data warehouse, is consumed in the conversion process.
A. 10 percent.
B. 20 percent.
C. 40 percent
D. 80 percent.
ANSWER: D

26. Detail data in single fact table is otherwise known as_____.
A. monoatomic data.
B. diatomic data.
C. atomic data.
D. multiatomic data.
ANSWER: C

27. _____test is used in an online transactional processing environment.
A. MEGA.
B. MICRO.
C. MACRO.
D. ACID.
ANSWER: D

28. _____ is a good alternative to the star schema.
A. Star schema.
B. Snowflake schema.
C. Fact constellation.
D. Star-snowflake schema.
ANSWER: C

29. The biggest drawback of the level indicator in the classic star-schema is that it limits_____.

A. quantify.
B. qualify.
C. flexibility.
D. ability.
ANSWER: C

30. A data warehouse is _____.
A. updated by end users.
B. contains numerous naming conventions and formats
C. organized around important subject areas.
D. contains only current data.
ANSWER: C

31. An operational system is _____.
A. used to run the business in real time and is based on historical data.
B. used to run the business in real time and is based on current data.
C. used to support decision making and is based on current data.
D. used to support decision making and is based on historical data.
ANSWER: B

32. The generic two-level data warehouse architecture includes _____.
A. at least one data mart.
B. data that can extracted from numerous internal and external sources.
C. near real-time updates.
D. far real-time updates.
ANSWER: C

33. The active data warehouse architecture includes _____
A. at least one data mart.
B. data that can extracted from numerous internal and external sources.
C. near real-time updates.
D. all of the above.
ANSWER: D

34. Reconciled data is _____.
A. data stored in the various operational systems throughout the organization.
B. current data intended to be the single source for all decision support systems.
C. data stored in one operational system in the organization.
D. data that has been selected and formatted for end-user support applications.
ANSWER: B

35. Transient data is _____.
A. data in which changes to existing records cause the previous version of the records to be eliminated.
B. data in which changes to existing records do not cause the previous version of the records to be eliminated.
C. data that are never altered or deleted once they have been added.
D. data that are never deleted once they have been added.
ANSWER: A

36. The extract process is _____.
A. capturing all of the data contained in various operational systems.
B. capturing a subset of the data contained in various operational systems.
C. capturing all of the data contained in various decision support systems.

D. capturing a subset of the data contained in various decision support systems.
ANSWER: B

37. Data scrubbing is _____.
A. a process to reject data from the data warehouse and to create the necessary indexes.
B. a process to load the data in the data warehouse and to create the necessary indexes.
C. a process to upgrade the quality of data after it is moved into a data warehouse.
D. a process to upgrade the quality of data before it is moved into a data warehouse
ANSWER: D

38. The load and index is _____.
A. a process to reject data from the data warehouse and to create the necessary indexes.
B. a process to load the data in the data warehouse and to create the necessary indexes.
C. a process to upgrade the quality of data after it is moved into a data warehouse.
D. a process to upgrade the quality of data before it is moved into a data warehouse.
ANSWER: B

39. Data transformation includes _____.
A. a process to change data from a detailed level to a summary level.
B. a process to change data from a summary level to a detailed level.
C. joining data from one source into various sources of data.
D. separating data from one source into various sources of data.
ANSWER: A

40. _____ is called a multifield transformation.
A. Converting data from one field into multiple fields.
B. Converting data from fields into field.
C. Converting data from double fields into multiple fields.
D. Converting data from one field to one field.
ANSWER: A

41. The type of relationship in star schema is _____.
A. many-to-many.
B. one-to-one.
C. one-to-many.
D. many-to-one.
ANSWER: C

42. Fact tables are _____.
A. completely demoralized.
B. partially demoralized.
C. completely normalized.
D. partially normalized.
ANSWER: C

43. _____ is the goal of data mining.
A. To explain some observed event or condition.
B. To confirm that data exists.
C. To analyze data for expected relationships.
D. To create a new data warehouse.
ANSWER: A

44. Business Intelligence and data warehousing is used for _____.
A. Forecasting.

B. Data Mining.
C. Analysis of large volumes of product sales data.
D. All of the above.
ANSWER: D

45. The data administration subsystem helps you perform all of the following, except_____.
A. backups and recovery.
B. query optimization.
C. security management.
D. create, change, and delete information.
ANSWER: D

46. The most common source of change data in refreshing a data warehouse is _____.
A. queryable change data.
B. cooperative change data.
C. logged change data.
D. snapshot change data.
ANSWER: A

47. _____ are responsible for running queries and reports against data warehouse tables.
A. Hardware.
B. Software.
C. End users.
D. Middle ware.
ANSWER: C

48. Query tool is meant for _____.
A. data acquisition.
B. information delivery.
C. information exchange.
D. communication.
ANSWER: A

49. Classification rules are extracted from _____.
A. root node.
B. decision tree.
C. siblings.
D. branches.
ANSWER: B

50. Dimensionality reduction reduces the data set size by removing _____.
A. relevant attributes.
B. irrelevant attributes.
C. derived attributes.
D. composite attributes.
ANSWER: B

51. _____ is a method of incremental conceptual clustering.
A. CORBA.
B. OLAP.
C. COBWEB.
D. STING.
ANSWER: C

52. Effect of one attribute value on a given class is independent of values of other attribute is called _____.
A. value independence.
B. class conditional independence.
C. conditional independence.
D. unconditional independence.
ANSWER: A

53. The main organizational justification for implementing a data warehouse is to provide _____.
A. cheaper ways of handling transportation.
B. decision support.
C. storing large volume of data.
D. access to data.
ANSWER: C

54. Maintenance of cache consistency is the limitation of _____.
A. NUMA.
B. UNAM.
C. MPP.
D. PMP.
ANSWER: C

55. Data warehouse architecture is based on _____.
A. DBMS.
B. RDBMS.
C. Sybase.
D. SQL Server.
ANSWER: B

56. Source data from the warehouse comes from _____.
A. ODS.
B. TDS.
C. MDDB.
D. ORDBMS.
ANSWER: A

57. _____ is a data transformation process.
A. Comparison.
B. Projection.
C. Selection.
D. Filtering.
ANSWER: D

58. The technology area associated with CRM is _____.
A. specialization.
B. generalization.
C. personalization.
D. summarization.
ANSWER: C

59. SMP stands for _____.
A. Symmetric Multiprocessor.
B. Symmetric Multiprogramming.
C. Symmetric Metaprogramming.

D. Symmetric Microprogramming.
ANSWER: A

60. _____ are designed to overcome any limitations placed on the warehouse by the nature of the relational data model.
A. Operational database.
B. Relational database.
C. Multidimensional database.
D. Data repository.
ANSWER: C

61. _____ are designed to overcome any limitations placed on the warehouse by the nature of the relational data model.
A. Operational database.
B. Relational database.
C. Multidimensional database.
D. Data repository.
ANSWER: C

62. MDDB stands for _____.
A. multiple data doubling.
B. multidimensional databases.
C. multiple double dimension.
D. multi-dimension doubling.
ANSWER: B

63. _____ is data about data.
A. Metadata.
B. Microdata.
C. Minidata.
D. Multidata.
ANSWER: A

64. _____ is an important functional component of the metadata.
A. Digital directory.
B. Repository.
C. Information directory.
D. Data dictionary.
ANSWER: C

65. EIS stands for _____.
A. Extended interface system.
B. Executive interface system.
C. Executive information system.
D. Extendable information system.
ANSWER: C

66. _____ is data collected from natural systems.
A. MRI scan.
B. ODS data.
C. Statistical data.
D. Historical data.
ANSWER: A

67. _____ is an example of application development environments.
A. Visual Basic.
B. Oracle.
C. Sybase.
D. SQL Server.
ANSWER: A

68. The term that is not associated with data cleaning process is _____.
A. domain consistency.
B. deduplication.
C. disambiguation.
D. segmentation.
ANSWER: D

69. _____ are some popular OLAP tools.
A. Metacube, Informix.
B. Oracle Express, Essbase.
C. HOLAP.
D. MOLAP.
ANSWER: A

70. Capability of data mining is to build _____ models.
A. retrospective.
B. interrogative.
C. predictive.
D. imperative.
ANSWER: C

71. _____ is a process of determining the preference of customer's majority.
A. Association.
B. Preferencing.
C. Segmentation.
D. Classification.
ANSWER: B

72. Strategic value of data mining is _____.
A. cost-sensitive.
B. work-sensitive.
C. time-sensitive.
D. technical-sensitive.
ANSWER: C

73. _____ proposed the approach for data integration issues.
A. Ralph Campbell.
B. Ralph Kimball.
C. John Raphlin.
D. James Gosling.
ANSWER: B

74. The terms equality and roll up are associated with _____.
A. OLAP.
B. visualization.
C. data mart.
D. decision tree.

ANSWER: C

75. Exceptional reporting in data warehousing is otherwise called as _____.
A. exception.
B. alerts.
C. errors.
D. bugs.
ANSWER: B

76. _____ is a metadata repository.
A. Prism solution directory manager.
B. CORBA.
C. STUNT.
D. COBWEB.
ANSWER: A

77. _____ is an expensive process in building an expert system.
A. Analysis.
B. Study.
C. Design.
D. Information collection.
ANSWER: D

78. The full form of KDD is _____.
A. Knowledge database.
B. Knowledge discovery in database.
C. Knowledge data house.
D. Knowledge data definition.
ANSWER: B

79. The first International conference on KDD was held in the year _____.
A. 1996.
B. 1997.
C. 1995.
D. 1994.
ANSWER: C

80. Removing duplicate records is a process called _____.
A. recovery.
B. data cleaning.
C. data cleansing.
D. data pruning.
ANSWER: B

81. _____ contains information that gives users an easy-to-understand perspective of the information stored in the data warehouse.
A. Business metadata.
B. Technical metadata.
C. Operational metadata.
D. Financial metadata.
ANSWER: A

82. _____ helps to integrate, maintain and view the contents of the data warehousing system.

A. Business directory.
B. Information directory.
C. Data dictionary.
D. Database.
ANSWER: B

83. Discovery of cross-sales opportunities is called _____.
A. segmentation.
B. visualization.
C. correction.
D. association.
ANSWER: D

84. Data marts that incorporate data mining tools to extract sets of data are called _____.
A. independent data mart.
B. dependent data marts.
C. intra-entry data mart.
D. inter-entry data mart.
ANSWER: B

85. _____ can generate programs itself, enabling it to carry out new tasks.
A. Automated system.
B. Decision making system.
C. Self-learning system.
D. Productivity system.
ANSWER: D

86. The power of self-learning system lies in _____.
A. cost.
B. speed.
C. accuracy.
D. simplicity.
ANSWER: C

87. Building the informational database is done with the help of _____.
A. transformation or propagation tools.
B. transformation tools only.
C. propagation tools only.
D. extraction tools.
ANSWER: A

88. How many components are there in a data warehouse?
A. two.
B. three.
C. four.
D. five.
ANSWER: D

89. Which of the following is not a component of a data warehouse?
A. Metadata.
B. Current detail data.
C. Lightly summarized data.
D. Component Key.
ANSWER: D

90. _____ is data that is distilled from the low level of detail found at the current detailed leve.
A. Highly summarized data.
B. Lightly summarized data.
C. Metadata.
D. Older detail data.
ANSWER: B

91. Highly summarized data is _____.
A. compact and easily accessible.
B. compact and expensive.
C. compact and hardly accessible.
D. compact.
ANSWER: A

92. A directory to help the DSS analyst locate the contents of the data warehouse is seen in _____.
A. Current detail data.
B. Lightly summarized data.
C. Metadata.
D. Older detail data.
ANSWER: C

93. Metadata contains atleast _____.
A. the structure of the data.
B. the algorithms used for summarization.
C. the mapping from the operational environment to the data warehouse.
D. all of the above.
ANSWER: D

94. Which of the following is not a old detail storage medium?
A. Phot Optical Storage.
B. RAID.
C. Microfinche.
D. Pen drive.
ANSWER: D

95. The data from the operational environment enter _____ of data warehouse.
A. Current detail data.
B. Older detail data.
C. Lightly summarized data.
D. Highly summarized data.
ANSWER: A

96. The data in current detail level resides till _____ event occurs.
A. purge.
B. summarization.
C. archieved.
D. all of the above.
ANSWER: D

97. The dimension tables describe the _____.
A. entities.
B. facts.
C. keys.

D. units of measures.
ANSWER: B

98. The granularity of the fact is the _____ of detail at which it is recorded.
A. transformation.
B. summarization.
C. level.
D. transformation and summarization.
ANSWER: C

99. Which of the following is not a primary grain in analytical modeling?
A. Transaction.
B. Periodic snapshot.
C. Accumulating snapshot.
D. All of the above.
ANSWER: B

100. Granularity is determined by _____.
A. number of parts to a key.
B. granularity of those parts.
C. both A and B.
D. none of the above.
ANSWER: C

101. _____ of data means that the attributes within a given entity are fully dependent on the entire primary key of the entity.
A. Additivity.
B. Granularity.
C. Functional dependency.
D. Dimensionality.
ANSWER: C

102. A fact is said to be fully additive if _____.
A. it is additive over every dimension of its dimensionality.
B. additive over atleast one but not all of the dimensions.
C. not additive over any dimension.
D. None of the above.
ANSWER: A

103. A fact is said to be partially additive if _____.
A. it is additive over every dimension of its dimensionality.
B. additive over atleast one but not all of the dimensions.
C. not additive over any dimension.
D. None of the above.
ANSWER: B

104. A fact is said to be non-additive if _____.
A. it is additive over every dimension of its dimensionality.
B. additive over atleast one but not all of the dimensions.
C. not additive over any dimension.
D. None of the above.
ANSWER: C

105. Non-additive measures can often combined with additive measures to create new _____.

A. additive measures.
B. non-additive measures.
C. partially additive.
D. All of the above.
ANSWER: A

106. A fact representing cumulative sales units over a day at a store for a product is a _____.
A. additive fact.
B. fully additive fact.
C. partially additive fact.
D. non-additive fact.
ANSWER: B

107. _____ of data means that the attributes within a given entity are fully dependent on the entire primary key of the entity.
A. Additivity.
B. Granularity.
C. Functional Dependency.
D. Dependency.
ANSWER: C

108. Which of the following is the other name of Data mining?
A. Exploratory data analysis.
B. Data driven discovery.
C. Deductive learning.
D. All of the above.
ANSWER: D

109. Which of the following is a predictive model?
A. Clustering.
B. Regression.
C. Summarization.
D. Association rules.
ANSWER: B

110. Which of the following is a descriptive model?
A. Classification.
B. Regression.
C. Sequence discovery.
D. Association rules.
ANSWER: C

111. A _____ model identifies patterns or relationships.
A. Descriptive.
B. Predictive.
C. Regression.
D. Time series analysis.
ANSWER: A

112. A predictive model makes use of _____.
A. current data.
B. historical data.
C. both current and historical data.
D. assumptions.

ANSWER: B

113. _____ maps data into predefined groups.
A. Regression.
B. Time series analysis
C. Prediction.
D. Classification.
ANSWER: D

114. _____ is used to map a data item to a real valued prediction variable.
A. Regression.
B. Time series analysis.
C. Prediction.
D. Classification.
ANSWER: B

115. In _____, the value of an attribute is examined as it varies over time.
A. Regression.
B. Time series analysis.
C. Sequence discovery.
D. Prediction.
ANSWER: B

116. In _____ the groups are not predefined.
A. Association rules.
B. Summarization.
C. Clustering.
D. Prediction.
ANSWER: C

117. Link Analysis is otherwise called as _____.
A. affinity analysis.
B. association rules.
C. both A & B.
D. Prediction.
ANSWER: C

118. _____ is a the input to KDD.
A. Data.
B. Information.
C. Query.
D. Process.
ANSWER: A

119. The output of KDD is _____.
A. Data.
B. Information.
C. Query.
D. Useful information.
ANSWER: D

120. The KDD process consists of _____ steps.
A. three.
B. four.

C. five.
D. six.
ANSWER: C

121. Treating incorrect or missing data is called as _____.
A. selection.
B. preprocessing.
C. transformation.
D. interpretation.
ANSWER: B

122. Converting data from different sources into a common format for processing is called as _____.
A. selection.
B. preprocessing.
C. transformation.
D. interpretation.
ANSWER: C

123. Various visualization techniques are used in _____ step of KDD.
A. selection.
B. transformaion.
C. data mining.
D. interpretation.
ANSWER: D

124. Extreme values that occur infrequently are called as _____.
A. outliers.
B. rare values.
C. dimensionality reduction.
D. All of the above.
ANSWER: A

125. Box plot and scatter diagram techniques are _____.
A. Graphical.
B. Geometric.
C. Icon-based.
D. Pixel-based.
ANSWER: B

126. _____ is used to proceed from very specific knowledge to more general information.
A. Induction.
B. Compression.
C. Approximation.
D. Substitution.
ANSWER: A

127. Describing some characteristics of a set of data by a general model is viewed as _____
A. Induction.
B. Compression.
C. Approximation.
D. Summarization.
ANSWER: B

128. _____ helps to uncover hidden information about the data.

A. Induction.
B. Compression.
C. Approximation.
D. Summarization.
ANSWER: C

129. _____ are needed to identify training data and desired results.
A. Programmers.
B. Designers.
C. Users.
D. Administrators.
ANSWER: C

130. Overfitting occurs when a model _____.
A. does fit in future states.
B. does not fit in future states.
C. does fit in current state.
D. does not fit in current state.
ANSWER: B

131. The problem of dimensionality curse involves _____.
A. the use of some attributes may interfere with the correct completion of a data mining task.
B. the use of some attributes may simply increase the overall complexity.
C. some may decrease the efficiency of the algorithm.
D. All of the above.
ANSWER: D

132. Incorrect or invalid data is known as _____.
A. changing data.
B. noisy data.
C. outliers.
D. missing data.
ANSWER: B

133. ROI is an acronym of _____.
A. Return on Investment.
B. Return on Information.
C. Repetition of Information.
D. Runtime of Instruction
ANSWER: A

134. The _____ of data could result in the disclosure of information that is deemed to be confidential.
A. authorized use.
B. unauthorized use.
C. authenticated use.
D. unauthenticated use.
ANSWER: B

135. _____ data are noisy and have many missing attribute values.
A. Preprocessed.
B. Cleaned.
C. Real-world.
D. Transformed.

ANSWER: C

136. The rise of DBMS occurred in early _____.
A. 1950's.
B. 1960's
C. 1970's
D. 1980's.
ANSWER: C

137. SQL stand for _____.
A. Standard Query Language.
B. Structured Query Language.
C. Standard Quick List.
D. Structured Query list.
ANSWER: B

138. Which of the following is not a data mining metric?
A. Space complexity.
B. Time complexity.
C. ROI.
D. All of the above.
ANSWER: D

139. Reducing the number of attributes to solve the high dimensionality problem is called as _____.
A. dimensionality curse.
B. dimensionality reduction.
C. cleaning.
D. Overfitting.
ANSWER: B

140. Data that are not of interest to the data mining task is called as _____.
A. missing data.
B. changing data.
C. irrelevant data.
D. noisy data.
ANSWER: C

141. _____ are effective tools to attack the scalability problem.
A. Sampling.
B. Parallelization
C. Both A & B.
D. None of the above.
ANSWER: C

142. Market-basket problem was formulated by _____.
A. Agrawal et al.
B. Steve et al.
C. Toda et al.
D. Simon et al.
ANSWER: A

143. Data mining helps in _____.
A. inventory management.
B. sales promotion strategies.

C. marketing strategies.
D. All of the above.
ANSWER: D

144. The proportion of transaction supporting X in T is called _____.
A. confidence.
B. support.
C. support count.
D. All of the above.
ANSWER: B

145. The absolute number of transactions supporting X in T is called _____.
A. confidence.
B. support.
C. support count.
D. None of the above.
ANSWER: C

146. The value that says that transactions in D that support X also support Y is called _____.
A. confidence.
B. support.
C. support count.
D. None of the above.
ANSWER: A

147. If T consist of 500000 transactions, 20000 transaction contain bread, 30000 transaction contain jam, 10000 transaction contain both bread and jam. Then the support of bread and jam is _____.
A. 2%
B. 20%
C. 3%
D. 30%
ANSWER: A

148. 7 If T consist of 500000 transactions, 20000 transaction contain bread, 30000 transaction contain jam, 10000 transaction contain both bread and jam. Then the confidence of buying bread with jam is _____.
A. 33.33%
B. 66.66%
C. 45%
D. 50%
ANSWER: D

149. The left hand side of an association rule is called _____.
A. consequent.
B. onset.
C. antecedent.
D. precedent.
ANSWER: C

150. The right hand side of an association rule is called _____.
A. consequent.
B. onset.
C. antecedent.
D. precedent.

ANSWER: A

151. Which of the following is not a desirable feature of any efficient algorithm?
A. to reduce number of input operations.
B. to reduce number of output operations.
C. to be efficient in computing.
D. to have maximal code length.
ANSWER: D

152. All set of items whose support is greater than the user-specified minimum support are called as _____.
A. border set.
B. frequent set.
C. maximal frequent set.
D. lattice.
ANSWER: B

153. If a set is a frequent set and no superset of this set is a frequent set, then it is called _____.
A. maximal frequent set.
B. border set.
C. lattice.
D. infrequent sets.
ANSWER: A

154. Any subset of a frequent set is a frequent set. This is _____.
A. Upward closure property.
B. Downward closure property.
C. Maximal frequent set.
D. Border set.
ANSWER: B

155. Any superset of an infrequent set is an infrequent set. This is _____.
A. Maximal frequent set.
B. Border set.
C. Upward closure property.
D. Downward closure property.
ANSWER: C

156. If an itemset is not a frequent set and no superset of this is a frequent set, then it is _____.
A. Maximal frequent set
B. Border set.
C. Upward closure property.
D. Downward closure property.
ANSWER: B

157. A priori algorithm is otherwise called as _____.
A. width-wise algorithm.
B. level-wise algorithm.
C. pincer-search algorithm.
D. FP growth algorithm.
ANSWER: B

158. The A Priori algorithm is a _____.
A. top-down search.

B. breadth first search.
C. depth first search.
D. bottom-up search.
ANSWER: D

159. The first phase of A Priori algorithm is _____.
A. Candidate generation.
B. Itemset generation.
C. Pruning.
D. Partitioning.
ANSWER: A

160. The second phaase of A Priori algorithm is _____.
A. Candidate generation.
B. Itemset generation.
C. Pruning.
D. Partitioning.
ANSWER: C

161. The _____ step eliminates the extensions of (k-1)-itemsets which are not found to be frequent, from being considered for counting support.
A. Candidate generation.
B. Pruning.
C. Partitioning.
D. Itemset eliminations.
ANSWER: B

162. The a priori frequent itemset discovery algorithm moves _____ in the lattice.
A. upward.
B. downward.
C. breadthwise.
D. both upward and downward.
ANSWER: A

163. After the pruning of a priori algorithm, _____ will remain.
A. Only candidate set.
B. No candidate set.
C. Only border set.
D. No border set.
ANSWER: B

164. The number of iterations in a priori _____.
A. increases with the size of the maximum frequent set.
B. decreases with increase in size of the maximum frequent set.
C. increases with the size of the data.
D. decreases with the increase in size of the data.
ANSWER: A

165. MFCS is the acronym of _____.
A. Maximum Frequency Control Set.
B. Minimal Frequency Control Set.
C. Maximal Frequent Candidate Set.
D. Minimal Frequent Candidate Set.
ANSWER: C

166. Dynamuc Itemset Counting Algorithm was proposed by ____.
A. Bin et al.
B. Argawal et at.
C. Toda et al.
D. Simon et at.
ANSWER: A

167. Itemsets in the _____ category of structures have a counter and the stop number with them.
A. Dashed.
B. Circle.
C. Box.
D. Solid.
ANSWER: A

168. The itemsets in the _____category structures are not subjected to any counting.
A. Dashes.
B. Box.
C. Solid.
D. Circle.
ANSWER: C

169. Certain itemsets in the dashed circle whose support count reach support value during an iteration move into the _____.
A. Dashed box.
B. Solid circle.
C. Solid box.
D. None of the above.
ANSWER: A

170. Certain itemsets enter afresh into the system and get into the _____, which are essentially the supersets of the itemsets that move from the dashed circle to the dashed box.
A. Dashed box.
B. Solid circle.
C. Solid box.
D. Dashed circle.
ANSWER: D

171. The itemsets that have completed on full pass move from dashed circle to _____.
A. Dashed box.
B. Solid circle.
C. Solid box.
D. None of the above.
ANSWER: B

172. The FP-growth algorithm has _____ phases.
A. one.
B. two.
C. three.
D. four.
ANSWER: B

173. A frequent pattern tree is a tree structure consisting of _____.
A. an item-prefix-tree.

B. a frequent-item-header table.
C. a frequent-item-node.
D. both A & B.
ANSWER: D

174. The non-root node of item-prefix-tree consists of _____ fields.
A. two.
B. three.
C. four.
D. five.
ANSWER: B

175. The frequent-item-header-table consists of _____ fields.
A. only one.
B. two.
C. three.
D. four.
ANSWER: B

176. The paths from root node to the nodes labelled 'a' are called _____.
A. transformed prefix path.
B. suffix subpath.
C. transformed suffix path.
D. prefix subpath.
ANSWER: D

177. The transformed prefix paths of a node 'a' form a truncated database of pattern which co-occur with a is called _____.
A. suffix path.
B. FP-tree.
C. conditional pattern base.
D. prefix path.
ANSWER: C

178. The goal of _____ is to discover both the dense and sparse regions of a data set.
A. Association rule.
B. Classification.
C. Clustering.
D. Genetic Algorithm.
ANSWER: C

179. Which of the following is a clustering algorithm?
A. A priori.
B. CLARA.
C. Pincer-Search.
D. FP-growth.
ANSWER: B

180. _____ clustering technique start with as many clusters as there are records, with each cluster having only one record.
A. Agglomerative.
B. divisive.
C. Partition.
D. Numeric.

ANSWER: A

181. _____ clustering techniques starts with all records in one cluster and then try to split that cluster into small pieces.
A. Agglomerative.
B. Divisive.
C. Partition.
D. Numeric.
ANSWER: B

182. Which of the following is a data set in the popular UCI machine-learning repository?
A. CLARA.
B. CACTUS.
C. STIRR.
D. MUSHROOM.
ANSWER: D

183. In _____ algorithm each cluster is represented by the center of gravity of the cluster.
A. k-medoid.
B. k-means.
C. STIRR.
D. ROCK.
ANSWER: B

184. In _____ each cluster is represented by one of the objects of the cluster located near the center.
A. k-medoid.
B. k-means.
C. STIRR.
D. ROCK.
ANSWER: A

185. Pick out a k-medoid algoithm.
A. DBSCAN.
B. BIRCH.
C. PAM.
D. CURE.
ANSWER: C

186. Pick out a hierarchical clustering algorithm.
A. DBSCAN
B. BIRCH.
C. PAM.
D. CURE.
ANSWER: A

187. CLARANS stands for _____.
A. CLARA Net Server.
B. Clustering Large Application RAnge Network Search.
C. Clustering Large Applications based on RANdomized Search.
D. CLustering Application Randomized Search.
ANSWER: C

188. BIRCH is a _____.

A. agglomerative clustering algorithm.
B. hierarchical algorithm.
C. hierarchical-agglomerative algorithm.
D. divisive.
ANSWER: C

189. The cluster features of different subclusters are maintained in a tree called _____.
A. CF tree.
B. FP tree.
C. FP growth tree.
D. B tree.
ANSWER: A

190. The _____ algorithm is based on the observation that the frequent sets are normally very few in number compared to the set of all itemsets.
A. A priori.
B. Clustering.
C. Association rule.
D. Partition.
ANSWER: D

191. The partition algorithm uses _____ scans of the databases to discover all frequent sets.
A. two.
B. four.
C. six.
D. eight.
ANSWER: A

192. The basic idea of the apriori algorithm is to generate_____ item sets of a particular size & scans the database.
A. candidate.
B. primary.
C. secondary.
D. superkey.
ANSWER: A

193. _____is the most well known association rule algorithm and is used in most commercial products.
A. Apriori algorithm.
B. Partition algorithm.
C. Distributed algorithm.
D. Pincer-search algorithm.
ANSWER: A

194. An algorithm called_____is used to generate the candidate item sets for each pass after the first.
A. apriori.
B. apriori-gen.
C. sampling.
D. partition.
ANSWER: B

195. The basic partition algorithm reduces the number of database scans to _____ & divides it into partitions.

A. one.
B. two.
C. three.
D. four.
ANSWER: B

196. _____and prediction may be viewed as types of classification.
A. Decision.
B. Verification.
C. Estimation.
D. Illustration.
ANSWER: C

197. _____can be thought of as classifying an attribute value into one of a set of possible classes.
A. Estimation.
B. Prediction.
C. Identification.
D. Clarification.
ANSWER: B

198. Prediction can be viewed as forecasting a_____value.
A. non-continuous.
B. constant.
C. continuous.
D. variable.
ANSWER: C

199. _____data consists of sample input data as well as the classification assignment for the data.
A. Missing.
B. Measuring.
C. Non-training.
D. Training.
ANSWER: D

200. Rule based classification algorithms generate _____ rule to perform the classification.
A. if-then.
B. while.
C. do while.
D. switch.
ANSWER: A

201. _____ are a different paradigm for computing which draws its inspiration from neuroscience.
A. Computer networks.
B. Neural networks.
C. Mobile networks.
D. Artificial networks.
ANSWER: B

202. The human brain consists of a network of _____.
A. neurons.
B. cells.
C. Tissue.

D. muscles.
ANSWER: A

203. Each neuron is made up of a number of nerve fibres called _____.
A. electrons.
B. molecules.
C. atoms.
D. dendrites.
ANSWER: D

204. The _____is a long, single fibre that originates from the cell body.
A. axon.
B. neuron.
C. dendrites.
D. strands.
ANSWER: A

205. A single axon makes _____ of synapses with other neurons.
A. ones.
B. hundreds.
C. thousands.
D. millions.
ANSWER: C

206. _____ is a complex chemical process in neural networks.
A. Receiving process.
B. Sending process.
C. Transmission process.
D. Switching process.
ANSWER: C

207. _____ is the connectivity of the neuron that give simple devices their real power. a. b. c. d.
A. Water.
B. Air.
C. Power.
D. Fire.
ANSWER: D

208. _____ are highly simplified models of biological neurons.
A. Artificial neurons.
B. Computational neurons.
C. Biological neurons.
D. Technological neurons.
ANSWER: A

209. The biological neuron's _____ is a continuous function rather than a step function.
A. read.
B. write.
C. output.
D. input.
ANSWER: C

210. The threshold function is replaced by continuous functions called _____ functions.
A. activation.

B. deactivation.
C. dynamic.
D. standard.
ANSWER: A

211. The sigmoid function also knows as _____functions.
A. regression.
B. logistic.
C. probability.
D. neural.
ANSWER: B

212. MLP stands for _____.
A. mono layer perception.
B. many layer perception.
C. more layer perception.
D. multi layer perception.
ANSWER: D

213. In a feed- forward networks, the conncetions between layers are _____ from input to output.
A. bidirectional.
B. unidirectional.
C. multidirectional.
D. directional.
ANSWER: B

214. The network topology is constrained to be _____.
A. feedforward.
B. feedbackward.
C. feed free.
D. feed busy.
ANSWER: A

215. RBF stands for _____.
A. Radial basis function.
B. Radial bio function.
C. Radial big function.
D. Radial bi function.
ANSWER: A

216. RBF have only _____ hidden layer.
A. four.
B. three.
C. two.
D. one.
ANSWER: D

217. RBF hidden layer units have a receptive field which has a _____; that is, a particular input value at which they have a maximal output.
A. top.
B. bottom.
C. centre.
D. border.

ANSWER: C

218. _____ training may be used when a clear link between input data sets and target output values does not exist.
A. Competitive.
B. Perception.
C. Supervised.
D. Unsupervised.
ANSWER: D

219. _____ employs the supervised mode of learning.
A. RBF.
B. MLP.
C. MLP & RBF.
D. ANN.
ANSWER: C

220. _____ design involves deciding on their centres and the sharpness of their Gaussians.
A. DR.
B. AND.
C. XOR.
D. RBF.
ANSWER: D

221. _____ is the most widely applied neural network technique.
A. ABC.
B. PLM.
C. LMP.
D. MLP.
ANSWER: D

222. SOM is an acronym of _____.
A. self-organizing map.
B. self origin map.
C. single organizing map.
D. simple origin map.
ANSWER: A

223. _____ is one of the most popular models in the unsupervised framework.
A. SOM.
B. SAM.
C. OSM.
D. MSO.
ANSWER: A

224. The actual amount of reduction at each learning step may be guided by _____.
A. learning cost.
B. learning level.
C. learning rate.
D. learning time.
ANSWER: C

225. The SOM was a neural network model developed by _____.
A. Simon King.

B. Teuvokohonen.
C. Tomoki Toda.
D. Julia.
ANSWER: B

226. SOM was developed during _____.
A. 1970-80.
B. 1980-90.
C. 1990 -60.
D. 1979 -82.
ANSWER: D

227. Investment analysis used in neural networks is to predict the movement of _____ from previous data.
A. engines.
B. stock.
C. patterns.
D. models.
ANSWER: B

228. SOMs are used to cluster a specific _____ dataset containing information about the patient's drugs etc.
A. physical.
B. logical.
C. medical.
D. technical.
ANSWER: C

229. GA stands for _____.
A. Genetic algorithm
B. Gene algorithm.
C. General algorithm.
D. Geo algorithm.
ANSWER: A

230. GA was introduced in the year _____.
A. 1955.
B. 1965.
C. 1975.
D. 1985.
ANSWER: C

231. Genetic algorithms are search algorithms based on the mechanics of natural_____.
A. systems.
B. genetics.
C. logistics.
D. statistics.
ANSWER: B

232. GAs were developed in the early _____.
A. 1970.
B. 1960.
C. 1950.
D. 1940.

ANSWER: A

233. The RSES system was developed in _____.
A. Poland.
B. Italy.
C. England.
D. America.
ANSWER: A

234. Crossover is used to _____.
A. recombine the population's genetic material.
B. introduce new genetic structures in the population.
C. to modify the population's genetic material.
D. All of the above.
ANSWER: A

235. The mutation operator _____.
A. recombine the population's genetic material.
B. introduce new genetic structures in the population.
C. to modify the population's genetic material.
D. All of the above.
ANSWER: B

236. Which of the following is an operation in genetic algorithm?
A. Inversion.
B. Dominance.
C. Genetic edge recombination.
D. All of the above.
ANSWER: D

237. . _____ is a system created for rule induction.
A. RBS.
B. CBS.
C. DBS.
D. LERS.
ANSWER: D

238. NLP stands for _____.
A. Non Language Process.
B. Nature Level Program.
C. Natural Language Page.
D. Natural Language Processing.
ANSWER: D

239. Web content mining describes the discovery of useful information from the _____contents.
A. text.
B. web.
C. page.
D. level.
ANSWER: B

240. Research on mining multi-types of data is termed as _____ data.
A. graphics.
B. multimedia.

C. meta.
D. digital.
ANSWER: B

241. _____ mining is concerned with discovering the model underlying the link structures of the web.
A. Data structure.
B. Web structure.
C. Text structure.
D. Image structure.
ANSWER: B

242. _____ is the way of studying the web link structure.
A. Computer network.
B. Physical network.
C. Social network.
D. Logical network.
ANSWER: C

243. The _____ propose a measure of standing a node based on path counting.
A. open web.
B. close web.
C. link web.
D. hidden web.
ANSWER: B

244. In web mining, _____ is used to find natural groupings of users, pages, etc.
A. clustering.
B. associations.
C. sequential analysis.
D. classification.
ANSWER: A

245. In web mining, _____ is used to know the order in which URLs tend to be accessed.
A. clustering.
B. associations.
C. sequential analysis.
D. classification.
ANSWER: C

246. In web mining, _____ is used to know which URLs tend to be requested together.
A. clustering.
B. associations.
C. sequential analysis.
D. classification.
ANSWER: B

247. _____ describes the discovery of useful information from the web contents.
A. Web content mining.
B. Web structure mining.
C. Web usage mining.
D. All of the above.
ANSWER: A

248. _____ is concerned with discovering the model underlying the link structures of the web.

A. Web content mining.
B. Web structure mining.
C. Web usage mining.
D. All of the above.
ANSWER: B

249. A link is said to be _____ link if it is between pages with different domain names.
A. intrinsic.
B. transverse.
C. direct.
D. contrast.
ANSWER: B

250. A link is said to be _____ link if it is between pages with the same domain name.
A. intrinsic.
B. transverse.
C. direct.
D. contrast.
ANSWER: A

Staff Name
LAXMI.SREE.B.R.