

**GUJARAT TECHNOLOGICAL UNIVERSITY****BE - SEMESTER-VI EXAMINATION – SUMMER 2025****Subject Code: 3161610****Date:30-05-2025****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**

- |            |                                                                                                                                                          |           |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Q.1</b> | (a) Compare data mart and data warehouse.                                                                                                                | <b>03</b> |
|            | (b) A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data – Justify.                                      | <b>04</b> |
|            | (c) What is Cuboid? Explain various OLAP operations on data cube with example.                                                                           | <b>07</b> |
| <b>Q.2</b> | (a) Differentiate Fact table vs. Dimension table.                                                                                                        | <b>03</b> |
|            | (b) briefly explain classification and prediction.                                                                                                       | <b>04</b> |
|            | (c) Explain the KDD process in detail.                                                                                                                   | <b>07</b> |
|            | <b>OR</b>                                                                                                                                                |           |
|            | (c) Explain the major issues in data mining.                                                                                                             | <b>07</b> |
| <b>Q.3</b> | (a) Briefly discuss the need for data preprocessing.                                                                                                     | <b>03</b> |
|            | (b) Explain the following terms with suitable example.<br>1) Data Integration 2) Data Transformation                                                     | <b>04</b> |
|            | (c) Draw the diagram and describe the architecture of a data mining system.                                                                              | <b>07</b> |
|            | <b>OR</b>                                                                                                                                                |           |
| <b>Q.3</b> | (a) Explain parametric and non-parametric methods of data reduction.                                                                                     | <b>03</b> |
|            | (b) What is data cleaning? How to handle the missing value in data cleaning?                                                                             | <b>04</b> |
|            | (c) What is noise? Describe the possible reasons for noisy data. Explain the different techniques to remove the noise from data.                         | <b>07</b> |
| <b>Q.4</b> | (a) Briefly explain Linear and Non-linear regression.                                                                                                    | <b>03</b> |
|            | (b) What is market basket analysis? Explain the two measures of rule interestingness: support and confidence.                                            | <b>04</b> |
|            | (c) Explain the steps of the Apriori Algorithm for mining Frequent Itemsets with Candidate Generation. Use a suitable example to illustrate your answer. | <b>07</b> |
|            | <b>OR</b>                                                                                                                                                |           |
| <b>Q.4</b> | (a) Discuss : training and test dataset.                                                                                                                 | <b>03</b> |
|            | (b) What is classification? Explain classification as a two-step process with a diagram.                                                                 | <b>04</b> |
|            | (c) Explain how the accuracy of a classifier/predictor can be measured.                                                                                  | <b>07</b> |
| <b>Q.5</b> | (a) Explain text mining using example.                                                                                                                   | <b>03</b> |
|            | (b) Write a short note on tree pruning.                                                                                                                  | <b>04</b> |
|            | (c) Explain the working of the k-Means clustering algorithm.                                                                                             | <b>07</b> |
|            | <b>OR</b>                                                                                                                                                |           |
| <b>Q.5</b> | (a) Write a note on web mining.                                                                                                                          | <b>03</b> |
|            | (b) Explain the following as attribute selection measures: (i) Information Gain (ii) Gain Ratio.                                                         | <b>04</b> |
|            | (c) Discuss Bayesian classification.                                                                                                                     | <b>07</b> |

\*\*\*\*\*