GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER–VI (NEW) EXAMINATION – WINTER 2023

Subject Code:3161610   **Date:13-12-2023**
Subject Name: Data Warehousing and 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)**	Explain any three OLAP operations with suitable example.	**03**
	(b)	Discuss Data Smoothing by Binning using suitable example.	**04**
	(c)	What is Data Mining? Why is it called Data Mining rather Knowledge Mining? Explain KDD process.	**07**
Q.2	**(a)**	Explain the major issues in Data Mining.	**03**
	(b)	Compare OLAP and OLTP systems.	**04**
	(c)	What is an 'Outlier'? How do Outliers impact the results of Mining? Explain any one method to detect Outliers.	**07**
		OR	
	(c)	Explain Linear and Non-linear Regression.	**07**
Q.3	**(a)**	Define Euclidean Distance, Manhattan Distance & Minkowski Distance.	**03**
	(b)	List and explain types of Concept Hierarchy.	**04**
	(c)	List out the different schema of Data Warehouse and explain one of the Data Warehouse schemas in detail with suitable diagram.	**07**
		OR	
Q.3	**(a)**	Define Supervised Learning, Unsupervised Learning & Data Reduction.	**03**
	(b)	Explain Five Number Summary with suitable database example.	**04**
	(c)	What do you mean by Data Mart? What are the different types of Data Mart?	**07**
Q.4	**(a)**	Explain about role of Principal Component of Analysis (PCA).	**03**
	(b)	What is Data Transformation? Explain the different Data Transformation approaches for Transforming Data.	**04**
	(c)	Define Sampling. Explain different type of Sampling Techniques with suitable example.	**07**
		OR	
Q.4	**(a)**	Define Bayes' Theorem.	**03**
	(b)	What is Data Cleaning? Describe the different methods of handling missing values during Data Cleaning.	**04**
	(c)	Define Noise. Explain the different techniques to remove the Noise from Data.	**07**
Q.5	**(a)**	Explain Information Gain and Gain Ratio.	**03**
	(b)	Write a short note on Web Mining.	**04**
	(c)	What is Decision Tree Induction? Write Basic Algorithm for inducing a Decision Tree from training tuples.	**07**
		OR	
Q.5	**(a)**	Explain Market Basket Analysis.	**03**
	(b)	Write a short note on Text Mining.	**04**
	(c)	Explain the steps of the Apriori algorithm for mining frequent item sets with candidate generation.	**07**

****