

GUJARAT TECHNOLOGICAL UNIVERSITY**SEMESTER– 4 EXAMINATION – WINTER 2012****Subject code: 640005****Date: 11/01/2013****Subject Name: Elective I – Data Warehousing & Data Mining (DWDM)****Time: 10:30 – 13:00****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) What is Data-warehouse? Explain it with its characteristics. **07**

(b) Define KDD process with its steps to generate knowledge **07**

Q.2 (a) Differentiate OLTP and OLAP. **07**

(b) Write a short note on Data-Mart. **07**

OR

(b) Explain Star schema with suitable example? **07**

Q.3 (a) Explain types of OLAP operations with suitable example. **07**

(b) Explain Data Mining architecture in detail. **07**

OR

Q.3 (a) Explain the terms Data Cleaning, Data Integration & Transformation with suitable example **07**

(b) Write a short note on usage of “Data Mining in Retail and Marketing”. **07**

Q.4 (a) Write a short note on Apriori algorithm and apply Apriori on following data-set and find out frequent combination/Item-set for 3 products where minimum support =2. **07**

Transaction Number	Items Purchased
10	A, C, D
20	B, C, E
30	A, B, C, E
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 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. **07**

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

City	X	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 detail. **07**
- (b)** Write a short note on “Naïve Bayesian Classification”. 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 **07**

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
