Program: BE Computer Engineering

Curriculum Scheme: Revised 2016

Examination: Third Year Semester VI

Course Code: CSC603 and Course Name: Data Warehousing & Mining

Time: 1 hour   Max. Marks: 50

==

3009_R16_Comp_VI_CSC603_QP4

==

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

Q1.	_____ is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management decisions.
Option A:	Data Mining.
Option B:	Data Warehousing.
Option C:	Web Mining.
Option D:	Text Mining.
Q2.	Data warehouse is
Option A:	The actual discovery phase of a knowledge discovery process
Option B:	The stage of selecting the right data for a KDD process
Option C:	A subject-oriented integrated time variant non-volatile collection of data in support of management
Option D:	Extraction of data
Q3.	The key used in operational environment may not have an element of_____.
Option A:	time.
Option B:	cost.
Option C:	frequency.
Option D:	quality.
Q4.	Data warehouse contains _____ data that is never found in the operational environment.
Option A:	normalized.
Option B:	informational.
Option C:	summary.
Option D:	denormalized.
Q5.	Data transformation includes _____.
Option A:	a process to change data from a detailed level to a summary level.

Option B:	a process to change data from a summary level to a detailed level.
Option C:	joining data from one source into various sources of data.
Option D:	separating data from one source into various sources of data.
Q6.	Where is the ETL process perform?
Option A:	Data Storage
Option B:	Data Warehouse
Option C:	Data Mines
Option D:	Data Files
Q7.	In asymmetric attibute
Option A:	No value is considered important over other values
Option B:	All values are equals
Option C:	Only non-zero value is important
Option D:	Range of values is important
Q8.	_____ is a system where operations like data extraction, transformation and loading operations are executed.
Option A:	Data staging
Option B:	Data integration
Option C:	ETL
Option D:	Transform
Q9.	Background knowledge is...
Option A:	It is a form of automatic learning.
Option B:	A neural network that makes use of a hidden layer
Option C:	The additional acquaintance used by a learning algorithm to facilitate the learning process
Option D:	Plain Neurons
Q10.	Dimensionality reduction reduces the data set size by removing _____
Option A:	composite attributes
Option B:	derived attributes
Option C:	relevant attributes
Option D:	irrelevant attributes
Q11.	Data scrubbing can be defined as
Option A:	Check field overloading
Option B:	Delete redundant tuples
Option C:	Use simple domain knowledge (e.g., postal code, spell-check) to detect errors and make corrections
Option D:	Analyzing data to discover rules and relationship to detect violators
Q12.	.... is a comparison of the general features of the target class data objects against the general features of objects from one or multiple contrasting classes

Option A:	Data Characterization
Option B:	Data Classification
Option C:	Data discrimination
Option D:	Data selection
Q13.	Which of the following applied on warehouse?
Option A:	write only
Option B:	read only
Option C:	both a & b
Option D:	Copy Data
Q14.	Which of the following is true for Classification?
Option A:	A subdivision of a set
Option B:	A measure of the accuracy
Option C:	The task of assigning a classification
Option D:	Unit of Accuracy
Q15.	When data are classified according to a single characteristic, it is called:
Option A:	Quantitative classification
Option B:	Qualitative classification
Option C:	Area classification
Option D:	Simple classification
Q16.	Which one of these is a tree based learner?
Option A:	Rule based
Option B:	Bayesian Belief Network
Option C:	Bayesian classifier
Option D:	Random Forest
Q17.	Which of the following classifications would best suit the student performance classification systems?
Option A:	If...then... analysis
Option B:	Market-basket analysis
Option C:	Regression analysis
Option D:	Cluster analysis
Q18.	The coefficient of correlation
Option A:	is the square of the coefficient of determination
Option B:	is the square root of the coefficient of determination
Option C:	is the same as r-square
Option D:	can never be negative
Q19.	When do you consider an association rule interesting?
Option A:	If it only satisfies min_support
Option B:	If it only satisfies min_confidence
Option C:	If it satisfies both min_support and min_confidence

Option D:	There are other measures to check
Q20.	What does FP growth algorithm do?
Option A:	It mines all frequent patterns through pruning rules with lesser support
Option B:	It mines all frequent patterns through pruning rules with higher support
Option C:	It mines all frequent patterns by constructing a FP tree
Option D:	It guesses Patterns
Q21.	The absolute number of transactions supporting X in T is called _____.
Option A:	confidence.
Option B:	support.
Option C:	support count.
Option D:	Confidence count
Q22.	The left hand side of an association rule is called _____.
Option A:	consequent.
Option B:	onset.
Option C:	antecedent.
Option D:	precedent.
Q23.	Spatial databases' are also known as
Option A:	Geodatabases
Option B:	Monodatabases
Option C:	Concurrent databases
Option D:	Web databases
Q24.	In web mining, _____ is used to know which URLs tend to be requested together.
Option A:	clustering.
Option B:	associations.
Option C:	sequential analysis.
Option D:	classification.
Q25.	Which of the following is not an appropriate way to compromise web servers?
Option A:	Misconfiguration in OS
Option B:	Using network vulnerabilities
Option C:	Misconfiguration in networks
Option D:	Bugs in OS which allow commands to run on web servers