Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER– VI (NEW) EXAMINATION – WINTER 2021

Subject Code:3160714 Date:02/12/2021

Subject Name:Data Mining

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.**
4. **Simple and non-programmable scientific calculators are allowed.**

			MARKS
Q.1	**(a)**	Justify the importance of data mining.	**03**
	(b)	Differentiate OLTP and data warehouse.	**04**
	(c)	Briefly discussed steps of KDD process.	**07**
Q.2	**(a)**	Explain data reduction and dimensionality reduction?	**03**
	(b)	What do you mean by correlation analysis? Justify its importance.	**04**
	(c)	List common task involved in the data pre-processing. Explain briefly any four tasks of data pre-processing with suitable example.	**07**

OR

	(c)	Define the following: concept description, support, confidence, strong association rules, data generalization, and unsupervised learning.	**07**
Q.3	**(a)**	How the classification is differs from the prediction? Explain phases of classification.	**03**
	(b)	Attribute income have minimum value of 12000 INR and maximum value of 98000 INR. Normalize income value of 73600 INR, (i) Using min-max normalization in the range of [0,1] (ii) Using z-score normalization. Take mean value of income as 54000 and standard deviation is 16000.	**04**
	(c)	Using Apriori algorithm, find all frequent itemsets for following transaction data. (Take min_sup=60% and min_conf=80%)	**07**

ID	Items
1	{M,O,N,K,E,Y}
2	{D,O,N,K,E,Y
3	{M,A,K,E}
4	{M,U,C,K,Y}
5	{C,O,O,K,I,E}

OR

Q.3	**(a)**	What is the use of proximity measures? Explain any one proximity measures with equation.	**03**
	(b)	Explain Bayesian learning and inference with suitable example.	**04**
	(c)	List the accuracy parameters used for the performance evaluation of classification and discuss any five parameters with appropriate example.	**07**
Q.4	**(a)**	Differentiate supervised and unsupervised learning.	**03**
	(b)	Explain logistic regression with appropriate example.	**04**

	(c)	Explain working of decision tree algorithm with suitable example.	**07**

OR

Q.4	**(a)**	Differentiate agglomerative and divisive methods of clustering.	**03**
	(b)	What do you mean by perceptron? Discuss single-layer and multi layer perceptron.	**04**
	(c)	Explain K-means clustering algorithm and prove that outlier adversely affect the performance of algorithm.	**07**
Q.5	**(a)**	Give strength and weakness of k-means in comparison of k-medoids algorithm.	**03**
	(b)	What is outlier? Why outlier mining is important?	**04**
	(c)	Write about different clustering approaches with their strength and weakness.	**07**

OR

Q.5	**(a)**	Briefly explain the spatial data mining and temporal mining.	**03**
	(b)	Discuss any four data mining features available in the WEKA.	**04**
	(c)	How data mining is useful for web mining. Discuss any four web mining applications.	**07**

****

Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

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

Subject Code:3160714 Date:08/06/2022

Subject Name:Data Mining

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.**
4. **Simple and non-programmable scientific calculators are allowed.**

			Marks
Q.1	**(a)**	What are the types of data?	**03**
	(b)	Compare descriptive and predictive data mining	**04**
	(c)	Draw and explain the data mining architecture.	**07**
Q.2	**(a)**	What is dimensionality reduction?	**03**
	(b)	What are the types of concept hierarchies?	**04**
	(c)	What is Data Cleaning? Describe various methods of Data Cleaning.	**07**
		OR	
	(c)	Discuss issues to be considered during data integration.	**07**
Q.3	**(a)**	What is meant by association rule?	**03**
	(b)	How is association rules mined from large databases?	**04**
	(c)	Explain the various criteria for the classification of frequent pattern mining.	**07**
		OR	
Q.3	**(a)**	List two interesting measures for association rules.	**03**
	(b)	What is meant by multidimensional association rules?	**04**
	(c)	Write short notes on Maximal Frequent Item Set &Closed Frequent Item Set.	**07**
Q.4	**(a)**	What is an outlier?	**03**
	(b)	What is Bayesian theorem?	**04**
	(c)	Demonstrate how Bayesian classification helps in predicting class membership probabilities.	**07**
		OR	
Q.4	**(a)**	Differentiate classification and prediction.	**03**
	(b)	What is the difference between “supervised” and unsupervised” learning scheme.	**04**
	(c)	Explain the issues regarding the classification and prediction.	**07**
Q.5	**(a)**	What is temporal mining?	**03**
	(b)	Explain web usage mining.	**04**
	(c)	Discuss the K-means clustering algorithm using examples.	**07**
		OR	
Q.5	**(a)**	What is multimedia mining?	**03**
	(b)	Explain web content mining.	**04**
	(c)	What do you meant by Clustering? Explain the requirements used in Clustering?	**07**

****