## **GUJARAT TECHNOLOGICAL UNIVERSITY**

Subject	code	M. E SEMESTER – II • EXAMINATION – WINTER 2012 e: 1720108 Date: 02/01/2013	
Subject	Nan	ne: Data Mining and Data Warehousing am – 01.00 pm Total Marks: 70	
Instruc		-	
		empt all questions.	
2. 3.	Make suitable assumptions wherever necessary.		
Q.1	(a)	<ol> <li>Write down short note on KDD process.</li> <li>Write down short note on Attribute Oriented Induction.</li> </ol>	05 02
	<b>(b)</b>	Explain with an example Snowflake and Galaxy Schema.	07
Q.2	(a)	Discuss why analytical data characterization is needed and how it can be performed.	07
	<b>(b)</b>	What is Classification? Compare the merits and demerits of eager classification versus lazy classification. <b>OR</b>	07
	<b>(b)</b>	<ol> <li>Explain bitmap indexing with an example.</li> <li>Define Histogram and Hold-out method with an example.</li> </ol>	03 04
Q.3	(a)	<ol> <li>Discuss major issues in data mining.</li> <li>In real-world data, tuples with missing values for some attributes are a Common occurrence. Describe various methods for handling this problem.</li> </ol>	03 04
	<b>(b)</b>	What is Data transformation? Briefly explain various steps of data transformation.	07
		OR	
Q.3	(a)	<ol> <li>Define following terms.</li> <li>Roll-up 2). Drill-down 3). Pivot</li> <li>Explain Prepruning and Postpruning with an example.</li> </ol>	03
	<b>(b)</b>	List and describe the five primitives for specifying a data mining task.	07
Q.4	(a)	<ol> <li>What is clustering? List and explain any two clustering methods.</li> <li>What is Binning? List and explain binning strategies.</li> </ol>	04 03
	<b>(b)</b>	Write down short note on Backpropagation.  OR	07
Q.4	(a)	<ol> <li>Explain different approaches to mining multilevel association rules.</li> <li>Define Data Warehouse, Data Mart and Virtual Data Warehouse.</li> </ol>	04 03
Q.4	<b>(b)</b>	Briefly explain Apriori Algorithms with an example.	07
Q.5	(a)	Briefly outline how to compute the dissimilarity between objects described by the following types of variables:  Binary variables, Nominal Variables, Ratio-scaled variables, and Ordinal variables.	07
	<b>(b)</b>	What are the challenges for effective resource and knowledge discovery in mining the world wide web?	07
Q.5	(a)	OR  Why naïve Bayesian classification is called "naïve"? Briefly outline the major idea of naïve Bayesian classification.	07
	<b>(b)</b>	What do you mean by text mining? Explain various issues involved in it.	07

\*\*\*\*\*\*