Enrolment No.\_\_

## GUJARAT TECHNOLOGICAL UNIVERSITY

SEMESTER- 4 EXAMINATION – WINTER 2012								
Subj	ect c	Date: 11/01/2013						
Subject Name: Elective I – Data Warehousing & Data Mining (DWDM)								
Time	e:10:	Total Marks: 70						
Instructions.								
1 Attempt all questions								
	2	Make suitable assumptions wherever necessary						
	3	Figures to the right indicate full marks						
		i gui es to the right multure fun murks.						
0.1	(a)	What is Data-warehouse? Explain it with its	07					
	()	characteristics.						
	<b>(b)</b>	Define KDD process with its steps to generate knowledge	07					
Q.2	(a)	Differentiate OLTP and OLAP.	07					
	<b>(b)</b>	Write a short note on Data-Mart.	07					
		OR						
	<b>(b)</b>	Explain Star schema with suitable example?	07					
Q.3	<b>(a)</b>	Explain types of OLAP operations with suitable example.	07					
	<b>(b)</b>	Explain Data Mining architecture in detail.	07					
		OR						
Q.3	(a)	Explain the terms Data Cleaning, Data Integration & 07						
		Transformation with suitable example						
	(b)	Write a short note on usage of "Data Mining in Retail and	07					
		Marketing".						
0.4	$(\cdot)$	<b>XX7 '/ 1 / / A ' ' 1 '/1 1 1 A ' '</b>	07					
Q.4	(a)	on following data set and find out frequent	07					
		combination/Itom sat for 3 products where minimum						
		support -2						
		Transaction Number Items Purchased						
		10 A C D						
		$\begin{array}{c c} 10 & R, C, D \\ \hline 20 & B, C, F \\ \end{array}$						
		$\begin{array}{c c} 20 \\ \hline 30 \\ \hline \end{array} \\ \hline A B C E \\ \hline \end{array}$						
		40 B E						
			~-					

(b) Write a short note types of clusters.

07

OR

Q.4 (a) Explain Association and Corelation with Lift(A,B). Define 07 the Lift(A,B) measurement between 2 products TV and DVD where there are total 5000 transactions carried out. Out of which TV is sold in 3000 transactions. DVD is sold in 2500 transactions. TV and DVD together sold in 2000 transactions. Find out the Lift(TV,DVD) correlation.

Q.4 (b) Write a short note on "Types of data in cluster analysis". 07

- Q.5 (a) What is classification? Also explain Supervised and 07 Unsupervised learning in detail.
  - (b) What is K-means clustering? Create 2 clusters for the given 07 data-set using K-Means method and Euclidean Distance. Assign City C1 and C2 as initial Centroid point for the 2 clusters.

City	Χ	Y	
C1	64	60	
C2	60	61	
C3	59	70	
C4	68	71	
		OR	

- Q.5 (a) Explain the terms "Tree Induction" and "Tree Pruning" in 07 detail.
  - (b) Write a short note on "Naïve Bayesian Classifiaction". 07
    Observe following table. Find out what will be the result for Mr. X = {Rank="Professor", Years=5}. Find out the promotion probability of Mr. X using Naïve Bayesian Classification

NAME	RANK	YEARS	Promotion
Mike	Assistant Prof	3	no
Mary	Assistant Prof	7	yes
Bill	Professor	2	yes
Jim	Associate Prof	7	yes
Dave	Assistant Prof	6	no
Anne	Associate Prof	3	no

\*\*\*\*\*