GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER–VI EXAMINATION – SUMMER 2025

Subject Code:3163209   **Date:02-06-2025**

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.**

			MARKS
Q.1	**(a)**	Define: Data Mart & Business Intelligence.	**03**
	(b)	Perform feature wise comparison between OLAP & OLTP.	**04**
	(c)	Draw and explain typical data mining architecture in detail with suitable diagram.	**07**
Q.2	**(a)**	State the benefits of integrating Business intelligence with Data Warehousing.	**03**
	(b)	State various data repositories on which data can be performed. Explain any one with example.	**04**
	(c)	Explain three architecture of data warehouse with the help of neat diagram.	**07**
		OR	
	(c)	Explain major issues in data mining.	**07**
Q.3	**(a)**	Define Support & Confidence.	**03**
	(b)	Explain nominal and ordinal attribute types with suitable example.	**04**
	(c)	Explain Bayesian classification with suitable example.	**07**
		OR	
Q.3	**(a)**	Define: Information gain and Tree Pruning.	**03**
	(b)	What is a decision tree algorithm, and how does it work?	**04**
	(c)	Find frequent itemset and generate strong association rules using Apriori algorithm for give transactional database. Consider minimum support = 2 is the occurrence of the items in the transactional datasets and minimum confidence is 75%.	**07**

TID	List of Items
T1	A, C, D
T2	B, C, E
T3	A, B, C, E
T4	B, E

Q.4	**(a)**	Explain ETL process with suitable diagram.	**03**
	(b)	Define clustering. What are the typical requirements of clustering in data mining.	**04**
	(c)	Suppose that the data for analysis includes the attribute Marks as: 8, 10, 15 20. Use min-max normalization to transform the values for Marks onto the range [0:1].	**07**

OR

Q.4	**(a)**	State various data transformation methods. Explain any one in detail.	**03**
	(b)	Two stocks A & B having values in one week as: - (2,5), (3,8), (5,10), (4,11), (6 14). It the stocks are affected by same industry trends; will their prices rise together or falls together. Prove using covariance analysis.	**04**
	(c)	Suppose we want to investigate the relationship between the number of hours studied and the score obtained in a test. We collect data from 9 students and obtain the following results:	**07**

Hours Studied	Score
2	50
3	60
4	70
5	75
6	80
7	85
8	90
9	95
10	100

Solve above data using linear regression to find the value of R-squared metric.

Q.5	**(a)**	What is Neural Network. State their types.	**03**
	(b)	Compare hierarchical and partitioning clustering.	**04**
	(c)	Draw & explain HDFS architecture with suitable example.	**07**

OR

Q.5	**(a)**	Explain data mining application fraud detection with example.	**03**
	(b)	What is Big data. State applications of big data.	**04**
	(c)	Describe Hadoop architecture with suitable diagram.	**07**

****