

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY
MCA - SEMESTER-IV • EXAMINATION – SUMMER 2015

Subject Code: 2640005

Date: 20/05/2015

Subject Name: Data Warehousing and Data Mining (DWDM)

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)** Define Following Full Forms: **07**
1. OLAM
 2. HOLAP
 3. ETL
 4. KDD
 5. AOI
 6. FP
 7. CART

- (b)** Discuss and Differentiate Query-Driven Approach and Update-Driven Approach **07**

- Q.2 (a)** Explain back- end tools and utilities for Data warehouse System also Justify why it is Important for Data warehouse Management **07**

- (b)** Explain Starnet Query Model for Multidimensional Database also Draw Starnet Model for following Dimensions **07**

[location, Customer, item, time]

OR

- (b)** What is Data Mining? Discuss Major Components for Data Mining **07**

- Q.3 (a)** What is Outliers? Explain Useful Tool for Identifying Outliers **07**

- (b)** What do you Understand Data Noisy? Explain Smoothing technique for Handle Data Noisy **07**

OR

- Q.3 (a)** Generate Mining Class Comparison for following Classes **07**

Target Class Graduate Students:

Name	Gender	Major	Birth Date	Birth Place	Residence	gpa
Jenifer Adleman	F	BCA	03/01/85	New York	USA Satellite	6.5
Rutherford	M	BBA	06/02/87	Vancouver	Canada, new road	6.3

Contrasting Class Undergraduate Students

Name	Gender	Major	Birth Date	Birth Place	Residence	gpa
Bob	M	12 Bio	03/01/90	Canada	Alt Canada	2.9
Amy	F	12 Che	06/02/91	Canada	BC Canada	2.3

Prepare at least Two Class Comparison for above table

Bio = Biology Che= Chemistry

- (b)** Explain different Strategies for Data Reduction **07**

Q.4 (a) Prepare Apriori Algorithm for finding Frequent item set in Database **07**

Database TDB $\text{Sup}_{\text{Min}} = 2$

Tid	Items
10	A,C,D
20	B,C,E
30	A,B,C,E
40	B,E

(b) Discuss Naïve Bayesian Classification for Classification and Prediction **07**

OR

Q.4 (a) Explain with example how Improving the Efficiency of Apriori **07**

(b) Discuss following in brief: **07**

1. Information Gain
2. Gain Ratio
3. Gini Index

Q.5 (a) Explain the Typical Requirements of Clustering in Data Mining **07**

(b) Discuss Financial Data Analysis For Data Mining Applications **07**

OR

Q.5 (a) Write steps for K-means partitioning Cluster Algorithm **07**

(b) How choose Data Mining? Explain with example of Commercial Data Mining Systems **07**
