University of Mumbai
Online Examination 2020
Program: BE Computer Engineering
Curriculum Scheme: Revised 2016
Examination: Third Year Semester VI
Course Code: CSC603 and Course Name: Data Warehouse and Mining

Time: 1hour   Max. Marks: 50

Note to the students:- All the Questions are compulsory and carry equal marks .

Q1.	Which of the following is direct application of frequent itemset mining?
Option A:	Social Network Analysis
Option B:	Market Basket Analysis
Option C:	Outlier Detection
Option D:	Intrusion Detection

Q2.	Supervised learning and unsupervised clustering both require at least one _____
Option A:	Hidden attribute
Option B:	Output attribute.
Option C:	Input attribute.
Option D:	Categorical attribute.

Q3.	In ETL, If errors are found during transformation then _____
Option A:	the transformation is continue for data backup.
Option B:	the transformation is start again from first step.
Option C:	the transformation is stopped.
Option D:	the loading start directly.

Q4.	What is incorrect considering density based clustering?
Option A:	Data points with outliers
Option B:	Data points with different densities

Option C:	Data points with round shape
Option D:	Data points with non convex shapes

Q5.	Which type of database can be used for optimizing data warehousing and OLAP (online analytical processing)?
Option A:	Unstructured
Option B:	Relational
Option C:	Structural
Option D:	Multidimensional

Q6.	What is the relation between candidate and frequent itemsets?
Option A:	A candidate itemset is always a frequent itemset
Option B:	A frequent itemset must be a candidate itemset
Option C:	No relation between the two
Option D:	Both are same

Q7.	Which of the below is not used in Web Mining
Option A:	Web content mining
Option B:	Web Structure Mining
Option C:	Web view mining
Option D:	Web Usage Mining

Q8.	A bitmap index is for the columns in which the ratio of the number of distinct values to the number of rows in the table, a ratio called as ___.
Option A:	degree of cardinality
Option B:	degree of freedom
Option C:	degree of index
Option D:	bitmap degree

Q9.	The ……………… allows the selection of the relevant information necessary for the data warehouse.
Option A:	top-down view
Option B:	data warehouse view
Option C:	data source view
Option D:	business query view

Q10.	Which of the following is not a kind of data warehouse application?
Option A:	Information processing
Option B:	Analytical processing
Option C:	Data mining
Option D:	Transaction processing

Q11.	Which method shows hierarchical data in a nested format?
Option A:	Treemaps
Option B:	Scatter plots
Option C:	Population pyramids
Option D:	Area charts

Q12.	Which is the classification improving technique?
Option A:	Holdout
Option B:	Cross validation
Option C:	Bootstrap
Option D:	Bagging

Q13.	Unsupervised classification is called as
Option A:	Dimension reduction
Option B:	Attribute Pruning
Option C:	Distance measurement
Option D:	Clustering

Q14.	Clickstream data is used in ___ mining
Option A:	Structure
Option B:	content
Option C:	usage
Option D:	site

Q15.	Important Component of spatial data is ___

Option A:	location
Option B:	weight
Option C:	color
Option D:	area

Q16.	Which data structure is used in Spatial data structures?
Option A:	B-Tree
Option B:	queue
Option C:	stack
Option D:	R-Tree

Q17.	What is not true about FP growth algorithms?
Option A:	It mines frequent itemsets without candidate generation.
Option B:	There are chances that FP trees may not fit in the memory
Option C:	FP trees are very expensive to build
Option D:	It expands the original database to build FP trees.

Q18.	An operational system is which of the following?
Option A:	A system that is used to run the business in real time and is based on current data
Option B:	A system that is used to run the business in real time and is based on historical data
Option C:	A system that is used to support decision making and is based on historical data
Option D:	A system that is used to support decision making and is based on current data

Q19.	What is true of the multidimensional model?
Option A:	It typically requires less disk storage
Option B:	It typically requires more disk storage
Option C:	Typical business queries requiring aggregate functions take more time
Option D:	Increasing the size of a dimension is difficult

Q20.	Which of the following is not a data mining functionality?
Option A:	Characterization and Discrimination
Option B:	Classification and regression

Option C:	Selection and interpretation
Option D:	Clustering and Analysis

Q21.	Identify the example of Nominal attribute.
Option A:	Temperature
Option B:	Gender
Option C:	Mass
Option D:	Salary

Q22.	Which of the following is not a data pre-processing methods?
Option A:	Data Visualization
Option B:	Data Discretization
Option C:	Data Cleaning
Option D:	Data Reduction

Q23.	Which process removes redundancies that may be present in a data model?
Option A:	Abstraction
Option B:	Granularization
Option C:	Standardization
Option D:	Normalization

Q24.	Which of the below is not used to improve the efficiency of apriori algorithm ?
Option A:	Transaction reduction
Option B:	Partitioning
Option C:	Hash-Based technique
Option D:	Tree prunning

Q25.	The dimension tables describe the _____.
Option A:	entities.
Option B:	facts.
Option C:	keys.
Option D:	units of measures.