

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY
BE SEM-VII Examination-Nov/Dec.-2011

Subject code: 171601

Date: 19/11/2011

Subject Name: Data warehousing and Data Mining

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) List and describe major issues in data mining. **07**
(b) Explain the methodologies for stream data processing and stream data Systems. **07**
- Q.2** (a) Short note: Information gain, Gain ratio, Gini index. **07**
(b) Write the typical requirements of clustering in data mining. **07**
- OR**
- (b) Explain k-means and k-medoids algorithms of clustering. **07**
- Q.3** (a) What is noise? Describe the possible reasons for noisy data. Explain the different techniques to remove the noise from data. **07**
(b) Explain the KDD process in details. **07**
- OR**
- Q.3** (a) List and describe the methods for handling the missing values in data cleaning. **07**
(b) Explain meta data repository. **07**
- Q.4** (a) Differentiate between OLTP and OLAP systems. **07**
(b) Explain rule based classification and case based reasoning in details. **07**
- OR**
- Q.4** (a) Write an algorithm for finding frequent item-sets using candidate generation. **07**
(b) Short Note: Distributive and Holistic measures. **07**
- Q.5** (a) Describe the list of techniques for improving the efficiency of Apriori-based mining. **07**
(b) Explain three-tier data warehouse architecture. **07**
- OR**
- Q.5** (a) What are the challenges for effective resource and knowledge discovery in mining the world wide web? **07**
(b) Explain data transformation in data mining. **07**
