Semester VI
CORE- 26 Elective (A) : Data Mining
Multiple Choice Questions.

1. 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

2. SQL helps to find _____.
   A. the interesting data
   B. hidden information
   C. intermediate data
   D. data under constraints that are already known
   ANSWER: D

3. Translation of problem to learning technique is called as _____.
   A. reengineering.
   B. translational engineering.
   C. representational engineering.
   D. learning algorithm.
   ANSWER: C

4. 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

5. _____and _____ are the important qualities of good learning algorithm.
   A. Consistent, Complete.
   B. Information content, Complex.
   C. Complete, Complex.
   D. Transparent, Complex.
   ANSWER: A

6. Redundancy refers to the elements of a message that can be derived from other parts of _____.
   A. different message.
   B. irrelevant message.
   C. same message.
   D. complete message.
   ANSWER: C

7. Metadata describes _____.
   A. contents of database.
   B. structure of contents of database.
   C. structure of database.

D. database itself.
ANSWER: B

8. The partition of overall data warehouse is _____.
   A. database.
   B. data cube.
   C. data mart.
   D. operational data.
   ANSWER: C

9. _____ is used to load the information from operational database.
   A. Replication technique.
   B. Reengineering technique.
   C. Engineering technique.
   D. Transformation engineering.
   ANSWER: A

10. _____ multiprocessing machines share same hard disk and internal memory.
   A. Massively parallel.
   B. Symmetric.
   C. Parallel.
   D. Asymmetric.
   ANSWER: B

11. A trivial result that is obtained by an extremely simple method is called _____.
   A. naive prediction.
   B. accurate prediction.
   C. correct prediction.
   D. wrong prediction.
   ANSWER: A

12. 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

13. OLAP stands for _____.
   A. Online Analytical Processing.
   B. Online Linear Analytical Processing.
   C. Online Animated Process.
   D. Online Analytical Problem.
   ANSWER: A

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

15. The intermediate unit in perceptron is _____.
   A. photoreceptors.
   B. associators.
   C. responders.
   D. receptors.
   ANSWER: B

16. OLAP is used to explore the _____ knowledge.

A. shallow.
B. deep.
C. multidimensional.
D. hidden.
  ANSWER: C

17. A natural way to visualize the process of training a self-organizing map is called _____.
    A. kohonen movie.
    B. kohonen map.
    C. frame.
    D. scatter diagram.
  ANSWER: A

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

19. Deep knowledge can be found only by using _____.
    A. clues.
    B. OLAP.
    C. SQL.
    D. algorithm
  ANSWER: A

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

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

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

23. In _____ approach data ware house is build first and all information needed is selected.
    A. top-down.
    B. client/server.
    C. bottom-up.
    D. DSS.
  ANSWER: A

24. The DB vendor who is able to operate massively parallel computers is _____.
    A. TCS.
    B. IBM.
    C. CTS.
    D. Wipro.

ANSWER: B

25. 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

26. _____ is a creative activity that has to be performed repeatedly in order to get best results.
    A. Cleaning
    B. Reporting
    C. Coding.
    D. Selection.
   ANSWER: C

27. _____ is an example for case based-learning.
    A. Decision trees.
    B. Neural networks.
    C. Genetic algorithm.
    D. K-nearest neighbor.
   ANSWER: D

28. _____ percentage of the interesting information can be obtained by using SQL.
    A. 80
    B. 70
    C. 40
    D. 50
   ANSWER: A

29. _____ 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

30. In K-nearest neighbor algorithm K stands for _____.
    A. number of neighbors that are investigated.
    B. number of iterations.
    C. number of total records.
    D. random number.
   ANSWER: A

31. The complexity of data mining algorithm is represented by _____.
    A. log n.
    B. 2n log n.
    C. n log n.
    D. 2 log n.
   ANSWER: C

32. Genetic algorithm was proposed by _____.
    A. John Holland.
    B. Johnson.
    C. Watson.
    D. Kohenon.
   ANSWER: A

33. _____ is the first stage in genetic algorithm.
    A. Evaluation of each string.

B. Selection of string.

C. Creation of population of string.

D. Genetic manipulation.

ANSWER: C

34. The _____ is one of genetic operators that are used to recombine the population of genetic material.

A. genetic operator.

B. mutation operator.

C. cross over operator.

D. encoding operator.

ANSWER: A

35. _____ is the heart of knowledge discovery in database process.

A. Selection.

B. Data ware house.

C. Data mining.

D. Creative coding.

ANSWER: D

36. _____ is a planning optimization application written for KLM

A. PILOTS.

B. CAPTAINS.

C. CUSTOMERS.

D. AIRLINES.

ANSWER: B

37. EIS stands for _____.

A. Executive Information System.

B. Exchange of Information System.

C. Extra Information System.

D. Extended Information system.

ANSWER: A

38. Foreign key constraints are also referred as _____.

A. consistency constraints.

B. referential integrity.

C. conditional integrity.

D. domain constraints.

ANSWER: B

39. The set of attribute in a database that refers to data in another table is called _____.

A. primary key.

B. candidate key.

C. foreign key.

D. super key.

ANSWER: C

40. The distance between two points that is calculated using Pythagoras theorem is _____.

A. cartesian distance.

B. eucledian distance.

C. extendable distance.

D. heuristic distance.

ANSWER: B

41. A database containing volatile data used for daily operation of an organization is _____.

A. historic data.

B. metadata.

C. knowledge.

D. operational data.

ANSWER: D

42. The system that can be used without knowledge of internal operation _____.
    A. black box.
    B. white box.
    C. case based learning.
    D. bias.
   ANSWER: A

43. _____ is the relationship between compressibility and learnability.
    A. Maximum description length principle.
    B. Minimum description length principle.
    C. Kolmogorov complexity.
    D. Voronoi principle.
   ANSWER: B

44. 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

45. DSS stands for _____.
    A. Deciding Support System.
    B. Decision Support System.
    C. Decision Software System.
    D. Decision System of System.
   ANSWER: B

46. Data mining algorithms require _____
    A. efficient sampling method.
    B. storage of intermediate results.
    C. capacity to handle large amounts of data.
    D. All of the above.
   ANSWER: D

47. The algorithm that need to access a table several times during execution is_____.
    A. n-table scan algorithm.
    B. zoom scan algorithm.
    C. hybrid algorithm.
    D. nearest neighbor search.
   ANSWER: A

48. A coding operation in which an attribute with cardinality n is replaced by n binary attributes is called as _____.
    A. falsification of table.
    B. enrichment of table.
    C. flattening of table.
    D. fuzzification of table.
   ANSWER: C

49. The un-normalized relation containing all attributes that exist in database is _____.
    A. actual relation.
    B. transparent relation.
    C. verified relation.
    D. universal relation.
   ANSWER: D

50. The technique of learning by generalizing from examples is _____.
    A. incremental learning.
    B. inductive learning.

   C. hybrid learning.

   D. generalized learning.

   ANSWER: B

51. The ever increasing amount of data is compared to that of infinite library by Jorge Louis Borges in his short stories namely _____.

   A. the library of Louis.

   B. the library of Borges.

   C. the library of Babel.

   D. the library of Boulevard.

   ANSWER: C

52. _____ itself has become a production factor of importance.

   A. Data.

   B. Information.

   C. Program.

   D. Algorithm.

   ANSWER: B

53. The _____ plays an important role in artificial intelligence.

   A. programming skill.

   B. scheduling.

   C. planning.

   D. learning capabilities.

   ANSWER: D

54. Knowledge discovery in database refers to _____.

   A. whole process of extraction of knowledge from data.

   B. selection of data.

   C. coding.

   D. cleaning the data.

   ANSWER: A

55. Data mining is used to refer _____ stage in knowledge discovery in database.

   A. selection.

   B. retrieving.

   C. discovery.

   D. coding.

   ANSWER: C

56. Query tools and data mining tools are _____.

   A. same.

   B. different.

   C. complementary.

   D. standard.

   ANSWER: C

57. In genetic algorithm the problem is considered in terms of _____.

   A. values.

   B. points in multidimensional space.

   C. node.

   D. strings of characters.

   ANSWER: D

58. In UK,_____ has applied data mining techniques to analyze viewing figures. a. a press .

   A. press

   B. BBC

   C. CNN

   D. NDT

   ANSWER: B

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

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

61. _____refers to the process of deriving high-quality information from text.
   A. Text Mining.
   B. Image Mining.
   C. Database Mining.
   D. Multimedia Mining.
   ANSWER: A

62. The process of selecting good hypothesis and improving the theory based on this is called _____.
   A. heuristic search
   B. hill climbing algorithm.
   C. incremental search.
   D. apriori algorithm
   ANSWER: B

63. _____ is the application of data mining techniques to discover patterns from the Web.
   A. Text Mining.
   B. Multimedia Mining.
   C. Web Mining.
   D. Link Mining.
   ANSWER: C

64. It is important to know the complexity of the _____ before developing any machine learning algorithm.
   A. data
   B. algorithm
   C. search space
   D. learning
   ANSWER: C

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

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

67. 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

68. 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

69. A _____ is necessary condition for KDDs effective implement.

A. data set.

B. database.

C. data warehouse.

D. data.

ANSWER: C

70. The first international KDD conference was held in the year _____.

A. 1995.

B. 1994.

C. 1993.

D. 1992.

ANSWER: A

71. AI stands for _____.

A. art of interest.

B. artificial interest.

C. art of intelligence.

D. artificial intelligence.

ANSWER: D

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

73. _____ could generate rule automatically.

A. KDD.

B. machine learning.

C. artificial intelligence.

D. expert system.

ANSWER: B

74. Intelligent miner is a mining tool from _____.

A. Clementine.

B. living stones.

C. IBM.

D. Wipro.

ANSWER: C

75. The organization such as _____ is in USA.

A. AT & T.

B. AD & T.

C. AA & T.

D. AT & D.

ANSWER: A

76. _____ is a mining tool from integral solutions.
   A. WEKA
   B. web miner.
   C. rapid miner.
   D. clementine.
   ANSWER: D

77. _____ % of KDD is about preparing data.
   A. 60.
   B. 70
   C. 80
   D. 90
   ANSWER: C

78. The _____ is one of the operation research techniques.
   A. association rules.
   B. k-nearest neighbor.
   C. decision trees.
   D. genetic algorithm.
   ANSWER: B

79. Everything that science discovers has only _____ value.
   A. standard.
   B. different.
   C. same.
   D. temporary.
   ANSWER: D

80. 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

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

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

83. Bias helps to _____.
   A. learn.
   B. complete the search.
   C. execute the search.
   D. constrain the search and utilizes KDD to analyze client files.
   ANSWER: D

84. 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

85. 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

86. The _____ forms the background knowledge in the inductive logic programming.
   A. prolog program.
   B. perl.
   C. python.
   D. ruby.
   ANSWER: A

87. In KDD process _____ % is about mining.
   A. 40.
   B. 30.
   C. 20.
   D. 10.
   ANSWER: C

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

89. A definition of a concept is _____ if it does not classify any negative examples as falling under the concept.
   A. complete.
   B. consistent.
   C. good.
   D. bad.
   ANSWER: B

90. Lot of kangaroo jumping around the country side is an example for _____.
   A. parallelism.
   B. concept learning.
   C. machine learning.
   D. data mining.
   ANSWER: A

91. 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

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

93. Data warehouse is only used for _____.
   A. operating the data.
   B. managing the data.
   C. decision making.
   D. queries.
   ANSWER: D

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

95. 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

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

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

98. The _____ represents the best choice for building a data warehouse.
   A. client/server.
   B. database.
   C. bottom up.
   D. visualization.
   ANSWER: A

99. The _____ is one of database that operates on massively parallel computer.
   A. sybase.
   B. SQL.
   C. postgre SQL.
   D. tandem.
   ANSWER: D

100. _____ is more recent expert system.
   A. Mycin.
   B. Gasoil.
   C. BMT.
   D. XCONVAX.
   ANSWER: B

101. A _____ is not the rule that govern the basic structure of data warehouse.
   A. time dependent.
   B. volatile.
   C. subject oriented.

D. integrated.
ANSWER: B

102. The metadata that is generated at the time of building a warehouse is called _____.
   A. Build time metadata.
   B. Usage metadata.
   C. Control metadata.
   D. structure metadata.
   ANSWER: A

103. The control metadata is used to _____.
   A. design a metadata.
   B. administrate the metadata.
   C. track the sequence and timing of warehouse events.
   D. control the data.
   ANSWER: C

104. A data warehouse is said to contain a time-varying collection of data because ___.
   A. its contents vary automatically with time.
   B. its lifespan is very limited.
   C. it contains historical data.
   D. its content has explicit stamp.
   ANSWER: C

105. A data warehouse is an integrated collection of data because _____.
   A. it is a collection of data of different data types.
   B. it is a collection of data derived from multiple sources.
   C. it is a relational database.
   D. it contains summarized data.
   ANSWER: B

106. Expert systems are _____.
   A. system that contain the knowledge of specialists.
   B. system that can think of their own.
   C. system that can work.
   D. system that can create the knowledge.
   ANSWER: A

107. _____ is an expert who analyzed the effect of using machine learning algorithm in setting up expert system.
   A. Borges.
   B. Popper.
   C. Bratko.
   D. Papert.
   ANSWER: C

108. The element that is not taken into consideration for cost justification for the implementation of KDD environment is _____.
   A. speed.
   B. cost.
   C. complexity.
   D. repetition.
   ANSWER: B

109. A _____ is an interactive system that enables decision makers to use database and models on a computer in order to solve ill structured problems.
   A. database.
   B. client/server.
   C. DSS.
   D. mainframe
   ANSWER: C

110. 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

111. DB/2 is a family of RDBMS marketed by _____.
   A. HCL.
   B. Wipro.
   C. IBM.
   D. Infosys.
   ANSWER: C

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

113. In data mining software that works on local workstation is used to _____.
   A. write coding.
   B. generate screen and reports for the end user.
   C. make decisions.
   D. find patterns.
   ANSWER: B

114. A _____ acts a bridge between data warehouse and database application.
   A. data mart.
   B. operational data.
   C. meta data.
   D. data cube.
   ANSWER: C

115. The _____ operation is used for reducing data cube by one or more dimensions.
   A. drilling.
   B. rolling.
   C. dicing.
   D. slicing.
   ANSWER: D

116. 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

117. KDD consists of _____ stages.
   A. four.
   B. five.
   C. six.
   D. seven.
   ANSWER: C

118. _____ is the first stage in KDD process.
   A. Data selection.
   B. Cleaning.

C. Mining.

D. Enrichment.

ANSWER: A

119. The term that is not associated with data cleaning process is _____.

A. domain consistance.

B. de-duplication.

C. disambiguation.

D. segmentation.

ANSWER: D

120. In _____ process of KDD additional information can be added to the existing data.

A. enrichment.

B. coding.

C. selecting.

D. cleaning.

ANSWER: A

121. _____ is a type of coding operation that occurs frequently in KDD context.

A. Filtering.

B. Visualization.

C. Flattening.

D. Replication.

ANSWER: C

122. SQL stands for _____.

A. simple query language.

B. structured query language.

C. strong query language.

D. simple language.

ANSWER: B

123. _____ is one of the traditional query tool.

A. MYSQL.

B. OLAP.

C. PL/SQL.

D. SQL.

ANSWER: D

124. 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

125. A/An_____ is an object oriented 3D tool kit which enables the user to explore 3D structure.

A. inventor.

B. tandim.

C. mantis.

D. extruder.

ANSWER: A

126. The field of research dedicated to the search for interesting projections of datasets are called _____.

A. projection pursuit.

B. research pursuit.

C. projection.

D. dataset pursuit.

ANSWER: A

127. Which of the following is correct order of empirical cycle of scientific research?
   A. Analysis, observation, prediction, theory.
   B. Analysis, prediction, theory, observation.
   C. Analysis, prediction, observation, theory.
   D. Analysis, observation, theory, prediction.
   ANSWER: B

128. 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

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

130. _____ 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

131. In coding stage of KDD process one should be conscious in _____.
   A. adding the data.
   B. deleting the data.
   C. coding.
   D. discovering patterns.
   ANSWER: B

132. 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

133. 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

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

135. _____ is used to find relationship between multidimensional data.
   A. K-nearest neighbor.
   B. Decision trees.
   C. Association rules.

D. OLAP tools.
ANSWER: D

136. 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

137. Association rules are always defined on_____.
   A. binary attribute.
   B. single attribute.
   C. relational database.
   D. multidimensional attribute.
   ANSWER: A

138. Neural network are modeled on the _____.
   A. neuron.
   B. network.
   C. human brain.
   D. machines.
   ANSWER: C

139. _____ built perceptron.
   A. Kohenon.
   B. Frank Rosenblatt.
   C. Minsley.
   D. Papert.
   ANSWER: B

140. The input unit of perceptron is called as_____.
   A. associators.
   B. responders.
   C. neuron.
   D. photo receptors.
   ANSWER: D

141. The intermediate layers in a back-propagation network consists of_____.
   A. photo receptors.
   B. responders.
   C. hidden nodes.
   D. associators.
   ANSWER: C

142. The area for exchange of views between biology and computer science is called as_____.
   A. evolutionary programming.
   B. evolutionary computing.
   C. evolutionary strategies.
   D. genetic algorithm.
   ANSWER: B

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

144. The _____is a knowledge that can be found by using pattern recognition algorithm.

A. hidden knowledge.
B. deep.
C. shallow.
D. multidimensional.
ANSWER: A

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

146. The performance of _____is better at problem solving techniques
   A. neural network.
   B. classification trees.
   C. decision tree.
   D. genetic knowledge.
ANSWER: D

147. The quality of output depends on _____.
   A. algorithm that is capable of learning rules.
   B. algorithm that can handle large data sets.
   C. algorithm that can learn incrementally.
   D. algorithm that performs well.
ANSWER: A

148. In CAPITANS _____ algorithm is used to predict the pilots behavior.
   A. neural network.
   B. genetic algorithm.
   C. association.
   D. decision tree.
ANSWER: B

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

150. 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

151. 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

152. 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

153. 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

154. 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

155. 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

156. 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

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

158. 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

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

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

161. 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

162. 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

163. 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

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

165. Which of the following table type belongs to snowflake schema
   A. Free.
   B. Dimension.
   C. Double.
   D. Replicator.
   ANSWER: B

166. The goal of data mining is _____.
   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

167. Which of the following information systems are used in the daily running of the business
   A. Operational planning systems.
   B. Transaction processing systems (TPS).
   C. Process control systems.
   D. Office automation systems (OAS).
   ANSWER: A

168. Which of the following is not one of the three Cs describing groupware
   A. Collaboration.
   B. Communication.
   C. Co-ordination.
   D. Compliance.
   ANSWER: D

169. Materials Requirements Planning (MRP) software is an example of an information systems application in which of the following areas
   A. Operations Management.
   B. Accounting.
   C. Marketing.

D. Human Resource.
ANSWER: A

170. Which of the following is not one of the three main components in a decision support system
   A. Model.
   B. Communications.
   C. Data.
   D. Dialogue.
ANSWER: B

171. Business Intelligence and data warehousing is used for which of the following
   A. Forecasting.
   B. Data Mining.
   C. Analysis of large volumes of product sales data.
   D. All of the above.
ANSWER: D

172. Which of the following is true about ERP software
   A. Simplified support and maintenance through a supplier.
   B. Extensive skills available for development.
   C. Supplier costs kept low.
   D. Always adaptable to how the business works.
ANSWER: A

173. Which of the following is an EIS
   A. Electronic information system.
   B. Executive information system.
   C. Extended information system.
   D. Electronic interface system.
ANSWER: B

174. What is the role of document image processing (DIP) systems
   A. Cash flow forecasting.
   B. Inventory control.
   C. Conversion into digital format.
   D. Producing customer invoices.
ANSWER: C

175. Which of the following uses a series of logically related two-dimensional tables or files to store information in the form of a database
   A. Database.
   B. Database management system.
   C. Data warehouse.
   D. None of the above.
ANSWER: D

176. What DBMS component contains facilities to help you develop transaction-intensive applications
   A. DBMS engine.
   B. Data definition subsystem.
   C. Application generation subsystem.
   D. Data administration subsystem.
ANSWER: C

177. Which of the following is a data manipulation tool
   A. File generators.
   B. Query by example tool.
   C. Structure question language.
   D. Structure query language.
ANSWER: B

178. 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

179. Which data administration subsystem periodically backs up information contained in a database
    A. Concurrency control facilities.
    B. Reorganization facilities.
    C. Backup and recovery facilities.
    D. Security management facilities.
    ANSWER: C

180. Which of the following is true of three-tier data warehouses
    A. Once created, the data marts will keep on being updated from the data warehouse at periodic times.
    B. Once created, the data marts will directly receive their new data from the operational databases.
    C. The data marts are different groups of tables in the data warehouse.
    D. A data mart becomes a data warehouse when it reaches a critical size.
    ANSWER: A

181. 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

182. 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

183. 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

184. Which of the following statements is not true about refreshing a data warehouse
    A. It is a process of managing timing differences between the updating of data sources and the related data
warehouse objects.
    B. Updates to dimension tables may occur at different times than the fact table.
    C. The data warehouse administrator has more control over the load time lag than the valid time lag.
    D. None of the above.
    ANSWER: D

185. The _____ is responsible for running queries and reports against data warehouse tables.
    A. hardware.
    B. software.
    C. end users.
    D. middleware.
    ANSWER: C

186. Query tool is meant for _____.
    A. data acquisition.
    B. information delivery.

C. information exchange.

D. communication.

ANSWER: A

187. 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

188. Which of the following clustering analysis method uses multi resolution approach

A. STUNT.

B. OPTICS.

C. CLIQUE.

D. Wave Cluster.

ANSWER: D

189. Which type of following clustering computes augmented cluster ordering

A. OPTICS.

B. CLIQUE.

C. STING.

D. CLUSTER.

ANSWER: A

190. Data compression is to compress the given data by encoding in terms of _____.

A. bytes.

B. bits.

C. cluster.

D. group.

ANSWER: C

191. Which of the following feature is supported by the physical design of data warehouse

A. Support backup and recovery.

B. Planning.

C. Analysis.

D. Coding.

ANSWER: A

192. Classification rules are extracted from_____.

A. root node.

B. decision tree.

C. siblings.

D. branches.

ANSWER: B

193. Dimensionality reduction reduces the data set size by removing _____.

A. relevant attributes.

B. irrelevant attributes.

C. derived attributes.

D. composite attributes.

ANSWER: B

194. _____is a method of incremental conceptual clustering.

A. CORBA.

B. OLAP.

C. COBWEB.

D. STING.

ANSWER: C

195. 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
   ANSWER: A

196. Which of the following are special programs that are stored on database and fired when certain predefined action occurs
   A. Snapshots.
   B. Relations.
   C. Triggers.
   D. Cursors.
   ANSWER: C

197. The _____ refers to extracting knowledge from larger amount of data.
   A. data abstraction.
   B. data warehouse.
   C. database.
   D. data mining.
   ANSWER: D

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

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

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

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

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

203. The _____is a data transformation process.
   A. comparison.
   B. projection.

C. selection.

D. filtering.

ANSWER: D

204. Technology area associated with CRM is _____.

A. specialization.

B. generalization.

C. personalization.

D. summarization.

ANSWER: C

205. SMP stands for _____.

A. symmetric multiprocessor.

B. symmetric multiprogramming.

C. symmetric meta programming.

D. symmetric microprogramming.

ANSWER: A

206. The _____is 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

207. MDDB stands for _____.

A. multiple data doubling.

B. multidimensional databases.

C. multiple double dimension.

D. multi-dimension doubling.

ANSWER: B

208. _____ is data about data.

A. Metadata.

B. Microdata.

C. Minidata

D. Multidata.

ANSWER: A

209. The _____ is an important functional component of the metadata.

A. digital directory.

B. repository.

C. information directory.

D. data dictionary.

ANSWER: C

210. EIS stands for _____.

A. extended interface system.

B. executive interface system.

C. executive information system.

D. extendable information system.

ANSWER: C

211. _____ is data collected from natural systems.

A. MRI scan.

B. ODS data.

C. Statistical data.

D. Historical data.

ANSWER: A

212. Which one of the following feature is provided by an effective data visualization tool
   A. Capability to update data.
   B. Capability to delete or alter data.
   C. Capability in append data.
   D. Capability to compare data
   ANSWER: D

213. The _____ are inexpensive desktop tools designed for end users.
   A. production tools.
   B. report writers.
   C. query tools.
   D. application tools.
   ANSWER: B

214. _____is an example of application development environments.
   A. Visual Basic.
   B. Oracle.
   C. Sybase.
   D. SQL Server.
   ANSWER: A

215. 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

216. _____are some popular OLAP tools.
   A. Metacube, informix.
   B. Oracle express, essbase.
   C. HOLAP.
   D. MOLAP.
   ANSWER: A

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

218. The _____is a process of determining the preference of customers majority.
   A. association.
   B. preferencing.
   C. segmentation.
   D. classification.
   ANSWER: B

219. Which of the following is a process to humanize the mass of data
   A. Visualizing.
   B. Segmentation.
   C. Discovery.
   D. Correction.
   ANSWER: A

220. Strategic value of data mining is_____.
   A. cost-sensitive.
   B. work-sensitive.

C. time-sensitive.
D. technical-sensitive.
  ANSWER: C

221. Which of the following is a data store that is a subsidiary of a data warehouse of integrated data
   A. Data mart.
   B. Data store.
   C. Database.
   D. Data house.
  ANSWER: A

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

223. The terms equality and roll up are associated with_____.
   A. OLAP.
   B. visualization.
   C. data mart.
   D. decision tree.
  ANSWER: C

224. Exceptional reporting in data warehousing is otherwise called as_____.
   A. exception.
   B. alerts.
   C. errors.
   D. bugs.
  ANSWER: B

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

226. Which year was the beginning of new generation of machine learning
   A. 1880.
   B. 1980.
   C. 1970.
   D. 1960.
  ANSWER: C

227. The _____ is an expensive process in building an expert system.
   A. analysis.
   B. study.
   C. design.
   D. information collection.
  ANSWER: D

228. The full form of KDD is_____.
   A. knowledge database.
   B. knowledge discovery in database.
   C. knowledge data house.
   D. knowledge data definition.
  ANSWER: C

229. Which of the following is not a open source data mining tool.
   A. WEKA
   B. R
   C. RapidMiner
   D. KnowledgeMiner
   ANSWER: D

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

231. The _____ 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

232. A/An _____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

233. Discovery of cross-sales opportunities are called_____.
   A. segmentation.
   B. visualization.
   C. correction.
   D. association.
   ANSWER: D

234. 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

235. 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

236. The _____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

237. The power of self-learning system lies in_____.
   A. cost.
   B. speed.

C. accuracy.

D. simplicity.

ANSWER: C

238. The human brain consists primarily of nerve cells called_____.

A. dendrites.

B. neurons.

C. synapse.

D. perceptron.

ANSWER: B

239. Which of the following is not a decision tree node?

A. Root node.

B. Internal node.

C. Leaf node.

D. predicted node.

ANSWER: D

240. _____ is the closeness of repeated measurements to one another.

A. Precision.

B. Bias.

C. Accuracy.

D. non-scientific.

ANSWER: A

241. Which of the following matrix consist asymmetric data?

A. Sparse data matrix.

B. Indentity matrix.

C. Confusion matrix.

D. Cross matrix.

ANSWER: A

242. A representation of data objects as columns and attributes as rows is called_____.

A. matrix.

B. data matrix.

C. table.

D. file.

ANSWER: B

243. Which of the following is not a data mining attribute?

A. nominal.

B. ordinal.

C. interval.

D. multiple.

ANSWER: D

244. Patterns of machine-language program are_____.

A. definitive theories.

B. hypothesis.

C. not-definitive theories.

D. quantitative.

ANSWER: B

245. Nominal and ordinal attributes are collectively referred to as_____ attributes.

A. qualitative.

B. perfect.

C. consistent.

D. optimized.

ANSWER: A

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

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

248. Which of the following is the data mining tool?
   A. C.
   B. Weka.
   C. C++.
   D. VB.
 ANSWER: B

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

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