GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER–VI(NEW) EXAMINATION – WINTER 2022

Subject Code:3161610   **Date:17-12-2022**

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.**

			MARKS
Q.1	**(a)**	Explain various features of Data Warehouse.	**03**
	(b)	Draw the three tier architecture of data warehouses.	**04**
	(c)	Explain KDD Process.	**07**
Q.2	**(a)**	Give the difference between OLAP and OLTP.	**03**
	(b)	Define the term "data mining". List the major issues in data mining.	**04**
	(c)	With the help of a suitable example, illustrate the OLAP operations: 'drill-down', 'roll-up', 'slice' and 'dice'	**07**
		OR	
	(c)	Explain data smoothing methods to divide given data into bins of size 3 by bin means, by bin medians and by bin boundaries. Consider the data: 10, 2, 19, 18, 20, 18, 25, 28, 22	**07**
Q.3	**(a)**	Differentiate Fact table vs. Dimension table	**03**
	(b)	What is noise? Explain the various noise removing methods.	**04**
	(c)	Explain Mean, Median, Mode, Variance, Standard Deviation with suitable database example	**07**
		OR	
Q.3	**(a)**	What is data cleaning? How to handle the missing value in data cleaning?	**03**
	(b)	Why pre-processing is required in data mining process? List the major steps of pre-processing.	**04**
	(c)	Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order):13, 15, 16, 16, 19, 20, 23, 29, 35, 41, 44, 53, 62, 69, 72 i) Use min-max normalization to transform the value 50 for age onto the range [0:0, 1:0] ii) Use z-score normalization to transform the value 50 for age, where the standard deviation of age is 20.64 years.	**07**
Q.4	**(a)**	What are the limitations of the Apriori approach for mining?	**03**
	(b)	What is Market Basket Analysis? Explain Association Rules with Confidence & Support.	**04**

(c) State the Apriori Property. Find frequent item-sets and association rules using Apriori algorithm on the following data set with minimum support count is 2 and minimum confidence=75%. **07**

Sr.No	TID	List of items_IDs
1	T100	I1,I2,I5
2	T200	I2,I4
3	T300	I2,I3
4	T400	I1,I2,I4
5	T500	I1,I3
6	T600	I2,I3
7	T700	I1,I3
8	T800	I1,I2,I3,I5
9	T900	I1,I2,I3

OR

Q.4 (a) What is classification? Give the difference between classification and predication **03**

(b) What is Information gain and Gain ratio? **04**

(c) Explain Baye's Theorem and Naïve Bayesian Classification. **07**

Q.5 (a) What is clustering? Why clustering is an un-supervised learning? **03**

(b) Write a short note on text mining. **04**

(c) Explain k-means algorithm of clustering. **07**

OR

Q.5 (a) Give the Difference between Spatial and Temporal Data Mining. **03**

(b) Briefly explain Linear and Non-linear regression. **04**

(c) Explain different types of Web Mining with example. **07**

****