

**GUJARAT TECHNOLOGICAL UNIVERSITY**  
**B. E. - SEMESTER – VII • EXAMINATION – WINTER 2012**

**Subject code: 171601****Date: 26/12/2012****Subject Name: Data Warehousing and Data****Time: 10.30 am - 01.00 pm****Total Marks: 70****Instructions:**

1. Attempt any five questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

- Q.1** (a) Define the term “Data Mining”. With the help of a suitable diagram explain the process of knowledge discovery from databases. **07**
- (b) Explain the pre-processing required to handle missing data and noisy data during the process of data mining. **07**
- Q.2** (a) With the help of a neat diagram explain the 3-tier architecture of a data warehouse. **07**
- (b) Clearly state the differences between “Data Warehouses” and “Operational Database Systems”. **07**
- OR**
- (b) Explain the ‘Star’ and ‘Snowflake’ schemas of data warehouse. **07**
- Q.3** (a) Explain the steps of the “Apriori Algorithm” for mining frequent itemsets. **07**
- (b) Explain with example how continuous numerical data values can be discretized. **07**
- OR**
- Q.3** (a) What is “Information Gain”? Explain the steps required to generate a Decision Tree from a training data set. **07**
- (b) Explain methods for data normalization. **07**
- Q.4** (a) Write the steps of the k-means clustering algorithm. Also state its limitations. **07**
- (b) Explain how the accuracy of a classifier can be measured. **07**
- OR**
- Q.4** (a) Write a short note on hierarchical clustering. **07**
- Q.4** (b) Explain “Linear Regression” using suitable example. **07**
- Q.5** (a) Explain the information retrieval methods used in text mining. **07**
- (b) Explain how the topology of a neural network is designed. **07**
- OR**
- Q.5** (a) Explain how a search engine automatically identifies authoritative web pages on a user’s search topic. **07**
- (b) Discuss applications of “Fuzzy Logic”. **07**

\*\*\*\*\*^\*\*\*\*\*