

University of Mumbai
Examination June 2021

Program: Information Technology

Curriculum Scheme: Rev 2016

Examination: Third Year Semester VI

Course Code: ITC602

and Course Name: Data Mining and Business Intelligence

Time: 2-hour

Max. Marks: 80

Q1.	Which of the following is the data mining task?
Option A:	Registering for an online course
Option B:	online money payment through a bank
Option C:	predicting if a student will pass an online course
Option D:	downloading the course certificate from the website
Q2.	Which of the following is not a preprocessing task?
Option A:	Data cleaning
Option B:	Data reduction
Option C:	Data integration
Option D:	Data manipulation
Q3.	Which is not a Data visualization technique?
Option A:	Pictures
Option B:	Graphs
Option C:	charts
Option D:	information graphics
Q4.	Which of the following is NOT example of ordinal attributes?
Option A:	Zip codes
Option B:	Ordered numbers
Option C:	Movie ratings
Option D:	Military ranks
Q5.	Calculate the median {1,2,2,2,3,4,5,6,7,8,9,9}
Option A:	2
Option B:	4
Option C:	8
Option D:	4.5
Q6.	Use smoothing by bin means to smooth (25,32,33)
Option A:	30,30,30
Option B:	32,32,32
Option C:	33,33,33
Option D:	25,25,25
Q7.	The _____ are the attributes of interest describing the tuple in Linear regression
Option A:	Response Variable
Option B:	Slope

University of Mumbai
Examination June 2021

Option C:	Coefficients
Option D:	Predictor Variables
Q8.	Height, width comes under which type of data?
Option A:	Finite
Option B:	Discrete
Option C:	Infinite
Option D:	Continuous
Q9.	difference between supervised learning and unsupervised learning is given by
Option A:	unlike unsupervised learning, supervised learning can be used to detect outliers
Option B:	no difference
Option C:	unlike supervised learning, unsupervised learning can form new classes
Option D:	unlike unsupervised learning, supervised learning needs labeled data
Q10.	Following algorithm is density-based clustering algorithm
Option A:	K-Medoids
Option B:	OPTICS
Option C:	DIANA
Option D:	BIRCH
Q11.	The process of hierarchical clustering is represented by-
Option A:	Decision Tree
Option B:	Dendrogram
Option C:	Flowchart
Option D:	Hierarchical model
Q12.	K medoids is an example of which type of clustering method?
Option A:	Hierarchical
Option B:	Density Based
Option C:	Partition
Option D:	Combination
Q13.	Suppose while performing DBSCAN we randomly choose a point which has less than MinPts number of points in its neighborhood. Which among the following is true for such a point?
Option A:	It is treated as noise, and not considered further in the algorithm
Option B:	It becomes part of its own cluster
Option C:	Depending upon other points, it may later turn out to be a core point
Option D:	Depending upon other points, it may be density connected to other points
Q14.	Which of the following methods explore structures in the data
Option A:	Contextual Outlier Detection and Collective Outlier Detection
Option B:	Classification Outlier Detection and Clustering Outlier Detection
Option C:	Parametric Outlier Detection and Non-Parametric Outlier Detection
Option D:	Distance based Outlier Detection and Density based Outlier Detection
Q15.	The frequency of occurrence of an itemset is also known as _____

University of Mumbai
Examination June 2021

Option A:	Support count
Option B:	Confidence
Option C:	Support
Option D:	Lift
Q16.	What is association rule mining?
Option A:	Same as frequent itemset mining
Option B:	Finding of strong association rules using frequent itemsets
Option C:	Using association to analyze correlation rules d.
Option D:	Finding classification rules using frequent itemsets
Q17.	Which of the following is the application of frequent itemset mining?
Option A:	Real time analysis
Option B:	Fraud analysis
Option C:	Network analysis
Option D:	Market Basket Analysis
Q18.	What do you mean by support($X \rightarrow Y$) is ____
Option A:	The number of transactions where XUY appears divided by the total number of transactions.
Option B:	Total number of transactions containing X or Y
Option C:	Total number of transactions containing X and Y
Option D:	Total number of transactions containing X but not Y
Q19.	Which of the following is not considered as ethical in business Intelligence
Option A:	Using customer data for better customer service
Option B:	Giving customer data to other companies for financial gain
Option C:	Accessing the users' data with for data mining
Option D:	Accessing user's location to give location specific ads n users mobile phone
Q20.	Decisions are ____ when they affect the entire organization or at least a substantial part of it for a long period of time.
Option A:	Operational
Option B:	Technical
Option C:	Tactical
Option D:	Strategic

University of Mumbai
Examination June 2021

Q2.	Solve any Two Questions out of Three	10 marks
A	What is data Preprocessing? Explain different methods	
B	What is clustering? The data for clustering is {2,4,10,12,3,20,11,25} where k=2 cluster the given data using k-means algorithm	
C	Explain multidimensional and multi-level association rules with example	

Q3	Solve any Two Questions out of Three	10 marks
A	Define business intelligence with examples	
B	Partition the given data into 4 bins using binning method and perform smoothing by bin mean, by median, by bin boundaries Data: 11,13,13,15,15,16,19,20,20,21,21,22,23,24,30,40,45,45,45,71,72,73,75	
C	What is an outlier? Describe methods that can be used for outlier analysis	