

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Program: TE Information Technology
Curriculum Scheme: Rev2012
Examination: Third Year Semester VI
Course Code: ITC 604 and Course Name: DMBI

Time: 1 hour

Max. Marks: 50

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

Q1.	Strategic value of data mining is
Option A:	cost-sensitive
Option B:	work-sensitive
Option C:	technical-sensitive
Option D:	time-sensitive
Q2.	Calculate median? { 3,6,6,6,8,8,9,11,13,15}
Option A:	9
Option B:	7
Option C:	8
Option D:	8.5
Q3.	An _____ is a heuristic for selecting the splitting criterion that “best” separates a given data partition, D, of class-labeled training tuples into individual classes
Option A:	Association rules
Option B:	Decision tree
Option C:	Clustering
Option D:	Attribute selection measure
Q4.	For unimodal frequency curves that are moderately skewed (asymmetrical), we have the following empirical relation
Option A:	mean - mode = 3*(mean + median)
Option B:	mean - mode = 2*(mean - median)
Option C:	mean - mode = 5*(mean - median)
Option D:	mean - mode = 3*(mean - median)
Q5.	A _____ is one of the most effective graphical methods for determining if there appears to be a relationship, pattern, or trend between two numerical attributes
Option A:	scatter plot
Option B:	quantile-quantile plot
Option C:	quantile plot
Option D:	loess curves
Q6.	_____ methods smooth a sorted data value by consulting its “neighborhood,” that is, the values around it. The sorted values are distributed into a number of “buckets,” or bins
Option A:	Regression
Option B:	Binning

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Option C:	Clustering
Option D:	Boxplot
Q7.	Info(D) is also known as the _____ of D
Option A:	Entropy
Option B:	Average
Option C:	Mean
Option D:	Mode
Q8.	$P(X H)$ is the _____ probability of X conditioned on H
Option A:	Prior
Option B:	Posterior
Option C:	Conditional
Option D:	Subjective
Q9.	_____ is used to split up your dataset into a train and test set
Option A:	cross validation method
Option B:	Bootstrap method
Option C:	Hold out method
Option D:	Random sampling
Q10.	_____ is when the dataset is randomly split up into K-groups
Option A:	cross validation
Option B:	hold out
Option C:	Bootstrap method
Option D:	error measure
Q11.	Which statement is not true about cluster analysis
Option A:	Cluster analysis is a technique for analyzing data when the criterion or dependent variable is categorical and the independent variables are interval in nature
Option B:	Objects in each cluster tend to be similar to each other and dissimilar to objects in the other clusters
Option C:	Cluster analysis is also called classification analysis or numerical taxonomy
Option D:	Groups or clusters are suggested by the data, not defined a priori
Q12.	The metric that is used to calculate cluster distances is called
Option A:	length metric
Option B:	Distance metric
Option C:	Cluster metric
Option D:	Linkage metrics
Q13.	BIRCH stands for
Option A:	Balanced Integral Reduction and Clustering using Hierarchies
Option B:	Balanced Iterative Recurrent Clustering using Hierarchies
Option C:	Balanced Iterative Reducing and Clustering using Hierarchies
Option D:	Balanced Integral Recurrent Clustering using Hierarchies
Q14.	Consider points $x_1 = (6,8)$; $x_2 = (5,7)$; $x_3 = (8,4)$. What would be cluster feature for

University of Mumbai
Examination 2020 under cluster 4 (PCE)

	{x1,x2,x3}
Option A:	(3,(10,19),(120,125))
Option B:	(3,(19,19),(125,129))
Option C:	(3,(9,10),(100,81))
Option D:	(3,(10,19),(125,129))
Q15.	When an algorithm uses the _____, to measure the distance between clusters, it is sometimes called a nearest-neighbor clustering algorithm
Option A:	minimum distance
Option B:	total distance
Option C:	maximum distance
Option D:	average distance
Q16.	Select application where Outlier Analysis is used
Option A:	Market Basket Analysis
Option B:	Credit card Fraud Detection
Option C:	Segmentation of Customer
Option D:	prediction of rain fall
Q17.	What does Apriori algorithm do
Option A:	Discovering interesting relations between variables in large databases
Option B:	Making a group of abstract objects into classes of similar objects
Option C:	Predict the class label of tuple whose label is not known
Option D:	Finding outliers in dataset
Q18.	Which Frequent Pattern Mining Algorithm requires more scans of data
Option A:	DBSCAN
Option B:	FP growth
Option C:	OPTICS
Option D:	Apriori
Q19.	The first step of Apriori algorithm is
Option A:	Itemset generation
Option B:	Pruning
Option C:	Candidate generation
Option D:	Partitioning
Q20.	What makes BI 2.0 different
Option A:	Semi structured data is taken care of
Option B:	Unstructured data is taken care of
Option C:	Dynamic querying of real-time corporate data
Option D:	Structured data is taken care of
Q21.	We have Market Basket data for 1,000 rental transactions at a Video Store. There are four videos for rent -- Video A, Video B, Video C and Video D. The probability that Video D will be rented given that Video C has been rented is known as
Option A:	Support

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Option B:	Confidence
Option C:	lift
Option D:	the basic probability
Q22.	Operational databases are designed to support _____ whereas decision support systems are designed to support _____
Option A:	transactional processing, data analysis
Option B:	data analysis, transactional processing
Option C:	independent data marts, dependent data marts
Option D:	dependent data marts, independent data marts
Q23.	What is the full form of DSS in Data Warehouse
Option A:	Decisive selection system
Option B:	Decision support solution
Option C:	Decision solution system
Option D:	Decision support system
Q24.	An _____ is a heuristic for selecting the splitting criterion that “best” separates a given data partition, D, of class-labeled training tuples into individual classes
Option A:	Clustering
Option B:	Decision tree
Option C:	Attribute selection measure
Option D:	Association rules
Q25.	What is meant by the term 'accuracy'
Option A:	The overall quality of the data
Option B:	The extent to which a value approaches its true value
Option C:	The lack of bias in the data
Option D:	The level of detail at which data is stored.

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Program: TE Information Technology

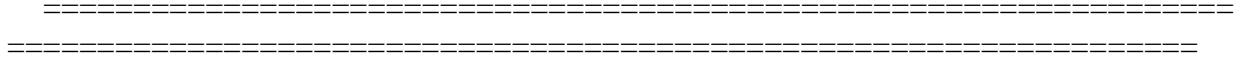
Curriculum Scheme: Rev2012

Examination: Third Year Semester VI

Course Code: ITC 604 and Course Name: DMBI

Time: 1 hour

Max. Marks: 50



Question	Correct Option (Enter either 'A' or 'B' or 'C' or 'D')
Q1.	D
Q2.	C
Q3.	D
Q4	D
Q5	A
Q6	B
Q7	A
Q8.	B
Q9.	C
Q10.	A
Q11.	A
Q12.	D
Q13.	C
Q14.	B
Q15.	A
Q16.	B
Q17.	A
Q18.	D
Q19.	C
Q20.	C
Q21.	B
Q22.	A
Q23.	D
Q24.	C
Q25.	B