

**GUJARAT TECHNOLOGICAL UNIVERSITY****BE - SEMESTER-VI (NEW) EXAMINATION – SUMMER 2024****Subject Code: 3160714****Date:22-05-2024****Subject Name: Data 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.

	<b>Marks</b>
<b>Q.1</b> (a) Define data mining. Describe three challenges to data mining regarding data mining methodology and user interaction issues.	<b>03</b>
(b) Explain the steps in knowledge discovery.	<b>04</b>
(c) Explain the various data mining issues.	<b>07</b>
<b>Q.2</b> (a) What are the smoothing techniques available to remove noise?	<b>03</b>
(b) Discuss normalization in detail.	<b>04</b>
(c) In real-world data, tuples with <i>missing values</i> for some attributes are a common occurrence. Describe various methods for handling this problem.	<b>07</b>
<b>OR</b>	
(c) Discuss data discretization and concept hierarchy generation.	<b>07</b>
<b>Q.3</b> (a) How are association rules mined from large databases?	<b>03</b>
(b) Give the difference between Boolean association rule and quantitative association rule.	<b>04</b>
(c) What are the limitations of the apriori approach for mining? Briefly describe the techniques to improve the efficiency of apriori algorithm.	<b>07</b>
<b>OR</b>	
<b>Q.3</b> (a) Describe two interesting measures for association rules.	<b>03</b>
(b) How Meta rules are useful in constraint based association mining.	<b>04</b>
(c) Write an algorithm for finding frequent item-sets using candidate generation.	<b>07</b>
<b>Q.4</b> (a) What are the difference between supervised learning and unsupervised learning?	<b>03</b>
(b) Write down short note on backpropagation.	<b>04</b>
(c) What is information gain? Explain the steps required to generate a decision tree from a training data set.	<b>07</b>
<b>OR</b>	
<b>Q.4</b> (a) Differentiate between linear regression and nonlinear regression.	<b>03</b>
(b) Explain various methods of evaluating accuracy of classifier.	<b>04</b>
(c) Write a short on: web content mining.	<b>07</b>
<b>Q.5</b> (a) Explain temporal mining.	<b>03</b>
(b) Differentiate between partitioning and hierarchical methods for clustering.	<b>04</b>
(c) Explain following clustering algorithm in details:	<b>07</b>
1) CLARA	
2) BIRCH	
<b>OR</b>	
<b>Q.5</b> (a) List out the applications of distributed and parallel data mining.	<b>03</b>
(b) Illustrate strength and weakness of k-mean in comparison with k-medoid algorithm.	<b>04</b>
(c) Explain the typical requirements of clustering in data mining.	<b>07</b>

\*\*\*\*\*