

Department of Computer Engineering

22621 DWM MCQ (Data Warehousing with Mining Techniques)

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. MDDDB 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 level.
- 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 at least _____.
- 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. archived.
 - 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 phase 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. Dynamic 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 algorithm.

- 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 known 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 connections 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

by Prof. LAXMI.SREE.B.R.

Happy Learning!
cwipedia.in