

GUJARAT TECHNOLOGICAL UNIVERSITY
BACHELOR OF ENGINEERING – SEMESTER 6 – EXAMINATION – SUMMER 2026

Subject Code: 3163209

Date: 03/06/2026

Subject Name: Data Mining and Business Intelligence

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) Discuss data mining issues. Explain any one with example. 03
- (b) Difference between star schema and snowflake schema. 04
- (c) Describe the three-tier architecture of Data Warehouse with suitable diagram in detail. 07

- Q.2**
- (a) Define Business Intelligence (BI). How does it differ from traditional data processing systems? 03
- (b) Dataset of student exam scores: 45, 48, 50, 52, 53, 55, 58, 60, 61, 65, 68, 70. Partition the data into 3 bins by equal-width partitioning. Smooth the data using the bin mean method. 04
- (c) Explain any four OLAP operations with suitable examples. 07

OR

- Q.2** (c) Training & Promotion data is given below: 07

	Promoted	Not promoted	Total
Received training	120	80	200
No training	60	140	200
Total	180	220	400

H_0 : Promotion is independent of training. H_1 : Promotion depends on training.

At: $\alpha = 0.05$, $\text{dof} = 1$, the $\chi^2 = 3.841$.

Compute Chi-square test value for given data and Check the independence of an employee who took training and got promotion.

- Q.3**
- (a) What is an outlier? How outliers can be processed, explain it with example. 03
- (b) Define the terms: Range, Quartiles, Inter-Quartile Range. Give equation for Inter-Quartile Range. 04
- (c) Explain Rule-Base Classification with suitable example in detail. 07

OR

- Q.3**
- (a) Define: Gini Index, Gain Ratio and Post Pruning. 03
- (b) Explain how to predict a class label using Naïve Bayesian classification with appropriate example. 04

- (c) Find the frequent itemset using Apriori algorithm for given transactional database. 07
 Generate strong association rules.
 Consider Minimum support count=2 and Minimum confidence=70%.

TID	Items
1	Phone, Cover, Charger
2	Phone, Earphones, Charger
3	Phone, Cover, Screen-Guard
4	Phone, Charger, Power-bank
5	Phone, Cover, Earphones

- Q.4** (a) List various forms of Data Pre-processing. Explain data reduction in detail with suitable examples. 03
 (b) Compare ordinal attributes and nominal attributes. 04
 (c) Normalize all the values of given dataset using Min-Max normalization. 07
 13, 18, 25, 30, 40.
 Transform onto scale as [0:1].

OR

- Q.4** (a) Explain the single layer perceptron with suitable diagram. 03
 (b) Transform the vector using HAAR Wavelet transformation: [2,4,6,7,3,1,0,1]. 04
 (c) Solve Linear regression for the values of x and their corresponding values of y are shown in the table below. 07
 Draw the regression line for $y=mx+c$ using linear regression. Also calculate the R-square (Least Squares) error for same.

X (Years Experience)	Y (Salary in Dollar)
3	30
8	57
9	64
13	72
3	36
6	43
11	59
21	90
1	20
16	83

- Q.5** (a) Explain the Euclidean distance metrics used in clustering with example. 03
 (b) Explain the steps involved in building a decision tree for classification. Give appropriate example. 04
 (c) Cluster the following 2D points into 2 clusters using K-means algorithm 07
 Points: (2,3), (3,4), (5,8), (8,8), (6,3), (7,4).
 Initial centroids given as $C1 = (2,3)$ and $C2 = (5,8)$

OR

- Q.5** (a) List two data mining applications in e-commerce with suitable examples. 03
 (b) Draw and explain the Map-Reduce architecture with an example of word count or log analysis. 04

- (c) Explain the architecture of HDFS. Draw a labeled diagram and describe the roles of Name-Node, Data-Node, and Secondary Name-Node.

07
