

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER-VI (NEW) EXAMINATION – SUMMER 2022

Subject Code:3161610

Date:10/06/2022

Subject Name:Data Warehousing and Mining

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) What is Data Warehousing? Explain its features.	03
	(b) With the help of a neat diagram explain the 3-tier architecture of Data warehouse.	04
	(c) What is Cube? Discuss various OLAP Operations on Data Cube.	07
Q.2	(a) Describe major issues in data mining.	03
	(b) Do feature wise comparison between OLAP and OLTP.	04
	(c) Define the term “Data Mining”. Why is it called data mining rather knowledge mining? Explain the process of knowledge discovery from databases with the help of a suitable diagram	07
OR		
	(c) Explain Star, Snowflake, and Fact Constellation Schema for Multidimensional Database with diagram.	07
Q.3	(a) List different methods for data discretization and explain any one in detail.	03
	(b) What is feature selection and explain methods for feature selection.	04
	(c) Explain the pre-processing required to handle missing data and noisy data during the process of data mining.	07
OR		
Q.3	(a) Explain the following as attribute selection measure: (i) Information Gain (ii) Gain Ratio	03
	(b) Discuss about dimensionality reduction in brief.	04
	(c) What is noise? Why data smoothing is required. Perform smoothing by bin means, by bin medians and by bin boundaries on the following data. Consider the data for price (in dollars): 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34	07
Q.4	(a) Why tree pruning useful in decision tree induction?	03
	(b) Explain Association Rules with Confidence & Support giving an example.	04
	(c) Explain how the accuracy of classifier/predictor can be measured.	07
OR		
Q.4	(a) Explain Mining Multiple-Level Association Rules using example.	03
	(b) How k-mean clustering method differs from k-medoid clustering method?	04

- (c) Find the frequent itemsets and generate association rules on below given data using apriori algorithm. Assume that minimum support threshold ($s = 33.33\%$) and minimum confident threshold ($c = 60\%$) **07**

Transaction ID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	Hot Dogs, Coke, Chips

- Q.5** (a) Difference between linear and logistic regression. **03**
(b) What is outlier? Discuss different methods for outlier detection. **04**
(c) What is web mining? Explain types of web mining. **07**

OR

- Q.5** (a) Difference between spatial mining and temporal mining. **03**
(b) Discuss k-NN classification algorithm. **04**
(c) Explain Linear Regression using suitable example along with its pros and cons. **07**
