GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER– VIII EXAMINATION – SUMMER 2020

Subject Code: 2170715   **Date:02/11/2020**

Subject Name: DATA MINING AND BUSINESS INTELLIGENCE

Time: 10:30 AM TO 01:00 PM   **Total Marks: 70**

Instructions:
1. **Attempt all questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**

Q.1	**(a)**	Define the following terms:	**03**
		1) Data warehouse.	
		2) Business Intelligence	
		3) Metadata in Data warehouse	
	(b)	List the major steps involved in the ETL process.	**04**
	(c)	Draw and explain the Data warehouse architecture.	**07**
Q.2	**(a)**	Briefly discuss the schemas for multidimensional databases.	**03**
	(b)	Differentiate between OLTP and OLAP.	**04**
	(c)	Explain Various Data Mining Functionalities with an example.	**07**
		OR	
	(c)	Explain the different issues in data mining.	**07**
Q.3	**(a)**	What is the need for preprocessing the data?	**03**
	(b)	How concept hierarchies are useful in data mining?	**04**
	(c)	Consider a transactional database where 1, 2, 3, 4, 5, 6, 7 are items.	**07**

ID	ITEMS
T_1	1, 2, 3, 5
T_2	1, 2, 3, 4, 5
T_3	1, 2, 3, 7
T_4	1, 3, 6
T_5	1, 2, 4, 5, 6

Suppose the minimum support is 60%. Find all frequent itemsets using Apriori algorithm.

OR

Q.3	**(a)**	What is dimensionality reduction?	**03**
	(b)	Explain about Data Transformation method with suitable example.	**04**
	(c)	Discuss the variations of the Apriori algorithm to improve the efficiency.	**07**
Q.4	**(a)**	What is meant by multidimensional association rules?	**03**
	(b)	Discuss the Information gain as attribute selection measure.	**04**
	(c)	Differentiate classification and prediction. State the issues regarding classification and prediction.	**07**
		OR	
Q.4	**(a)**	What is meant by Maximal Frequent Item Set?	**03**
	(b)	Discuss the Gain ratio as attribute selection measure.	**04**

	(c)	Why is naïve Bayesian classification called "naïve"? Briefly outline the major ideas of naïve Bayesian classification.	**07**
Q.5	**(a)**	List out the General applications of Clustering.	**03**
	(b)	What is Big Data? What is big data analytic?	**04**
	(c)	How the data mining will be used in the retail industry?	**07**

OR

Q.5	**(a)**	Define the web mining.	**03**
	(b)	Discuss the main features of Hadoop Distributed File System.	**04**
	(c)	How the data mining will be used in the telecommunication industry?	**07**

****