## GUJARAT TECHNOLOGICAL UNIVERSITY BE - SEMESTER-VII(OLD) • EXAMINATION – WINTER 2016

Date: 29/11/2016

Subject Code: 171601

**Subject Name: Data warehousing and 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. 07 **Q.1** (a) What are different sources of information? Explain the term Data, Information and Knowledge with suitable example. Explain OLAP Model and its advantages. Explain the various OLAP Scheme 07 and operations. **Q.2** What is data cleaning? Describe the different methods of handling missing 07 values during data cleaning. Explain Data Marts. What are the different types of data marts? Give 07 advantages of data marts over data warehouses. What do you understand by the terms data characterization in content to concept **(b)** 07 description? 0.3 Explain Star, Snowflake, Fact Constellation Schema for Multidimensional **07** (a) Database. **(b)** Explain Baye's Theorm and Naïve Bayesian Classification. **07** 0.3 What do you mean by frequent item-set mining for market basket analysis? 07 Explain apriori algorithm for the same with suitable example. **(b)** What do you mean by Decision Tree? Describe ID3 algorithm of decision tree. 07 Why is it unsuitable for data mining application? 0.4 What do you meant by multilevel association rule? Give different methods to **07** find frequent item sets at different levels by suitable example. What is associative classification? Why is associative classification able to 07 achieve higher classification accuracy than a classical decision tree method? Explain how associative classification can be used for text document classification. OR 0.4 Present conditions under which density-based clustering is more suitable than **07** partitioning-based clustering and hierarchical clustering. Give some sample data sets to support your arguments. (b) Explain Linear Regression and Non-linear Regression techniques of prediction. 07 Q.5 (a) What is web log? What are the access limitations in web mining? Explain it. 07 Explain the information retrieval methods used in text mining. **(b)** 07 OR What is Clustering and its attributes? Explain K-Mean clustering algorithm. Q.5 **07** (a) What is time series database? How to characterize the time series data using **07 (b)** trend analysis?

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