University of Mumbai
Examination June 2021
Examinations Commencing from 1ˢᵗ June 2021
Program: **Information Technology**
Curriculum Scheme: Rev-2016
Examination: TE Semester VI
Course Code: ITC602 and Course Name: DMBI

Time: 2 hour   Max. Marks: 80

===

Q1.	Choose the correct option for following questions. All the Questions are compulsory and carry equal marks
1.	Which is conclusively produced by Hierarchical Clustering?
Option A:	final estimation of cluster centroids
Option B:	tree showing how nearby things are to each other
Option C:	assignment of each point to clusters
Option D:	all of these
2.	Data cleaning is
Option A:	Large collection of data mostly stored in a computer system
Option B:	The removal of noise errors and incorrect input from a database
Option C:	The systematic description of the syntactic structure of a specific database. It describes the structure of the attributes the tables and foreign key relationships.
Option D:	None of these
3.	Business intelligence (BI) is a broad category of application programs which includes _____
Option A:	Decision support
Option B:	Data mining
Option C:	OLAP
Option D:	All of the above
4.	Which of the following areas are affected by BI?
Option A:	Revenue
Option B:	CRM
Option C:	Sales
Option D:	All of the mentioned
5.	Point out the wrong statement.
Option A:	Data is factual information for analysis
Option B:	BI is a category of database software that provides an interface to help users quickly and interactively scrutinize the results in a variety of dimensions of the data
Option C:	Customer relationship management (CRM) entails all aspects of interaction that a company has with its customer
Option D:	None of the above

6.	Data mining **is?**
Option A:	time variant non-volatile collection of data
Option B:	The actual discovery phase of a knowledge
Option C:	The stage of selecting the right data
Option D:	None of these
7.	——- 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
8.	—— is the out put of KDD
Option A:	Query
Option B:	Data
Option C:	information
Option D:	Useful Information
9.	What is noise?
Option A:	context of KDD and data mining
Option B:	component of a network
Option C:	aspects of a data warehouse
Option D:	None of these
10.	The learning which is used to find the hidden pattern in unlabeled data is called?
Option A:	Unsupervised learning
Option B:	Semi-supervised
Option C:	Supervised learning
Option D:	Reinforcement learning
11.	Which of the following terms is used as a synonym for data mining?
Option A:	knowledge discovery in databases
Option B:	parallel processing in databases
Option C:	data warehousing
Option D:	regression analysis
12.	Agglomerative clustering falls under which type of clustering method?
Option A:	partition
Option B:	hierarchical
Option C:	Divisive clustering
Option D:	none of the above
13.	K means and K-medioids are example of which type of clustering method?
Option A:	Hierarchical
Option B:	probabilistic
Option C:	Divisive
Option D:	partition

14.	A collection of one or more items is called as _____
Option A:	itemset
Option B:	Support
Option C:	Confidence
Option D:	Support value
15.	Frequency of occurrence of an itemset is called as _____
Option A:	Support
Option B:	Confidence
Option C:	Support count
Option D:	rule
16.	**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 mines all frequent patterns by constructing an itemsets
17.	**What do you mean by support(A)?**
Option A:	Total number of transactions containing A
Option B:	Total Number of transactions not containing A
Option C:	Number of transactions containing A / Total number of transactions
Option D:	Number of transactions not containing A / Total number of transactions
18.	**How do you calculate Confidence (A -> B)?**
Option A:	Support(A ∩ B) / Support (A)
Option B:	Support(A ∩ B) / Support (B)
Option C:	Support(A ∪ B) / Support (A)
Option D:	Support(A ∪ B) / Support (B)
19.	**Which of the following is the direct application of frequent itemset mining?**
Option A:	Social Network Analysis
Option B:	Market Basket Analysis
Option C:	Outlier Detection
Option D:	**Intrusion Detection**
20.	**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 so

Q2	Solve any Two Questions out of Three	10 marks each
A	Explain steps in KDD in detail?	

B	What is clustering? Explain k-means clustering algorithm. Suppose the data for clustering- {2,4,10,12,3,20,11,25} consider K=2, Cluster the given data set.
C	Explain sequence mining in transactional DB.

Q3	**Solve any Two Questions out of Three** 10 marks each
A	Explain multidimensional and multilevel association rules with an example.
B	Define "business Intelligence" and "Decision Support System" with example
C	Explain Apriori Algorithm to identify the frequent set than extract strong association rule from these set? Min support=30% and confidence=75% <table><tr><td>TID</td><td>ITEMS</td></tr><tr><td>1</td><td>A,B,D,E,F</td></tr><tr><td>2</td><td>B,C,E</td></tr><tr><td>3</td><td>A,B,D,E</td></tr><tr><td>4</td><td>A,B,C,E</td></tr><tr><td>5</td><td>A,B,C,D,E,F</td></tr><tr><td>6</td><td>B,C,D</td></tr><tr><td>7</td><td>A,B,D,E</td></tr></table>