GUJARAT TECHNOLOGICAL UNIVERSITY
Bachelor of Engineering - SEMESTER - VI EXAMINATION - WINTER 2025

Subject Code: 3163209   **Date: 27-11-2025**

Subject Name: Data Mining and Business Intelligence

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)	Explain the difference between OLTP and OLAP systems with suitable examples.	03
(b)	Describe the architecture and key components of a Data Warehouse.	04
(c)	Discuss the role of Metadata in a Data Warehouse and explain how it supports Business Intelligence operations.	07
Q.2 (a)	Define KDD and explain its relationship with Data Mining.	03
(b)	List and explain various data preprocessing techniques with examples.	04
(c)	Design a data preprocessing flow for handling missing and noisy data.	07

OR

		Marks
(c)	Explain Mean, Median, Mode, and Variance& Standard Deviation in brief.	07
Q.3 (a)	What are frequent itemsets? Explain support and confidence with example.	03
(b)	Explain Apriori algorithm in brief.	04

(c) Consider the following set of transactions. Let min_sup = 40% and min_conf = 60%. **07**
1. Find all frequent itemsets using Apriori algorithm.
2. Generate strong association rules.

Transaction ID	Items Purchased
1	{Milk, Bread, Diapers}
2	{Bread, Coffee, Diapers}
3	{Milk, Diapers, Eggs}
4	{Milk, Bread, Coffee}
5	{Bread, Diapers, Eggs}

OR

(a) Define support and confidence. Why are these measures important in Association Rule Mining? **03**

(b) Describe steps in the FP-Growth algorithm. **04**

(c) i) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order): **07**
13, 15, 16, 16, 19, 20, 23, 29, 35, 41, 44, 53, 62, 69, 72
Use min-max normalization to transform the value 45 for age onto the
range [0:0, 1:0]
ii) Suppose that the mean and standard deviation of the values for attribute income are 54,000 Rs. and 16,000 Rs., respectively. then find z score value of 73,600 Rs salary.

Q.4 (a) Explain the difference between classification and prediction with examples. **03**

(b) Describe Bayesian classification with a suitable example. **04**

(c) Explain Decision Tree algorithm (ID3) to classify dataset. **07**

OR

(a) What is clustering? Differentiate between partitioning and hierarchical clustering. **03**

(b) Explain K-means clustering algorithm with suitable example. **04**

(c) Define outlier analysis. Why outlier mining is important? Briefly describe the different approaches for outlier detection. **07**

Q.5 (a) What is Big Data? Explain the characteristics and importance of Big Data Analytics.   **03**

(b) Explain how Hadoop handles distributed storage and processing using HDFS and MapReduce.   **04**

(c) Discuss real-world applications of Data Mining in Business Intelligence, such as fraud detection and customer segmentation.   **07**

OR

(a) Explain any four applications of Business Intelligence.   **03**

(b) Discuss Hadoop ecosystem components: Pig, Hive, and HBase.   **04**

(c) Explain how MapReduce can be used to perform Market Basket Analysis on a large dataset.   **07**

***