Enrolment No./Seat No_____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE- SEMESTER–VI (NEW) EXAMINATION – WINTER 2024

Subject Code:3160714 Date:02-12-2024

Subject Name:Data Mining

Time:02:30 PM TO 05:00 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.**

Q.1 **(a)** Define each of the following data mining functionalities: characterization, discrimination, regression. **03**

(b) How is a data warehouse different from database? **04**

(c) Explain KDD process. **07**

Q.2 **(a)** How to handle missing values? Explain. **03**

(b) How to handle noisy data? **04**

(c) Consider a database, **D**, consisting of 9 transactions. **07**
Suppose min. support count required is **2.**
Let minimum confidence required is **70%**.
Find out the frequent itemset using Apriori algorithm.

TID	List of Items
T100	I1, I2, I5
T100	I2, I4
T100	I2, I3
T100	I1, I2, I4
T100	I1, I3
T100	I2, I3
T100	I1, I3
T100	I1, I2 ,I3, I5
T100	I1, I2, I3

OR

(c) A database has five transactions. Let min sup=60% and min conf =80%. **07**

Tid	Item brought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Find all frequent item sets using Apriori and FP-growth, respectively. Compare the efficiency of the two mining processes.

Q.3 **(a)** Explain market basket analysis. **03**

(b) Explain Linear regression. **04**

(c) Explain decision tree algorithm. **07**

OR

Q.3	**(a)**	Explain WEKA tool.	**03**
	(b)	Explain logistic regression.	**04**
	(c)	Explain CART Classification Method**.**	**07**
Q.4	**(a)**	Compare classification and Clustering.	**03**
	(b)	Which metrics used for evaluating classifier performance?	**04**
	(c)	Explain Principal Component Analysis.	**07**

OR

Q.4	**(a)**	Compare classification and prediction.	**03**
	(b)	Explain outlier detection.	**04**
	(c)	Explain Backpropagation algorithm.	**07**
Q.5	**(a)**	Write applications of clustering graph and network data.	**03**
	(b)	What is Web log structure? And discuss issues regarding web logs.	**04**
	(c)	Explain PAM clustering Algorithm.	**07**

OR

Q.5	**(a)**	Write similarity measures for clustering graph and network data.	**03**
	(b)	Explain Web Structure mining.	**04**
	(c)	Write Applications of Distributed and parallel Data Mining.	**07**

****