Seat No.: _____   Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER–VI (NEW) EXAMINATION – SUMMER 2023
Subject Code:3163209   **Date:18-07-2023**
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, Enterprise Warehouse & Virtual Warehouse	**03**
	(b)	A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data – Justify.	**04**
	(c)	What are different sources of information? Explain the term data, information and knowledge with suitable example.	**07**
Q.2	**(a)**	Differentiate Fact table vs. Dimension table.	**03**
	(b)	Define noise. Explain binning methods for data smoothing.	**04**
	(c)	Explain different OLAP operation with example.	**07**

<div align="center">OR</div>

	(c)	What is Data Mining? Explain Data mining as one step of Knowledge Discovery Process.	**07**
Q.3	**(a)**	List and describe the methods for handling the missing values in data cleaning.	**03**
	(b)	What is market basket analysis? Explain the two measures of rule interestingness: support and confidence.	**04**
	(c)	State the Apriori Property. Generate large itemsets and association rules using Apriori algorithm on the following data set with minimum support value and minimum confidence value set as 50% and 75% respectively.	**07**

TID	Items Purchased
T101	Cheese, Milk, Cookies
T102	Butter, Milk, Bread
T103	Cheese, Butter, Milk, Bread
T104	Butter, Bread

<div align="center">OR</div>

Q.3	**(a)**	Discuss click-stream analysis using data mining.	**03**
	(b)	Explain the Min-max data normalization method with suitable example.	**04**
	(c)	Explain three-tier Data Warehouse Architecture.	**07**
Q.4	**(a)**	Discuss following terms. 1) Supervised learning 2) Correlation analysis 3) Tree pruning	**03**
	(b)	Differentiate Association vs. Classification.	**04**
	(c)	Explain Baye's Theorm and calculate Naïve Bayesian Classification for given example:	**07**

Sr. no.	Outlook	Temperature	Humidity	Wind	Play
1	Sunny	hot	high	FALSE	No
2	Sunny	hot	high	TRUE	No
3	Overcast	hot	high	FALSE	Yes
4	Rain	mild	high	FALSE	Yes
5	Rain	cool	normal	FALSE	Yes
6	Rain	cool	normal	TRUE	No
7	Overcast	cool	normal	TRUE	Yes
8	Sunny	mild	high	FALSE	No
9	Sunny	cool	normal	FALSE	Yes
10	Rain	mild	normal	FALSE	Yes
11	Sunny	mild	normal	TRUE	Yes
12	Overcast	mild	high	TRUE	Yes
13	Overcast	hot	normal	FALSE	Yes
14	Rain	mild	high	TRUE	No

OR

Q.4 (a) Draw the topology of a multilayer, feed-forward Neural Network.   **03**

(b) Explain data mining application for fraud detection.   **04**

(c) Define linear and nonlinear regression using figures. Calculate the value of Y for X=100 based on Linear regression prediction method.   **07**

X	Y
4	390
9	580
10	650
14	730
4	410
7	530
12	600
22	790
1	350
3	400
8	590
11	640
5	450
6	520
10	690
11	690
16	770
13	700
13	730
10	640

Q.5 (a) Describe web mining using example.   **03**

(b) What is Big Data? What is big data analytic?   **04**

(c) Define the term "Information Gain". Explain the steps of the ID3 Algorithm for generating Decision Tree.   **07**

OR

Q.5 (a) Define: 1) Data Node 2) Name Node 3) Text mining   **03**

(b) Explain partitioning and hierarchical methods of clustering.   **04**

(c) What is distributed file system? Explain HDFS architecture in detail.   **07**

****