

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER-VII (NEW) EXAMINATION – SUMMER 2021

Subject Code:2170715

Date:06/08/2021

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.

		MARKS
Q.1	(a) Explain how data mining can play a crucial role in the success of a business.	03
	(b) Explain the Min-max data normalization method with suitable example.	04
	(c) What is KDD? With the help of a neat diagram explain the steps in KDD.	07
Q.2	(a) Explain the difference between a 'data warehouse' and a 'data mart'.	03
	(b) Explain the 'Star' and 'Snowflake' data warehouse schemas with suitable illustrations.	04
	(c) What is a Concept Hierarchy? With the help of a suitable example discuss concept hierarchy generation for a multi-dimensional data model.	07
	OR	
	(c) With the help of a neat diagram explain the three-tier architecture of a data warehouse.	07
Q.3	(a) Explain what is meant by "outlier" in the data mining process.	03
	(b) What are the reasons for missing values in real world data? Describe various methods for handling this problem.	04
	(c) Explain the steps of the Apriori Algorithm for mining Frequent Itemsets with Candidate Generation. Use a suitable example to illustrate your answer.	07
	OR	
Q.3	(a) What is "data visualization"? Explain its significance.	03
	(b) Give differences between OLAP and OLTP.	04
	(c) What are the limitations of the Apriori algorithm? Discuss ANY TWO methods to improve the efficiency of Apriori-based mining.	07
Q.4	(a) Explain what is meant by "activation function" in a Neural Network.	03
	(b) Explain the terms "support" and "confidence" with respect to association rule mining.	04

- (c) Define the term “Information Gain”. Explain the steps of the ID3 Algorithm for generating Decision Tree. **07**
- OR**
- Q.4** (a) Briefly explain ROLAP and MOLAP servers. **03**
 (b) Discuss the main characteristics of Bayesian Classification. **04**
 (c) Write a note on ‘Linear Regression’ and discuss its usefulness as a prediction method in data mining. **07**
- Q.5** (a) Explain the role of data scientists. **03**
 (b) What is Big Data? Explain the Four V’s of Big Data. **04**
 (c) Write a short note on Text Mining. **07**
- OR**
- Q.5.** (a) What is data clustering? How is it different from classification? **03**
 (b) Write a short note on Hadoop Distributed File System. **04**
 (c) Discuss the applications of data mining in Retail industry and Customer Relationship Management. **07**
