

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY
MCA - SEMESTER– IV • EXAMINATION – WINTER - 2016

Subject Code:640005

Date:26/10/ 2016

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.**

- Q.1** **(a)** i) Give any four differences between OLAP and OLTP. **02**
 ii) Explain meaning of 4 keywords used in the definition of Data Warehouse. **02**
 iii) Define Concept hierarchy with example. **03**

- (b)** i) Explain in brief Knowledge Discovery from Database with diagram. **05**
 ii) Explain Star schema and Snow flake schema with suitable example. **02**

- Q.2** **(a)** Explain outlier analysis with its importance. **07**
 (b) Why real world data is dirty? How to clean the Noisy data? Elaborate it with appropriate example **07**

OR

- (b)** What do you mean by data reduction? Explain in brief any two data reduction strategies. **07**

- Q.3** **(a)** What is SelfExp, PathExp and InExp? How these measures are used in discovery driven exploration of data cubes? **05**
 (b) Discuss Decision Tree Induction algorithm. **05**
 (c) Define Terms: Data Mart, Loose coupling, Tight coupling, Prediction **04**

OR

- Q.3** **(a)** Explain Bagging and Boot strap methods. **05**
 (b) Discuss Bayes' Theorem. **05**
 (c) Define Terms: Data Cube, Semi-tight coupling, no-coupling, Classification **04**

- Q.4** **(a)** What is Apriori property? How the Apriori property is used in finding frequent itemset. Explain Join and Prune step **07**
 (b) Justify the Statement with proper example: "Strong rules are not necessarily interesting". **07**

OR

- Q.4** **(a)** Discuss Major issues in Data Mining. **07**
 (b) Justify the statement: "All patterns derived by Data Mining algorithm are Interesting". **07**

- Q.5** **(a)** Write down the k-medoid algorithm for clustering and explain its working. How it is better than k-mean clustering algorithm? **07**
 (b) Write Short Note: **07**
 i) Data Mining in Retail Industry
 ii) Types of OLAP Servers

OR

- Q.5** **(a)** Explain K-mean clustering method with its advantages and disadvantages. **07**
 (b) Write Short Note: **07**
 iii) Role of Data Mining in Financial Data Analysis
 iv) OLAP Operations
