Enrolment No. /Seat No.: _____

GUJARAT TECHNOLOGICAL UNIVERSITY

MCA INTEGRATED – SEMESTER VI- EXAMINATION –SUMMER-2025

Subject Code: 2668603 Date: 06/06/2025

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. **Use of simple calculators and non-programmable scientific calculators are permitted.**

Q.1	**(a)**	Define what a Data Warehouse is and explain its key characteristics.	**07**
	(b)	Describe the multi-tiered architecture of Data Warehousing.	**07**
Q.2	**(a)**	Briefly outline the processes involved in Extraction, Transformation, and Loading (ETL) and the role of the Metadata Repository.	**07**
	(b)	Explain the different models of Data Warehousing, including Enterprise Warehouse, Datamart, and Virtual Warehouse.	**07**
		OR	
	(b)	Define Data Mining and discuss its importance in today's data-driven environment.	**07**
Q.3	**(a)**	Discuss the applications targeted by Data Mining, particularly in Business Intelligence and Web Search Engines.	**07**
	(b)	Describe the types of data that can be mined and the various patterns that can be identified through Data Mining.	**07**
		OR	
Q.3	**(a)**	Define Market-Basket Analysis and explain its significance in understanding consumer behavior.	**07**
	(b)	Explain the Apriori algorithm for Frequent Itemset Mining, detailing its process and importance.	**07**
Q.4	**(a)**	Discuss how association rules are generated from frequent itemsets and the significance of this process.	**07**
	(b)	Define Classification and its importance in data mining and predictive analytics.	**07**
		OR	
Q.4	**(a)**	Explain Bayesian Classification, focusing on Bayes' Theorem and the Naïve Bayesian Classification approach.	**07**
	(b)	Discuss Rule-based Classification, detailing how IF-THEN rules are used and the processes of Rule Extraction and Rule Induction using a Sequential Covering Algorithm.	**07**
Q.5	**(a)**	Define Cluster Analysis and explain its significance in data mining and pattern recognition. What are the essential requirements for effective cluster analysis?	**07**
	(b)	Provide an overview of basic clustering methods, including the main characteristics of Partitioning Methods such as K-Means and K-Medoids.	**07**
		OR	
Q.5	**(a)**	Compare and contrast Hierarchical Methods, specifically Agglomerative versus Divisive clustering, and describe the process of Hierarchical Clustering.	**07**
	(b)	Explain various Distance Measures used in clustering algorithms and discuss advanced methods like BIRCH and Chameleon for hierarchical clustering.	**07**

****