GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER–VI EXAMINATION – SUMMER 2025

Subject Code: 3164305   **Date: 22-05-2025**

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**

			MARKS
Q.1*	**(a)**	Discuss the concept hierarchy.	**03**
	(b)	Differentiate between OLTP and OLAP.	**04**
	(c)	With the help of a neat diagram explain the 3-tier architecture of a data warehouse.	**07**
Q.2	**(a)**	Explain various features of a data warehouse.	**03**
	(b)	Discuss possible ways for the integration of a Data Mining system with a Database or Data Warehouse system.	**04**
	(c)	Explain Star, Snowflake, and Fact Constellation schemas of a data warehouse with a suitable example.	**07**
		OR	
	(c)	With the help of a suitable example, illustrate the OLAP operations: 'drill-down,' 'roll-up,' 'slice,' and 'dice.'.	**07**
Q.3	**(a)**	Explain Variance and Standard Deviation.	**03**
	(b)	List and describe methods for handling missing values in data cleaning.	**04**
	(c)	List and Explain Data Mining Task Primitives.	**07**
		OR	
Q.3	**(a)**	Why data preprocessing is required?	**03**
	(b)	Explain min-max and Z-score normalization.	**04**
	(c)	Enlist data reduction strategies and explain any two.	**07**
Q.4	**(a)**	Differentiate Clustering and classification.	**03**
	(b)	Discuss Data Smoothing by Binning using a suitable example.	**04**
	(c)	Discuss the key steps involved in the apriori algorithm for discovering frequent itemsets.	**07**
		OR	
Q.4	**(a)**	Explain Support and Confidence measures using proper example.	**03**
	(b)	Compare supervised with unsupervised learning.	**04**
	(c)	What are the limitations of the Apriori algorithm? Discuss any two methods to improve the efficiency of Apriori-based mining.	**07**
Q.5	**(a)**	Explain web mining using example.	**03**
	(b)	How does k-means clustering differ from k-medoids clustering?	**04**
	(c)	Explain the concept of pruning in decision trees and its impact on model performance.	**07**
		OR	
Q.5	**(a)**	Explain Spatial mining using example.	**03**
	(b)	Explain Linear regression with an example.	**04**
	(c)	Explain k-Means clustering algorithm in detail.	**07**