

GUJARAT TECHNOLOGICAL UNIVERSITY**BE- SEMESTER-VI EXAMINATION – WINTER 2025****Subject Code:3161610****Date:25-11-2025****Subject Name:Data Warehousing 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. Simple and non-programmable scientific calculators are allowed.

MARKS

- Q.1** (a) What is the importance of data analysis? **03**
 (b) What is data warehousing? List the characteristics of data warehouse. **04**
 (c) Explain the three tier architecture of data warehouse. **07**
- Q.2** (a) What is OLAP? How OLAP is differ from OLTP? **03**
 (b) What is data cube? Explain rollup operation on data cube. **04**
 (c) Explain different schema of data warehouse with suitable diagram. **07**

OR

- (c) Define: mean, median, mode, variance and standard deviation. **07**
- Q.3** (a) List different methods for data discretization and explain any one in detail. **03**
 (b) Why data smoothing is required? Perform smoothing by bin means, by bin medians and by bin boundaries on the given data with bin size is 3. Consider the data for price (in dollars): 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 33. **04**
 (c) Explain the process of extracting the knowledge from the database. **07**

OR

- Q.3** (a) Explain Association Rules with Confidence & Support. **03**
 (b) Write a note on data cleaning process for missing value and noisy data treatment. **04**
 (c) Explain various methods for normalization. **07**
- Q.4** (a) What are the limitations of the Apriori approach for mining? **03**
 (b) What is data mining? How data mining techniques useful in super market? **04**
 (c) State the Apriori Property. Find frequent item-sets and association rules using Apriori algorithm on the following data set with minimum support is 60% and minimum confidence=80%. **07**

Sr.No	TID	List of items
1	T100	M, O, N, K, E, Y
2	T200	D, O, N, K, E, Y
3	T300	M, A, K, E
4	T400	M, U, C, K, Y
5	T500	C, O, O, K, I, E

OR

- Q.4** (a) Give the difference between supervised and un-supervised learning. **03**
 (b) Explain attribute selection measures in decision tree. **04**
 (c) Explain Baye's Theorem and Naïve Bayesian Classification. **07**

- Q.5** (a) Give the importance of the pruning in association rule mining. **03**
(b) What is outlier? Discuss different methods for outlier detection. **04**
(c) Briefly explain Linear and Non-linear regression. **07**

OR

- Q.5** (a) Explain text mining in brief. **03**
(b) What is web mining? Explain types of web mining. **04**
(c) Explain k-mean clustering algorithm. **07**
