Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–VII (NEW) EXAMINATION – WINTER 2017

Subject Code: 2170715 Date:18/11/2017

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 measures for finding rule interestingness (support, confidence) with example.	**03**
	(b)	Differentiate between OLTP and OLAP.	**04**
	(c)	What is Data Mining? Write down short note on KDD process.	**07**
Q.2	**(a)**	Use min-max normalization method to *normalize* the following group of data by setting min = 0 and max = 1 200, 300, 400, 600, 1000	**03**
	(b)	Describe various methods for handling missing data values.	**04**
	(c)	Using Apriori algorithm, generate frequent item sets (min_sup >= 33.3%) for the following transaction database. Trans_id Itemlist T1 {A, B, D, K} T2 {A, B, C, D, E} T3 {A, B, C, E} T4 {B, D} T5 {A, C} T6 {B, D}	**07**
		OR	
	(c)	Compare association and classification. Briefly explain associative classification with suitable example.	**07**
Q.3	**(a)**	Suppose a group of sales price records has been sorted as follows: 6, 9, 12, 13, 15, 25, 50, 70, 72, 92, 204, 232 Partition them into three bins by equal-frequency (equi-depth) partitioning method. Perform data smoothing by bin mean.	**03**
	(b)	Define Big Data. Discuss various applications of Big Data.	**04**
	(c)	What are the major issues in Data Mining?	**07**
		OR	
Q.3	**(a)**	Explain Prepruning and Postpruning with an example.	**03**
	(b)	Define the following terms: Business Intelligence, Data Mart, Closed frequent itemset, Outlier Analysis	**04**
	(c)	Why naïve Bayesian classification is called “naïve”? Describe naïve Bayesian classification with example.	**07**
Q.4	**(a)**	What is classification and prediction? List out Issues regarding Classification and prediction.	**03**
	(b)	Explain Star schema and Snowflake schema with example.	**04**
	(c)	Explain Hadoop storage – HDFS.	**07**

OR

Q.4	**(a)**	Explain the following terms: Numerosity reduction, Data Integration, Data transformation	**03**
	(b)	Explain data mining application for fraud detection.	**04**
	(c)	Explain sampling methods for data reduction.	**07**
Q.5	**(a)**	Explain various OLAP operations.	**07**
	(b)	Explain basic concepts of text mining and web mining.	**07**

OR

Q.5	**(a)**	What is an attribute selection measure? Explain different attribute selection measures with example.	**07**
	(b)	Explain Data warehouse architecture.	**07**

****