

GUJARAT TECHNOLOGICAL UNIVERSITY

BE- SEMESTER- VI EXAMINATION – SUMMER 2026

Subject Code:3161610

Date:01-06-2026

Subject Name: Data Warehousing and 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) Explain various types of metadata in data warehouse. **03**
(b) Give the benefits of data warehousing. **04**
(c) What is cuboid? Explain following OLAP Operations with examples. **07**
(1) drill down (2) roll up (3) slice (4) dice
- Q.2** (a) Differentiate OLAP and OLTP **03**
(b) What is data mining? What are the major challenges of mining huge amount of data? **04**
(c) Draw and explain data warehouse architecture. **07**
- OR**
- (c) Explain the various stage of KDD process. **07**
- Q.3** (a) Write importance of preprocessing in data mining. **03**
(b) Explain different types of data on which mining can be performed. **04**
(c) Explain various schema for multidimensional database. **07**
- OR**
- Q.3** (a) Give the importance of dimensionality reduction. **03**
(b) What is normalization? Explain min-max normalization with suitable example. **04**
(c) Explain data smoothing methods to divide given data into bins of size 3 by bin means, by bin medians and by bin boundaries. **07**
Consider the data: 10, 2, 19, 18, 20, 18, 25, 28, 22
- Q.4** (a) Differentiate Fact table vs. Dimension table. **03**
(b) Explain support and confidence with example. **04**
(c) Find frequent item-sets and association rules using Apriori algorithm on the following data set with minimum support count is 2 and minimum confidence is 75%. **07**

Sr. No	TID	List of items
1	T1	A,B,E
2	T2	B,D
3	T3	B,C
4	T4	A,B,D
5	T5	A,C
6	T6	B,C
7	T7	A,C
8	T8	A,B,C,E
9	T9	A,B,C

OR

Q.4	(a) Differentiate supervised and unsupervised learning.	03
	(b) Explain various issues of classification.	04
	(c) Explain Information Gain, Gain Ratio, Gini Index	07
Q.5	(a) Give differences between linear and logistic regression.	03
	(b) What is outlier? Discuss different methods for outlier detection.	04
	(c) Explain k-means clustering algorithm.	07
OR		
Q.5	(a) Give the differences between Spatial and Temporal Data Mining.	03
	(b) Write short note on text mining.	04
	(c) What is web mining? Explain types of web mining.	07

.