

GUJARAT TECHNOLOGICAL UNIVERSITY**BE- SEMESTER-VII (NEW) EXAMINATION – WINTER 2020****Subject Code:2170715****Date:28/01/2021****Subject Name:Data Mining and Business Intelligence****Time:10:30 AM TO 12:30 PM****Total Marks: 56****Instructions:**

1. Attempt any **FOUR** questions out of **EIGHT** questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

MARKS

- Q.1**
- (a) Differentiate OLAP and OLTP. **03**
- (b) Define Schema. Explain the following schemas with suitable example. **04**
1) Star 2) Snowflakes 3) Constellations
- (c) Define noise data. Enlist the reasons for the presence of noise in data collection. Explain the methods to deal with noise. **07**
- Q.2**
- (a) Define market basket analysis. Explain *support* and *confidence* with suitable example for finding the rules. **03**
- (b) Explain the following terms. **04**
1) Bias 2) Variance 3) Generalization 4) Outlier
- (c) Define the Apriori Property. Generate candidate itemsets, frequent itemsets and association rules using Apriori algorithm on the following data set with minimum support count is **2** and minimum confidence is **60%**. **07**

TID	Items
T1	BREAD, BUTTER, TOAST
T2	BUTTER, MILK
T3	BUTTER, BUSCUIT
T4	BREAD, BUTTER, MILK
T5	BREAD, BUSCUIT
T6	BUTTER, BUSCUIT
T7	BREAD, BUSCUIT
T8	BREAD, BUTTER, BUSCUIT, TOAST
T9	BREAD, BUTTER, BUSCUIT

- Q.3**
- (a) Minimum salary is 20,000/- Rs and Maximum salary is 1,70,000/- Rs. Map the salary 1,00,000/- Rs in new Range of (60,000 , 2,60,000) Rs using min-max normalization method. **03**
- (b) Define time series database. Explain how to characterize time series data using trend analysis. **04**
- (c) Define data cube and explain any 3 operations on it. **07**
- Q.4**
- (a) If Mean salary is 54,000Rs and standard deviation is 16,000 Rs then find z score value of 73,600 Rs salary. **03**
- (b) Briefly explain linear and non-linear regression. **04**
- (c) Define Map and Reduce operations with suitable example. **07**
- Q.5**
- (a) Discuss the following terms: **03**
1) Tree Pruning 2) Information Gain 3) Spatiotemporal Databases
- (b) Discuss the major issues/challenges in data mining. **04**
- (c) Enlist the steps of K-Mean clustering algorithm. Explain it with suitable example. **07**

Q.6	(a) Discuss the following terms: 1) Correlation analysis 2) Gain Ration 3) Sequence Databases	03
	(b) Differentiate classification and prediction.	04
	(c) Enlist the steps of ID3 decision tree generation algorithm. Explain it with suitable example and generate the tree.	07
Q.7	(a) Explain Web structure and Web usage mining.	03
	(b) Draw and explain the topology of a multilayer, feed-forward Neural Network.	04
	(c) Define big data and big data analytics. Explain big data distributed file system.	07
Q.8	(a) Draw and explain Hadoop architecture.	03
	(b) Discuss Hash-based technique to improve efficiency of Apriori algorithm.	04
	(c) Explain Baye's Theorm and Naïve Bayesian Classification with suitable example.	07
