Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–VI(NEW) EXAMINATION – WINTER 2022

Subject Code:3160714 Date:16-12-2022

Subject Name:Data Mining

Time:02:30 PM TO 05:00 PM Total Marks:70

Instructions:

1. **Attempt all questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**
4. **Simple and non-programmable scientific calculators are allowed.**

			Marks
Q.1	**(a)**	Compare descriptive and predictive data mining.	**03**
	(b)	Explain the data mining functionalities.	**04**
	(c)	Explain major requirements and challenges in data mining.	**07**
Q.2	**(a)**	What do you mean by concept hierarchy?	**03**
	(b)	Explain the smoothing techniques.	**04**
	(c)	What is Data Cleaning? Describe various methods of Data Cleaning.	**07**
		OR	
	(c)	Explain about the different Data Reduction techniques.	**07**
Q.3	**(a)**	What are the techniques to improve the efficiency of Apriori algorithm?	**03**
	(b)	What is an Itemset? What is a Frequent Itemset?	**04**
	(c)	For the given data	**07**

Transaction ID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	Hot Dogs, Coke, Chips

Find the frequent itemsets and generate association rules on this. Assume that minimum support threshold (s = 33.33%) and minimum confident threshold (c = 60%).

OR

Q.3	**(a)**	Describe the different classifications of Association rule mining.	**03**
	(b)	What is meant by Reduced Minimum Support?	**04**
	(c)	Explain the steps of the “Apriori Algorithm” for mining frequent itemsets with suitable example.	**07**
Q.4	**(a)**	What are Bayesian Classifiers?	**03**
	(b)	What are the hierarchical methods used in classification?	**04**
	(c)	Describe in detail about Rule based Classification.	**07**

OR

Q.4	**(a)**	What is attribute selection measure?	**03**
	(b)	What is the difference between supervised and unsupervised learning scheme.	**04**
	(c)	Describe the issues regarding classification and prediction. Write an algorithm for decision tree.	**07**
Q.5	**(a)**	List the requirements of clustering in data mining.	**03**
	(b)	Differentiate Agglomerative and Divisive Hierarchical Clustering?	**04**
	(c)	Write a short note: Web content mining.	**07**

OR

Q.5	**(a)**	What is meant by hierarchical clustering?	**03**
	(b)	Illustrate strength and weakness of k-mean in comparison with k-medoid algorithm.	**04**
	(c)	Write a short note: Web usage mining.	**07**

****