

GUJARAT TECHNOLOGICAL UNIVERSITY
M.SC(IT)- INTEGRATED – SEMESTER V- EXAMINATION –SUMMER-2025

Subject Code: 1350503**Date: 15/05/2025****Subject Name: Data Warehouse and Mining****Time:02:30 PM TO 05:00 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.

		Marks
Q.1	(a) What is Data Warehouse? Who & Why it need?	03
	(b) How to manage data warehouse? Explain with its methodology.	04
	(c) Draw ETL architecture & Explain in details.	07
Q.2	(a) Define the term “data mining”. List the major issues in data mining.	03
	(b) What are the limitations of the Apriori approach for mining?	04
	(c) State the Apriori Property. Find frequent item-sets and association rules using Apriori algorithm on the following data set with minimum support count is 2 and minimum confidence=75%.	07

Sr.No	TID	List of items_IDs
1	T100	I1,I2,I5
2	T200	I2,I4
3	T300	I2,I3
4	T400	I1,I2,I4
5	T500	I1,I3
6	T600	I2,I3
7	T700	I1,I3
8	T800	I1,I2,I3,I5
9	T900	I1,I2,I3

OR

	(c) Explain multi-dimensional association rule in details.	07
Q.3	(a) What is classification? Give the difference between classification and predication.	03
	(b) What is Information gain and Gain ratio?	04
	(c) Explain Baye’s Theorem and Naïve Bayesian Classification.	07

OR

Q.3	(a) What is Market Basket Analysis? Explain Association Rules with Confidence & Support.	03
	(b) Give the difference between OLAP and OLTP.	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) What is clustering? Why clustering is an un-supervised learning? **03**
(b) Write a short note on text mining. **04**
(c) Explain k-means algorithm of clustering. **07**
- OR**
- Q.4** (a) Give the Difference between Spatial and Temporal Data Mining. **03**
(b) Briefly explain Linear and Non-linear regression. **04**
(c) Explain neural network based algorithms in details. **07**
- Q.5** (a) What is DBSCAN? Explain in brief. **03**
(b) Explain clustering partitioning method in brief. **04**
(c) Explain different types of Web Mining with example. **07**
- OR**
- Q.5** (a) Why web mining is needed? Explain. **03**
(b) Explain web structure mining in brief. **04**
(c) What is logistic regression? Explain in details. **07**
