

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Program: TE Information Technology

Curriculum Scheme: Rev2016

Examination: Third Year Semester VI

Course Code: ITC602 and Course Name: Data Mining and Business Intelligence

Time: 1 hour

Max. Marks: 50

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

Q1.	_____ 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 discrimination
Option B:	Data Classification
Option C:	Data selection
Option D:	Data Characterization
Q2.	Which statement is not TRUE regarding a data mining task?
Option A:	Clustering is a descriptive data mining task
Option B:	Classification is a predictive data mining task
Option C:	Regression is a descriptive data mining task
Option D:	Deviation detection is a predictive data mining task
Q3.	_____ is a summarization of the general characteristics or features of a target class of data.
Option A:	Data selection
Option B:	Data Characterization
Option C:	Data Classification
Option D:	Data discrimination
Q4.	What is meant by discrete data?
Option A:	One that allows Only infinite set of values
Option B:	One that allows Only finite set of values
Option C:	One that allows real numbers only
Option D:	One that allows float values only.
Q5.	The data are replaced by alternative, smaller representations using parametric models is called____
Option A:	Data Reduction
Option B:	Integration
Option C:	Numerosity reduction,
Option D:	Dimensionality reduction
Q6.	Nominal and ordinal attributes can be collectively referred to as_____ attributes
Option A:	perfect
Option B:	qualitative
Option C:	consistent
Option D:	optimized

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Q7.	The correlation coefficient for two real-valued attributes is -0.85 . What does this value tell you?
Option A:	As the value of one attribute increases the value of the second attribute also increases.
Option B:	The attributes are not linearly related
Option C:	The attributes show a linear relationship
Option D:	As the value of one attribute decreases the value of the second attribute increases.
Q8.	What is the difference between a bar chart and a histogram?
Option A:	There are no gaps between the bars on a histogram.
Option B:	Bar charts represent numbers, whereas histograms represent percentages.
Option C:	A histogram does not show the entire range of scores in a distribution.
Option D:	Bar charts are circular, whereas histograms are square.
Q9.	Classification accuracy is _____
Option A:	A subdivision of a set of examples into a number of classes
Option B:	The task of assigning a classification to a set of examples
Option C:	Group of similar objects that differ significantly from other objects
Option D:	Measure of the accuracy, of the classification of a concept that is given by a certain theory
Q10.	A decision tree is a tree in which every node is either a _____ or a decision node
Option A:	leaf node
Option B:	Root node
Option C:	leaf and Root node
Option D:	Output
Q11.	You are given data about seismic activity in Japan, and you want to predict a magnitude of the next earthquake, this is in an example of
Option A:	Unsupervised learning
Option B:	Dissimilarity reduction
Option C:	Supervised learning
Option D:	Serration
Q12.	Predicting with trees evaluate _____ within each group of data.
Option A:	dissimilarity
Option B:	homogeneity
Option C:	equality
Option D:	heterogeneity
Q13.	Which of the following is true about Residuals ?
Option A:	Medium is better
Option B:	Lower is better
Option C:	Higher is better
Option D:	A or B depend on the situation
Q14.	What is meant by the term 'data quality'?

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Option A:	The resolution of the data.
Option B:	The inherent quality of the data as characterized by its accuracy, precision, bias, level of error, etc.
Option C:	The lineage of the data.
Option D:	The generalization present in the source data.
Q15.	Clustering is also called _____ in some applications because clustering partitions large data sets into groups according to their similarity
Option A:	Data segmentation
Option B:	Data Reduction
Option C:	Data transformation
Option D:	Data normalization
Q16.	A binary variable is _____ if both of its states are equally valuable and carry the same weight.
Option A:	Nominal
Option B:	Subjective
Option C:	Symmetric
Option D:	Asymmetric
Q17.	When an algorithm uses the _____, to measure the distance between clusters, it is sometimes called a nearest-neighbor clustering algorithm.
Option A:	maximum distance
Option B:	average distance
Option C:	mean distance
Option D:	minimum distance
Q18.	Which one is not clustering method?
Option A:	ID3
Option B:	density
Option C:	hierarchical
Option D:	partitioning
Q19.	Which of the following is not null invariant measure(that does not considers null transactions)?
Option A:	all_confidence
Option B:	max_confidence
Option C:	cosine measure
Option D:	lift
Q20.	Why is correlation analysis important?
Option A:	To make apriori memory efficient
Option B:	To weed out uninteresting frequent itemsets
Option C:	To find large number of interesting itemsets
Option D:	To restrict the number of database iterations
Q21.	What is not true about FP growth algorithms?
Option A:	It mines frequent itemsets without candidate generation.

University of Mumbai
Examination 2020 under cluster 4 (PCE)

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.
Q22.	The _____ step eliminates the extensions of (k-1)-itemsets which are not found to be frequent, from being considered for counting support
Option A:	Partitioning
Option B:	Candidate generation
Option C:	Pruning
Option D:	Itemset eliminations
Q23.	_____ in business intelligence allows huge data and reports to be read in a single graphical interface.
Option A:	Dashboard
Option B:	Reports
Option C:	OLAP
Option D:	Warehouse
Q24.	Which of the following is not a component of Business intelligence system?
Option A:	Data Source
Option B:	Data Ware House
Option C:	MTO
Option D:	Business intelligence methodologies
Q25.	Consider building blocks of BI system in the order of first to last and select the correct option
Option A:	Data warehouse, Data exploration, Data sources, Data mining, Optimization and Decisions
Option B:	Data exploration, Data sources , Data warehouse, Data mining, Optimization and Decisions
Option C:	Data sources, Data exploration, Data warehouse, , Data mining, Optimization and Decisions
Option D:	Data sources, Data warehouse, Data exploration, Data mining, Optimization and Decisions

University of Mumbai
Examination 2020 under cluster 4 (PCE)

Program: TE Information Technology

Curriculum Scheme: Rev2016

Examination: Third Year Semester VI

Course Code: ITC602 and Course Name: Data Mining and Business Intelligence

Time: 1 hour

Max. Marks: 50

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