

# National Testing Agency

<b>Question Paper Name :</b>	Data Mining 26th March 2021 Shift2
<b>Subject Name :</b>	Data Mining
<b>Creation Date :</b>	2021-03-26 20:01:31
<b>Duration :</b>	180
<b>Number of Questions :</b>	100
<b>Total Marks :</b>	100
<b>Display Marks:</b>	Yes

## Data Mining

<b>Group Number :</b>	1
<b>Group Id :</b>	512452201
<b>Group Maximum Duration :</b>	0
<b>Group Minimum Duration :</b>	120
<b>Show Attended Group? :</b>	No
<b>Edit Attended Group? :</b>	No
<b>Break time :</b>	0
<b>Group Marks :</b>	100
<b>Is this Group for Examiner? :</b>	No

## Data Mining-1

<b>Section Id :</b>	512452877
<b>Section Number :</b>	1
<b>Section type :</b>	Online
<b>Mandatory or Optional :</b>	Mandatory
<b>Number of Questions :</b>	100
<b>Number of Questions to be attempted :</b>	100

**Section Marks :** 100  
**Mark As Answered Required? :** Yes  
**Sub-Section Number :** 1  
**Sub-Section Id :** 512452995  
**Question Shuffling Allowed :** Yes

**Question Number : 1 Question Id : 51245217051 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which one of the following processes is used by large retailers to study trends?

1. Data mining
2. Data selection
3. POS
4. Data conversion

**Options :**

- 51245255003. 1
- 51245255004. 2
- 51245255005. 3
- 51245255006. 4

**Question Number : 2 Question Id : 51245217052 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The Data Warehouse is

1. read only
2. write only
3. read write only
4. none of the above

**Options :**

51245255007. 1  
51245255008. 2  
51245255009. 3  
51245255010. 4

**Question Number : 3 Question Id : 51245217053 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Expansion for DSS in DW is

1. Decision Support system
2. Decision Single System
3. Data Storable System
4. Data Support System

**Options :**

51245255011. 1  
51245255012. 2  
51245255013. 3  
51245255014. 4

**Question Number : 4 Question Id : 51245217054 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which one of the following is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management decisions?

1. Data Mining
2. Data Warehousing
3. Web Mining
4. Text Mining

**Options :**

51245255015. 1

51245255016. 2  
51245255017. 3  
51245255018. 4

**Question Number : 5 Question Id : 51245217055 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

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

1. subject-oriented
2. time-variant
3. integrated
4. all of the above

**Options :**

51245255019. 1  
51245255020. 2  
51245255021. 3  
51245255022. 4

**Question Number : 6 Question Id : 51245217056 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The time horizon in Data warehouse is usually

1. 1-2 years
2. 3-4 years
3. 5-6 years
4. 5-10 years

**Options :**

51245255023. 1

51245255024. 2

51245255025. 3

51245255026. 4

**Question Number : 7 Question Id : 51245217057 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The data is stored, retrieved & updated in

1. OLAP
2. OLTP
3. SMTP
4. FTP

**Options :**

51245255027. 1

51245255028. 2

51245255029. 3

51245255030. 4

**Question Number : 8 Question Id : 51245217058 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Self Organising maps are an example of:

1. Unsupervised learning
2. Supervised learning
3. Reinforcement learning
4. Missing data imputation

**Options :**

51245255031. 1

51245255032. 2

51245255033. 3

51245255034. 4

**Question Number : 9 Question Id : 51245217059 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

\_\_\_\_\_describes the data contained in the data warehouse.

1. Relational data
2. Operational data
3. Metadata
4. Informational data

**Options :**

- 51245255035. 1
- 51245255036. 2
- 51245255037. 3
- 51245255038. 4

**Question Number : 10 Question Id : 51245217060 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which of the following process includes data cleaning, data integration, data transformation, data selection, data mining, pattern evaluation and knowledge presentation?

1. KDD process
2. ETL process
3. KTL process
4. None of the above

**Options :**

- 51245255039. 1
- 51245255040. 2
- 51245255041. 3

51245255042. 4

**Question Number : 11 Question Id : 51245217061 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Association rules are always defined on

1. Binary attributes
2. Single attribute
3. Relational database
4. Multidimensional attributes

**Options :**

- 51245255043. 1
- 51245255044. 2
- 51245255045. 3
- 51245255046. 4

**Question Number : 12 Question Id : 51245217062 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

\_\_\_\_\_ is data about data.

1. Metadata
2. Microdata
3. Minidata
4. Multidata

**Options :**

- 51245255047. 1
- 51245255048. 2
- 51245255049. 3
- 51245255050. 4

**Question Number : 13 Question Id : 51245217063 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The capability of data mining is to build what kind of models?

1. Retrospective
2. Interrogative
3. Predictive
4. Imperative

**Options :**

- 51245255051. 1
- 51245255052. 2
- 51245255053. 3
- 51245255054. 4

**Question Number : 14 Question Id : 51245217064 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

k-nearest neighbor is one of the

1. learning technique
2. OLAP tool
3. Purest search technique
4. Data warehousing tool

**Options :**

- 51245255055. 1
- 51245255056. 2
- 51245255057. 3
- 51245255058. 4

**Question Number : 15 Question Id : 51245217065 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**



Data transformation includes which of the following?

1. A process to change data from a detailed level to a summary level
2. A process to change data from a summary level to a detailed level
3. Joining data from one source into various sources of data
4. Separating data from one source into various sources of data

**Options :**

51245255059. 1

51245255060. 2

51245255061. 3

51245255062. 4

**Question Number : 16 Question Id : 51245217066 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which one of the following predicts future trends & behaviours, allowing business managers to make proactive, knowledge-driven decisions?

1. Data warehouse
2. Data mining
3. Datamarts
4. Metadata

**Options :**

51245255063. 1

51245255064. 2

51245255065. 3

51245255066. 4

**Question Number : 17 Question Id : 51245217067 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The heart of the Data Warehouse is

1. Data mining database servers
2. Data warehouse database servers
3. Data mart database servers
4. Relational data base servers

**Options :**

- 51245255067. 1
- 51245255068. 2
- 51245255069. 3
- 51245255070. 4

**Question Number : 18 Question Id : 51245217068 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

\_\_\_\_\_ is held in the catalog of the warehouse database system

1. Application level metadata
2. Algorithmic level metadata
3. Departmental level metadata
4. Core warehouse metadata

**Options :**

- 51245255071. 1
- 51245255072. 2
- 51245255073. 3
- 51245255074. 4

**Question Number : 19 Question Id : 51245217069 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

\_\_\_\_\_databases are owned by particular departments or business groups

1. Informational
2. Operational
3. Both informational and operational
4. Flat

**Options :**

- 51245255075. 1
- 51245255076. 2
- 51245255077. 3
- 51245255078. 4

**Question Number : 20 Question Id : 51245217070 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Data can be updated in \_\_\_\_environment

1. Data warehouse
2. Data mining
3. Operational
4. Informational

**Options :**

- 51245255079. 1
- 51245255080. 2
- 51245255081. 3
- 51245255082. 4

**Question Number : 21 Question Id : 51245217071 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

What does ETL stand for

1. Execute tramit and load
2. Extract transform and load
3. Excute Transform and load
4. All of the above

**Options :**

- 51245255083. 1
- 51245255084. 2
- 51245255085. 3
- 51245255086. 4

**Question Number : 22 Question Id : 51245217072 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The generic two-level data warehouse architecture includes which of the following?

1. At least one data mart
2. Data that can extracted from numerous internal and external sources
3. Near real-time updates
4. All of the above

**Options :**

- 51245255087. 1
- 51245255088. 2
- 51245255089. 3
- 51245255090. 4

**Question Number : 23 Question Id : 51245217073 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Data scrubbing is

1. A process to reject data from the data warehouse and to create the necessary indexes
2. A process to load the data in the data warehouse and to create the necessary indexes
3. A process to upgrade the quality of data after it is moved into a data warehouse
4. A process to upgrade the quality of data before it is moved into a data warehouse

**Options :**

- 51245255091. 1
- 51245255092. 2
- 51245255093. 3
- 51245255094. 4

**Question Number : 24 Question Id : 51245217074 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The process of removing the deficiencies and loopholes in the data is called as

1. Aggregation of data
2. Extracting of data
3. Cleaning up of data
4. Loading of data

**Options :**

- 51245255095. 1
- 51245255096. 2
- 51245255097. 3
- 51245255098. 4

**Question Number : 25 Question Id : 51245217075 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which of the following is not related to dimension table attributes?

1. Descriptive
2. Equally unavailable
3. Complete
4. Indexed

**Options :**

- 51245255099. 1
- 51245255100. 2
- 51245255101. 3
- 51245255102. 4

**Question Number : 26 Question Id : 51245217076 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The FP-growth algorithm has how many phases?

1. One
2. Two
3. Three
4. Four

**Options :**

- 51245255103. 1
- 51245255104. 2
- 51245255105. 3
- 51245255106. 4

**Question Number : 27 Question Id : 51245217077 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The number of iterations in a priori,

1. increase with the size of the maximum frequent set
2. decrease with increase in size of the maximum frequent set
3. increase with the size of the data
4. decrease with the increase in size of the data

**Options :**

- 51245255107. 1
- 51245255108. 2
- 51245255109. 3
- 51245255110. 4

**Question Number : 28 Question Id : 51245217078 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

After the pruning of a priori algorithm, what set will remain?

1. Only candidate set
2. No candidate set
3. Only border set
4. No border set

**Options :**

- 51245255111. 1
- 51245255112. 2
- 51245255113. 3
- 51245255114. 4

**Question Number : 29 Question Id : 51245217079 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The second phase of A Priori algorithm is

1. Candidate generation
2. Itemset generation
3. Pruning
4. Partitioning

**Options :**

- 51245255115. 1
- 51245255116. 2
- 51245255117. 3
- 51245255118. 4

**Question Number : 30 Question Id : 51245217080 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The first phase of A Priori algorithm is

1. Candidate generation
2. Itemset generation
3. Pruning
4. Partitioning

**Options :**

- 51245255119. 1
- 51245255120. 2
- 51245255121. 3
- 51245255122. 4

**Question Number : 31 Question Id : 51245217081 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**



Which one of the following refers to a comparison of the general features of the target class data objects against the general features of objects from one or multiple contrasting classes?

1. Data Characterization
2. Data Classification
3. Data discrimination
4. Data selection

**Options :**

- 51245255123. 1
- 51245255124. 2
- 51245255125. 3
- 51245255126. 4

**Question Number : 32 Question Id : 51245217082 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Strategic value of data mining is

1. Cost-sensitive
2. Work-sensitive
3. Time-sensitive
4. Technical-sensitive

**Options :**

- 51245255127. 1
- 51245255128. 2
- 51245255129. 3
- 51245255130. 4

**Question Number : 33 Question Id : 51245217083 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which one of the following refers to the process of finding a model that describes and distinguishes data classes or concepts?

1. Data Characterization
2. Data Classification
3. Data discrimination
4. Data selection

**Options :**

- 51245255131. 1
- 51245255132. 2
- 51245255133. 3
- 51245255134. 4

**Question Number : 34 Question Id : 51245217084 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The absolute number of transactions supporting X in T is called

1. Confidence
2. Support
3. Support count
4. None of the above

**Options :**

- 51245255135. 1
- 51245255136. 2
- 51245255137. 3
- 51245255138. 4

**Question Number : 35 Question Id : 51245217085 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The proportion of transactions supporting X in T is called

1. Confidence
2. Support
3. Support count
4. All of the above

**Options :**

- 51245255139. 1
- 51245255140. 2
- 51245255141. 3
- 51245255142. 4

**Question Number : 36 Question Id : 51245217086 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Multi-dimensional knowledge is

1. A class of learning algorithms that try to derive a Prolog program from examples
2. A table with n independent attributes can be seen as an n-dimensional space
3. A prediction made using an extremely simple method, such as always predicting the same output
4. None of the above

**Options :**

- 51245255143. 1
- 51245255144. 2
- 51245255145. 3
- 51245255146. 4

**Question Number : 37 Question Id : 51245217087 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Data cleaning is

1. Large collection of data mostly stored in a computer system
2. The removal of noise errors and incorrect input from a database
3. The systematic description of the syntactic structure of a specific database. It describes the structure of the attributes the tables and foreign key relationships
4. All of the above

**Options :**

- 51245255147. 1
- 51245255148. 2
- 51245255149. 3
- 51245255150. 4

**Question Number : 38 Question Id : 51245217088 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The difference between supervised learning and unsupervised learning is that

1. unlike unsupervised learning, supervised learning needs labeled data
2. unlike unsupervised learning, supervised learning can be used to detect outliers
3. there is no difference
4. unlike supervised learning, unsupervised learning can form new classes

**Options :**

- 51245255151. 1
- 51245255152. 2
- 51245255153. 3
- 51245255154. 4

**Question Number : 39 Question Id : 51245217089 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Identify the example of sequence data

1. Weather forecast
2. Data matrix
3. Market basket data
4. Genomic data

**Options :**

- 51245255155. 1
- 51245255156. 2
- 51245255157. 3
- 51245255158. 4

**Question Number : 40 Question Id : 51245217090 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which of the following is not a data mining task?

1. Feature Subset Detection
2. Association Rule Discovery
3. Regression
4. Sequential Pattern Discovery

**Options :**

- 51245255159. 1
- 51245255160. 2
- 51245255161. 3
- 51245255162. 4

**Question Number : 41 Question Id : 51245217091 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Enrichment means

1. Adding external data
2. Deleting data
3. Cleaning Data
4. Selecting the data

**Options :**

- 51245255163. 1
- 51245255164. 2
- 51245255165. 3
- 51245255166. 4

**Question Number : 42 Question Id : 51245217092 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The decision support system is used only for

1. Cleaning
2. Coding
3. Selecting
4. Queries

**Options :**

- 51245255167. 1
- 51245255168. 2
- 51245255169. 3
- 51245255170. 4

**Question Number : 43 Question Id : 51245217093 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Learning is

1. The process of finding the right formal representation of a certain body of knowledge in order to represent it in a knowledge-based system
2. It automatically maps an external signal space into a system's internal representational space. They are useful in the performance of classification tasks.
3. A process where an individual learns how to carry out a certain task when making a transition from a situation in which the task cannot be carried out to a situation in which the same task under the same circumstances can be carried out.
4. None of the above

**Options :**

- 51245255171. 1
- 51245255172. 2
- 51245255173. 3
- 51245255174. 4

**Question Number : 44 Question Id : 51245217094 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The output of KDD is

1. Data
2. Information
3. Query
4. Useful information

**Options :**

- 51245255175. 1
- 51245255176. 2
- 51245255177. 3
- 51245255178. 4

**Question Number : 45 Question Id : 51245217095 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Deep knowledge can be found only by using

1. Clues
2. OLAP
3. SQL
4. Algorithm

**Options :**

- 51245255179. 1
- 51245255180. 2
- 51245255181. 3
- 51245255182. 4

**Question Number : 46 Question Id : 51245217096 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Metadata describes

1. Content of database
2. Structure of content of database
3. Structure of database
4. Database itself

**Options :**

- 51245255183. 1
- 51245255184. 2
- 51245255185. 3
- 51245255186. 4

**Question Number : 47 Question Id : 51245217097 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**



**Correct Marks : 1 Wrong Marks : 0**

Which of the following is NOT a component of a data warehouse?

1. Metadata
2. Current detail data
3. Lightly summarized data
4. Component Key

**Options :**

- 51245255187. 1
- 51245255188. 2
- 51245255189. 3
- 51245255190. 4

**Question Number : 48 Question Id : 51245217098 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The core of the multidimensional model is the ....., which consists of a large set of facts and a number of dimensions.

1. Multidimensional cube
2. Dimensions cube
3. Data cube
4. Data model

**Options :**

- 51245255191. 1
- 51245255192. 2
- 51245255193. 3
- 51245255194. 4

**Question Number : 49 Question Id : 51245217099 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The partition of overall data warehouse is

1. Database
2. Data cube
3. Data mart
4. Operational data

**Options :**

- 51245255195. 1
- 51245255196. 2
- 51245255197. 3
- 51245255198. 4

**Question Number : 50 Question Id : 51245217100 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which one of the following is an example for case based learning?

1. Decision tree
2. Neural network
3. Genetic algorithm
4. K-Nearest neighbor

**Options :**

- 51245255199. 1
- 51245255200. 2
- 51245255201. 3
- 51245255202. 4

**Question Number : 51 Question Id : 51245217101 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

A data warehouse

1. is updated by end users
2. contains numerous naming conventions and formats
3. is organized around important subject areas
4. contains only current data

**Options :**

- 51245255203. 1
- 51245255204. 2
- 51245255205. 3
- 51245255206. 4

**Question Number : 52 Question Id : 51245217102 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Business Intelligence and Data Warehousing is used for

1. Forecasting
2. Data Mining
3. Analysis of large volumes of product sales data
4. All of the above

**Options :**

- 51245255207. 1
- 51245255208. 2
- 51245255209. 3
- 51245255210. 4

**Question Number : 53 Question Id : 51245217103 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

k-nearest neighbour is one of the

1. Learning technique
2. OLAP tool
3. Purest search technique
4. Data warehousing tool

**Options :**

- 51245255211. 1
- 51245255212. 2
- 51245255213. 3
- 51245255214. 4

**Question Number : 54 Question Id : 51245217104 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The complexity of data mining algorithm is represented by

1.  $\log n$
2.  $2n \log n$
3.  $n \log n$
4.  $2 \log n$

**Options :**

- 51245255215. 1
- 51245255216. 2
- 51245255217. 3
- 51245255218. 4

**Question Number : 55 Question Id : 51245217105 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which one of the following are designed to overcome any limitations placed on the warehouse by the nature of the relational data model?

1. Operational database
2. Relational database
3. Multidimensional database
4. Data repository

**Options :**

- 51245255219. 1
- 51245255220. 2
- 51245255221. 3
- 51245255222. 4

**Question Number : 56 Question Id : 51245217106 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which of the following process includes data cleaning, data integration, data selection, data transformation, data mining, pattern evolution and knowledge presentation?

1. KDD process
2. ETL process
3. KTL process
4. MDX process

**Options :**

- 51245255223. 1
- 51245255224. 2
- 51245255225. 3
- 51245255226. 4

**Question Number : 57 Question Id : 51245217107 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The intermediate unit in perception is

1. Photoreceptors
2. Associators
3. Responders
4. Receptors

**Options :**

- 51245255227. 1
- 51245255228. 2
- 51245255229. 3
- 51245255230. 4

**Question Number : 58 Question Id : 51245217108 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

In KDD and data mining, noise is referred to as

1. Repeated data
2. Complex data
3. Meta data
4. Random errors in database

**Options :**

- 51245255231. 1
- 51245255232. 2
- 51245255233. 3
- 51245255234. 4

**Question Number : 59 Question Id : 51245217109 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

An OLAP tool provides for

1. Multidimensional analysis
2. Roll-up and drill-down
3. Slicing and dicing
4. Rotation

**Options :**

- 51245255235. 1
- 51245255236. 2
- 51245255237. 3
- 51245255238. 4

**Question Number : 60 Question Id : 51245217110 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Data independence means

1. Data is defined separately and not included in programs
2. Programs are not dependent on the physical attributes of data
3. Programs are not dependent on the logical attributes of data
4. All of the above

**Options :**

- 51245255239. 1
- 51245255240. 2
- 51245255241. 3
- 51245255242. 4

**Question Number : 61 Question Id : 51245217111 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Challenges of clustering include

1. High dimensionality of data
2. Scalability
3. Noisy data
4. All of the above

**Options :**

- 51245255243. 1
- 51245255244. 2
- 51245255245. 3
- 51245255246. 4

**Question Number : 62 Question Id : 51245217112 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which is NOT a type of clustering?

1. Density based
2. Similarity based
3. Decision driven
4. Partition Based

**Options :**

- 51245255247. 1
- 51245255248. 2
- 51245255249. 3
- 51245255250. 4

**Question Number : 63 Question Id : 51245217113 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**



Which one of the following is NOT a clustering technique?

1. Agglomerative
2. Derivative
3. Partitioning
4. Density Based

**Options :**

- 51245255251. 1
- 51245255252. 2
- 51245255253. 3
- 51245255254. 4

**Question Number : 64 Question Id : 51245217114 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

In the k-means clustering algorithm, the distance between cluster centroid to each object is calculated using \_\_\_\_\_ method.

1. Cluster distance
2. Euclidean distance
3. Cluster width
4. None of the above

**Options :**

- 51245255255. 1
- 51245255256. 2
- 51245255257. 3
- 51245255258. 4

**Question Number : 65 Question Id : 51245217115 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

k-means clustering is also referred to as

1. Non-hierarchical clustering
2. Optimizing partitioning
3. Divisive clustering
4. Agglomerative clustering

**Options :**

- 51245255259. 1
- 51245255260. 2
- 51245255261. 3
- 51245255262. 4

**Question Number : 66 Question Id : 51245217116 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

A decision tree is a tree in which every node is either a \_\_\_\_ or a decision tree

1. Leaf node
2. Root node
3. Sub node
4. All of the above

**Options :**

- 51245255263. 1
- 51245255264. 2
- 51245255265. 3
- 51245255266. 4

**Question Number : 67 Question Id : 51245217117 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which algorithm is used to find correlations among different attributes in a data set?

1. Associative algorithm
2. Association algorithm
3. Time series algorithm
4. Series algorithm

**Options :**

- 51245255267. 1
- 51245255268. 2
- 51245255269. 3
- 51245255270. 4

**Question Number : 68 Question Id : 51245217118 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The process of taking operational data from one or more sources and mapping it field by field onto a new data structure in the data warehouse is

1. Transformation
2. Cleansing
3. Integration
4. Scrubbing

**Options :**

- 51245255271. 1
- 51245255272. 2
- 51245255273. 3
- 51245255274. 4

**Question Number : 69 Question Id : 51245217119 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

What is meant by discrete data?

1. One that allows only finite set of values
2. One that allows real numbers only
3. One that allows float values only
4. All of the above

**Options :**

- 51245255275. 1
- 51245255276. 2
- 51245255277. 3
- 51245255278. 4

**Question Number : 70 Question Id : 51245217120 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The purpose of validation partition in data mining is

1. It builds candidate models
2. It evaluates candidate models
3. It evaluates the final models
4. None of the above

**Options :**

- 51245255279. 1
- 51245255280. 2
- 51245255281. 3
- 51245255282. 4

**Question Number : 71 Question Id : 51245217121 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The process of forming general concept definitions from examples of concept to be learned

1. Deduction
2. Abduction
3. Induction
4. Conjunction

**Options :**

- 51245255283. 1
- 51245255284. 2
- 51245255285. 3
- 51245255286. 4

**Question Number : 72 Question Id : 51245217122 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Data mining is best described as the process of

1. Identifying patterns data
2. Deducting relationships in data
3. Representing data
4. Simulating trends in data

**Options :**

- 51245255287. 1
- 51245255288. 2
- 51245255289. 3
- 51245255290. 4

**Question Number : 73 Question Id : 51245217123 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Supervised learning and unsupervised clustering, both require at least one

1. Hidden attribute
2. Input/output attribute
3. Categorical attribute
4. All of the above

**Options :**

- 51245255291. 1
- 51245255292. 2
- 51245255293. 3
- 51245255294. 4

**Question Number : 74 Question Id : 51245217124 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

At which level can we create dimensional models?

1. Detailed models level
2. Architecture models level
3. Implementation level
4. Testing level

**Options :**

- 51245255295. 1
- 51245255296. 2
- 51245255297. 3
- 51245255298. 4

**Question Number : 75 Question Id : 51245217125 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which of the following is the collection of data objects that are similar to one another within the same group?

1. Partitioning
2. Cluster
3. Table
4. Data source

**Options :**

- 51245255299. 1
- 51245255300. 2
- 51245255301. 3
- 51245255302. 4

**Question Number : 76 Question Id : 51245217126 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which of the following is true for classification?

1. A subdivision of a set
2. A measure of the accuracy
3. The task of assigning a classification
4. All of the above

**Options :**

- 51245255303. 1
- 51245255304. 2
- 51245255305. 3
- 51245255306. 4

**Question Number : 77 Question Id : 51245217127 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Prediction is

1. The result of the application of a theory or a rule in a specific case
2. One of several possible enters within a database table that is chosen by the designer as the primary means of accessing the data in the table
3. Discipline in statistics that studies ways to find the most interesting projections of multi-dimensional spaces
4. None of the above

**Options :**

- 51245255307. 1
- 51245255308. 2
- 51245255309. 3
- 51245255310. 4

**Question Number : 78 Question Id : 51245217128 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Naive prediction is

1. A class of learning algorithms that try to derive a Prolog program from examples
2. A table with n independent attributes can be seen as an n- dimensional space
3. A prediction made using an extremely simple method, such as always predicting the same output
4. All of the above

**Options :**

- 51245255311. 1
- 51245255312. 2
- 51245255313. 3
- 51245255314. 4

**Question Number : 79 Question Id : 51245217129 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**



Rule based classification algorithms generate \_\_\_\_\_ rule to perform the classification.

1. if-then
2. while
3. do while
4. switch

**Options :**

- 51245255315. 1
- 51245255316. 2
- 51245255317. 3
- 51245255318. 4

**Question Number : 80 Question Id : 51245217130 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Classification rules are extracted from the \_\_\_\_\_.

1. root node
2. decision tree
3. siblings
4. branches

**Options :**

- 51245255319. 1
- 51245255320. 2
- 51245255321. 3
- 51245255322. 4

**Question Number : 81 Question Id : 51245217131 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

One popular classification technique in Business Intelligence (BI) reporting is

1. Cluster analysis only
2. Regression analysis only
3. RFM analysis only
4. Both cluster analysis and regression analysis

**Options :**

51245255323. 1

51245255324. 2

51245255325. 3

51245255326. 4

**Question Number : 82 Question Id : 51245217132 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which operation is mainly used to access the live data online and analyse it?

1. ROLAP
2. OLAP
3. MOLAP
4. HOLAP

**Options :**

51245255327. 1

51245255328. 2

51245255329. 3

51245255330. 4

**Question Number : 83 Question Id : 51245217133 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

In data mining, the slice operation performs a selection on

1. One dimension
2. Two dimension
3. Multi dimension
4. None of the above

**Options :**

- 51245255331. 1
- 51245255332. 2
- 51245255333. 3
- 51245255334. 4

**Question Number : 84 Question Id : 51245217134 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The pre-processing step used in data mining is

1. cleaning
2. integration
3. transformation
4. reduction

**Options :**

- 51245255335. 1
- 51245255336. 2
- 51245255337. 3
- 51245255338. 4

**Question Number : 85 Question Id : 51245217135 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Total number of stages in KDD is

1. 3
2. 4
3. 5
4. 6

**Options :**

- 51245255339. 1
- 51245255340. 2
- 51245255341. 3
- 51245255342. 4

**Question Number : 86 Question Id : 51245217136 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which one of the following can be thought of as classifying an attribute value into one of a the set of possible classes?

1. Estimation
2. Prediction
3. Identification
4. Clarification

**Options :**

- 51245255343. 1
- 51245255344. 2
- 51245255345. 3
- 51245255346. 4

**Question Number : 87 Question Id : 51245217137 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

\_\_\_\_\_and prediction may be viewed as types of classification

1. Decision
2. Verification
3. Estimation
4. Illustration

**Options :**

- 51245255347. 1
- 51245255348. 2
- 51245255349. 3
- 51245255350. 4

**Question Number : 88 Question Id : 51245217138 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Pick out a hierarchical clustering algorithm

1. DBSCAN
2. BIRCH
3. PAM
4. CURE

**Options :**

- 51245255351. 1
- 51245255352. 2
- 51245255353. 3
- 51245255354. 4

**Question Number : 89 Question Id : 51245217139 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which clustering technique starts with as many clusters as there are records, with each cluster having only one record?

1. Agglomerative
2. Divisive
3. Partition
4. Numeric

**Options :**

- 51245255355. 1
- 51245255356. 2
- 51245255357. 3
- 51245255358. 4

**Question Number : 90 Question Id : 51245217140 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The goal of which one of the following is to discover both the dense and sparse regions of a data set?

1. Association rule
2. Classification
3. Clustering
4. Genetic Algorithm

**Options :**

- 51245255359. 1
- 51245255360. 2
- 51245255361. 3
- 51245255362. 4

**Question Number : 91 Question Id : 51245217141 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

The process of upgrading the existing quality of data before it is moved into a data warehouse is known as

1. Data Scrubbing
2. Data Polishing
3. Data Auditing
4. Data Aggregation

**Options :**

- 51245255363. 1
- 51245255364. 2
- 51245255365. 3
- 51245255366. 4

**Question Number : 92 Question Id : 51245217142 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

In OLAP cubes, data (measures) are categorized by

1. Length
2. Angles
3. Dimensions
4. Breadth

**Options :**

- 51245255367. 1
- 51245255368. 2
- 51245255369. 3
- 51245255370. 4

**Question Number : 93 Question Id : 51245217143 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

An OLAP system is market-oriented and is used for data analysis by knowledge workers, including managers, executives, and analysts.

1. True
2. False
3. May be
4. None of the above

**Options :**

- 51245255371. 1
- 51245255372. 2
- 51245255373. 3
- 51245255374. 4

**Question Number : 94 Question Id : 51245217144 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which one of the following is NOT a Data Transformation strategy?

1. Aggregation
2. Normalization
3. Generalization
4. Compression

**Options :**

- 51245255375. 1
- 51245255376. 2
- 51245255377. 3
- 51245255378. 4

**Question Number : 95 Question Id : 51245217145 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**



A subset of the data warehouse that contains all the data about a particular topic is known as

1. Data Pocket
2. Metadata
3. Data Mart
4. Data Mining

**Options :**

- 51245255379. 1
- 51245255380. 2
- 51245255381. 3
- 51245255382. 4

**Question Number : 96 Question Id : 51245217146 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

In web mining, which one of the following is used to know the order in which URLs tend to be accessed?

1. Clustering
2. Associations
3. Sequential analysis
4. Classification

**Options :**

- 51245255383. 1
- 51245255384. 2
- 51245255385. 3
- 51245255386. 4

**Question Number : 97 Question Id : 51245217147 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

In web mining, which one of the following is used to know which URLs tend to be requested together?

1. Clustering
2. Associations
3. Sequential analysis
4. Classification

**Options :**

- 51245255387. 1
- 51245255388. 2
- 51245255389. 3
- 51245255390. 4

**Question Number : 98 Question Id : 51245217148 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which one of the following describes the discovery of useful information from web content?

1. Web content mining
2. Web structure mining
3. Web usage mining
4. All of the above

**Options :**

- 51245255391. 1
- 51245255392. 2
- 51245255393. 3
- 51245255394. 4

**Question Number : 99 Question Id : 51245217149 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which one of the following is concerned with discovering the model underlying the link structures of the web?

1. Web content mining
2. Web structure mining
3. Web usage mining
4. All of the above

**Options :**

- 51245255395. 1
- 51245255396. 2
- 51245255397. 3
- 51245255398. 4

**Question Number : 100 Question Id : 51245217150 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No Correct Marks : 1 Wrong Marks : 0**

Which one of the following is the way of studying the web link structure?

1. Computer network
2. Physical network
3. Social network
4. Logical network

**Options :**

- 51245255399. 1
- 51245255400. 2
- 51245255401. 3
- 51245255402. 4