

GUJARAT TECHNOLOGICAL UNIVERSITY**BE – SEMESTER- VII EXAMINATION-SUMMER 2023****Subject Code: 2170715****Date: 26/06/2023****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 Mining. Explain KDD process in detail.	03
	(b) Explain the following terms with suitable example. 1) Data Integration 2) Data Transformation	04
	(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) Compare association and classification.	03
	(b) Define Schema. Explain the following schemas with suitable example. 1) Star 2) Snowflakes 3) Constellations	04
	(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	PENCIL, SHARPNR, ERASER
T2	SHARPNR, INK
T3	SHARPNR, NOTEBOOK
T4	PENCIL, SHARPNR, INK
T5	PENCIL, NOTEBOOK
T6	SHARPNR, NOTEBOOK
T7	PENCIL, NOTEBOOK
T8	PENCIL, SHARPNR, NOTEBOOK, ERASER
T9	PENCIL, SHARPNR, NOTEBOOK

OR

- (c) Using a Euclidian distance generate 2 clusters using k-means clustering algorithm for the following dataset. **07**

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

Assume mean as the following.

	Individual	Mean Vector (Centroid)
Group1	1	(1.0, 1.0)
Group2	4	(5.0, 7.0)

- Q.3** (a) Use min-max normalization method, map the value 100000 by setting new range of (60000, 260000) with previous min= 20000 and max=170000. **03**
- (b) Explain spatial mining using example. **04**
- (c) Define time series database. Explain how to characterize time series data using trend analysis. **07**

OR

- Q.3** (a) If Mean salary is 54,000 Rs and standard deviation is 16,000 Rs then find z score value of 73,600 Rs salary. **03**
- (b) Briefly explain logistic and multivariate regression. **04**
- (c) Explain Map-Reduce operation for word count problem. **07**
- Q.4** (a) Define Big Data. Explain 4Vs properties of Big Data. **03**
- (b) Discuss the major issues/challenges in data mining. **04**
- (c) Explain Hadoop Distributed File System (HDFS) storage. **07**

OR

- Q.4** (a) Discuss the following terms: **03**
1) Business Intelligence 2) Gain Ratio 3) Data Warehouse
- (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.5** (a) Explain Web structure and Web usage mining. **03**
- (b) Draw and explain the topology of a multilayer, feed-forward Neural Network. **04**
- (c) State and explain attribute selection measure with suitable example. **07**

OR

- Q.5** (a) Explain how continuous numeric attribute can be discretized. **03**
- (b) Discuss Hash-based technique to improve efficiency of Apriori algorithm. **04**
- (c) Explain Baye's Theorem and Naïve Bayesian Classification with suitable example. **07**
