

GUJARAT TECHNOLOGICAL UNIVERSITY**M. E. - SEMESTER – II • EXAMINATION – WINTER 2012****Subject code: 1720108****Date: 02/01/2013****Subject Name: Data Mining and Data Warehousing****Time: 10.30 am – 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.

- Q.1** (a) 1. Write down short note on KDD process. **05**
 2. Write down short note on Attribute Oriented Induction. **02**
 (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) What is Classification? Compare the merits and demerits of eager classification versus lazy classification. **07**
OR
 (b) 1. Explain bitmap indexing with an example. **03**
 2. Define Histogram and Hold-out method with an example. **04**
- Q.3** (a) 1. Discuss major issues in data mining. **03**
 2. In real-world data, tuples with missing values for some attributes are a Common occurrence. Describe various methods for handling this problem. **04**
 (b) What is Data transformation? Briefly explain various steps of data transformation. **07**
OR
- Q.3** (a) 1. Define following terms. **03**
 1). Roll-up 2). Drill-down 3). Pivot
 2. Explain Prepruning and Postpruning with an example. **04**
 (b) List and describe the five primitives for specifying a data mining task. **07**
- Q.4** (a) 1. What is clustering? List and explain any two clustering methods. **04**
 2. What is Binning? List and explain binning strategies. **03**
 (b) Write down short note on Backpropagation. **07**
OR
- Q.4** (a) 1. Explain different approaches to mining multilevel association rules. **04**
 2. Define Data Warehouse, Data Mart and Virtual Data Warehouse. **03**
Q.4 (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: **07**
 Binary variables, Nominal Variables, Ratio-scaled variables, and Ordinal variables.
 (b) What are the challenges for effective resource and knowledge discovery in mining the world wide web? **07**
OR
- Q.5** (a) Why naïve Bayesian classification is called “naïve”? Briefly outline the major idea of naïve Bayesian classification. **07**
 (b) What do you mean by text mining? Explain various issues involved in it. **07**
