Enrolment No./Seat No_____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–VI EXAMINATION – SUMMER 2025

Subject Code: 3160714 Date: 28-05-2025

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

			Marks
Q.1	**(a)**	What is the difference between discrimination and classification?	**03**
	(b)	Describe the steps involved in data mining when viewed as a process of knowledge discovery.	**04**
	(c)	Draw and explain the data mining architecture.	**07**
Q.2	**(a)**	Discuss issues to consider during data integration.	**03**
	(b)	Describe various methods for handling the missing values in the dataset.	**04**
	(c)	Use the methods below to normalize the following group of data: 200, 300, 400, 600, 1000 (a) min-max normalization by setting min = 0 and max = 1 (b) Calculate z-score normalization (c) z-score normalization using the mean absolute deviation instead of standard deviation (d) normalization by decimal scaling	**07**
		OR	
	(c)	For the following data (in increasing order) for the attribute age: 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. (a) Use smoothing by bin means to smooth these data, using a bin depth of 3. Illustrate your steps. Comment on the effect of this technique for the given data. (b) How might you determine outliers in the data? (c) What other methods are there for data smoothing?	**07**
Q.3	**(a)**	What is meant by association rule?	**03**
	(b)	Write a short note on: Web content mining.	**04**
	(c)	Explain about the Apriori algorithm for finding frequent item sets with an example.	**07**
		OR	
Q.3	**(a)**	What is meant by Maximal Frequent Item Set?	**03**
	(b)	Write a short note on: Web usage mining.	**04**
	(c)	Which patterns are interesting? Explain based on Pattern Evaluation Methods.	**07**

Q.4	**(a)**	What is the difference between “supervised” and unsupervised” learning?	**03**
	(b)	What are the attribute selection measures? Explain any one in details.	**04**
	(c)	Discuss about Decision tree induction algorithm with an example.	**07**

OR

Q.4	**(a)**	What do you meant by Bayesian Classification?	**03**
	(b)	Describe the issues regarding classification and prediction.	**04**
	(c)	Discuss about Neural Network-Based Algorithms.	**07**

Q.5	**(a)**	What are the requirements of clustering?	**03**
	(b)	Discuss about Outlier Detection.	**04**
	(c)	Discuss about k-nearest neighbor classification algorithm with an example?	**07**

OR

Q.5	**(a)**	State the categories of clustering methods?	**03**
	(b)	Explain the concept hierarchy.	**04**
	(c)	Explain in detail about partitioned Clustering method.	**07**

****