

GUJARAT TECHNOLOGICAL UNIVERSITY**BE- SEMESTER- VI EXAMINATION – SUMMER 2026****Subject Code:3164305****Date:22-05-2026****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 ETL.	03
	(b) Discuss star and snowflake schema.	04
	(c) Compare OLAP and OLTP.	07
Q.2	(a) Discuss data warehouse and data mart.	03
	(b) Mention operations of data cube and discuss any two in detail.	04
	(c) Discuss KDD process.	07
	OR	
	(c) Discuss architecture of a typical data mining system.	07
Q.3	(a) Discuss CUR decomposition.	03
	(b) Discuss outlier analysis using suitable example.	04
	(c) Discuss how to integrate or couple the DM system with a database (DB) system and/or a data warehouse (DW) system	07
	OR	
Q.3	(a) Discuss Dimensionality Reduction.	03
	(b) Discuss concept hierarchy generation using suitable example.	04
	(c) In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem.	07
Q.4	(a) Compare Classification vs. prediction.	03
	(b) Discuss Linear and nonlinear regression.	04
	(c) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.	07
	Give the five-number summary of the data.	
	OR	
Q.4	(a) Discuss Decision Tree.	03
	(b) Discuss evaluation measures of a classifier or predictor	04
	(c) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70	07
	What is the mode of the data? Comment on the data's modality. Also find the first quartile (Q1) and the third quartile (Q3) of the data.	
Q.5	(a) Discuss Text Data Mining.	03
	(b) Evaluation of Clustering Algorithm.	04

	(c) Dicuss Apriori algorithm.	07
	OR	
Q.5	(a) Discuss Spatial Data Mining.	03
	(b) Discuss correlation analysis.	04
	(c) Discuss k-means clustering.	07
