Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER– VII (New) EXAMINATION – WINTER 2019

Subject Code: 2170715 Date: 30/11/2019

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)**	Define data mining and list its features.	**03**
	(b)	Differentiate between OLTP and OLAP.	**04**
	(c)	Describe the steps involved in data mining when viewed as a process of knowledge discovery.	**07**
Q.2	**(a)**	Can BI is used for DM? Or vice versa? Justify.	**03**
	(b)	Explain in detail the extract/transform/load (ETL) design of an automated warehouse.	**04**
	(c)	Explain Mean, Median, Mode, Variance, Standard Deviation & five number summary with suitable database example.	**07**
		OR	
	(c)	Is Graphical visualization is better than text data ?Justify your answer and explain different data visualization technique.	**07**
Q.3	**(a)**	Explain why data warehouses are needed for developing business solutions from today's perspective. Discuss the role of data marts.	**03**
	(b)	Draw and Explain Snowflakes and Fact constellations Schema.	**04**
	(c)	Define outlier analysis? Why outlier mining is important? Briefly describe the different approaches: statistical-based outlier detection, distance-based outlier detection and deviation-based outlier detection.	**07**
		OR	
Q.3	**(a)**	Discuss Following: (i) Meta Data (ii) Virtual Warehouse	**03**
	(b)	Briefly outline the major steps of decision tree classification. Why tree pruning useful in decision tree induction?	**04**
	(c)	In data pre-processing why we need data smoothing? Discuss data smoothing by Binning.	**07**
Q.4	**(a)**	Draw the topology of a multilayer, feed-forward Neural Network.	**03**
	(b)	Describe Concept Hierarchy? List and briefly explain types of Concept Hierarchy	**04**
	(c)	In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem.	**07**
		OR	
Q.4	**(a)**	Define "clustering"? Mention any two applications of clustering.	**03**
	(b)	Briefly explain Linear and Non-linear regression.	**04**
	(c)	Consider the following dataset and find frequent item sets and generate association rules for them using Apriori Algorithm.	**07**

TID	items
T1	I1, I2 , I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

minimum support count is 2 minimum confidence is 60% .

Q.5 **(a)** Differentiate Fact table vs. Dimension table **03**

(b) What is market basket analysis? Explain the two measures of rule interestingness: support and confidence **04**

(c) Briefly explain the life-cycle of Data Analytics and discuss the role of data scientists. **07**

OR

Q.5 **(a)** Explain text mining using example. **03**

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

(c) Discuss the main features of Hadoop Distributed File System. **07**

****