

GUJARAT TECHNOLOGICAL UNIVERSITY**BE - SEMESTER-VI (NEW) EXAMINATION – WINTER 2023****Subject Code:3160714****Date:11-12-2023****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.

		Marks
Q.1	(a) Define Data Warehouse. State it's features.	03
	(b) Differentiate between OLAP and OLTP.	04
	(c) Explain in detail different steps of KDD process.	07
Q.2	(a) Why to preprocess the data in data Mining?	03
	(b) Explain Binning method with the help of example.	04
	(c) Explain following terms related to Association Rule Mining: Itemset, Support Count, support, and Association rule.	07

Transaction ID	Items
1	Bread, Milk
2	Bread, Chocolate, Pepsi, Eggs
3	Milk, Chocolate, Pepsi, Coke
4	Bread, Milk, Chocolate, Pepsi
5	Bread, Milk, Chocolate, Coke

For given example find support & confidence for

{Milk, Chocolate} \Rightarrow Pepsi.

{Milk, Pepsi} \rightarrow {Chocolate}

{Chocolate, Pepsi} \rightarrow {Milk}

OR

- (c) Solve the following problem using Apriori algorithm. **07**
 Find the frequent itemsets and generate association rules on this.
 Assume that minimum support threshold ($s = 33.33\%$), minimum confident threshold ($c = 60\%$), minimum support count=2.

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.3** (a) Define the following terms in Data Transformation: **03**
- i. Smoothing
 - ii. Normalization

iii. Discretization

- (b) Differentiate between Classification and Prediction. **04**
(c) Explain Decision Tree Classification algorithm with the help of example. **07**

OR

- Q.3** (a) Differentiate between supervised learning and unsupervised learning. **03**
(b) What is Regression? Explain Linear Regression in short. **04**
(c) Explain Naïve Bayes Classifier with example. **07**
- Q.4** (a) What do you mean by Tree Pruning? Explain with example. **03**
(b) Explain the following as attribute selection measure: **04**
(i) Information Gain
(ii) Gain Ratio
(c) What do you mean by learning-by-observation? Explain k-Means clustering algorithm in detail. **07**

OR

- Q.4** (a) Define Data Cube. Explain any two operations on it. **03**
(b) Differentiate between Partition method and Hierarchical method of Clustering. **04**
(c) What are the requirements of Clustering in Data Mining? **07**
- Q.5** (a) How K-Mean clustering method differs from K-Medoid clustering method? **03**
(b) Draw and explain the topology of a multilayer, feed-forward Neural Network. **04**
(c) Explain the major issues in data mining. **07**

OR

- Q.5** (a) Give difference between text mining and web mining. **03**
(b) Why Hadoop is important? **04**
(c) What is web log? Explain web structure mining and web usage mining in detail. **07**
