

GUJARAT TECHNOLOGICAL UNIVERSITY

BE- SEMESTER- VI EXAMINATION – SUMMER 2026

Subject Code:3160714

Date:29-05-2026

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.

- Q.1**
- (a) Define Schema. Explain the following schemas with suitable example. 03
1) Star 2) Snowflakes 3) Constellations
- (b) A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data – Justify. 04
- (c) Describe the steps involved in data mining when viewed as a process of knowledge discovery. 07
- Q.2**
- (a) What is Concept Hierarchy? List and explain types of Concept Hierarchy. 03
- (b) Differentiate OLAP and OLTP. 04
- (c) Explain the major issues in data mining 07
- OR**
- (c) Explain the need for data smoothing during pre-processing and discuss data smoothing by Binning. 07
- Q.3**
- (a) Briefly explain Linear and Non-linear regression. 03
- (b) Explain measures for finding rule interestingness (support,confidence) with example. 04
- (c) Compare association and classification. Briefly explain associative classification with suitable example. 07
- OR**
- Q.3**
- (a) Discuss following terms. 03
1) Supervised learning 2) Correlation analysis
- (b) Write a note on incremental Association Rule Mining. 04
- (c) Explain Hash-based technique to improve efficiency of Apriori algorithm. 07
- Q.4**
- (a) Define “clustering”? Mention any two applications of clustering. 03
- (b) Draw the topology of a multilayer, feed-forward Neural Network. 04
- (c) Write and discuss the algorithm which is used to generate frequent itemsets using an iterative level-wise approach based on candidate generation. 07
- OR**
- Q.4**
- (a) Explain web mining using example. 03
- (b) What is classification? Explain classification as a two step process with diagram. 04
- (c) Discuss the main features of Hadoop Distributed File System. 07
- Q.5**
- (a) Explain Hadoop storage – HDFS. 03
- (b) How K-Mean clustering method differs from K-Medoid clustering method? 04
- (c) Explain mapreduce. Explain any example using mapreduce. 07

OR

- Q.5**
- | | | |
|------------|---|-----------|
| (a) | Explain data mining application for fraud detection. | 03 |
| (b) | Dicuss the Applications of Distributed and parallel Data Mining. | 04 |
| (c) | What is web log? Explain web structure mining and web usage mining in detail. | 07 |
