GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER–VI (NEW) EXAMINATION – SUMMER 2023

Subject Code:3164305   **Date:06-07-2023**

Subject Name: Data Warehousing and 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.

Q.1	**(a)**	What are the different steps involved in the KDD process for data mining?	**07**
	(b)	What is a data warehousing schema? Explain the difference between star schema and snowflake schema.	**07**
Q.2	**(a)**	What are support and confidence measures in association rule mining, and how are they calculated?	**07**
	(b)	Explain three tier data warehouse architecture in details.	**07**
		OR	
	(b)	What is regression? Explain the difference between simple linear regression and multiple linear regression.	**07**
Q.3	**(a)**	Discuss the key steps involved in the apriori algorithm for discovering frequent itemsets.	**07**
	(b)	Explain the concept of pruning in decision trees and its impact on model performance.	**07**
		OR	
Q.3	**(a)**	Discuss Bayesian classification.	**07**
	(b)	What is dimensionality reduction, and what is its main objective in data mining?	**07**
Q.4	**(a)**	What are some common techniques for handling missing data?	**07**
	(b)	Discuss the concept of hierarchical clustering and the different linkage methods used to form clusters.	**07**
		OR	
Q.4	**(a)**	Discuss the concept of backpropagation and how it is used to train neural network classifiers.	**07**
	(b)	Compare and contrast hierarchical clustering and partition-based clustering algorithms.	**07**
Q.5	**(a)**	Write a short note on spatial data mining.	**07**
	(b)	What is clustering in data mining, and what are its main objectives?	**07**
		OR	
Q.5	**(a)**	Describe the k-means clustering algorithm and its characteristics.	**07**
	(b)	What is web mining, and what are its applications?	**07**

****