Enrollment No: _____ Exam Seat No:_____

C.U.SHAH UNIVERSITY

Winter Examination-2015

Subject Name : Data Warehousing & Data Mining

Subject Code :5CS05DWD1 **Branch :MCA**

Semester :5 **Date :7/12/2015** **Time : 2:30 To 5:30** **Marks :70**

Instructions:

(1) Use of Programmable calculator and any other electronic instrument is prohibited.
(2) Instructions written on main answer book are strictly to be obeyed.
(3) Draw neat diagrams and figures (if necessary) at right places.
(4) Assume suitable data if needed.

SECTION – I

Q-1 **Attempt the Following questions** **(07)**

a.	What is Data mart?	**1**
b.	Define Support & Confidence.	**1**
c.	What is Multi-Dimensional Analysis?	**1**
d.	Define Association Rule Mining.	**1**
e.	What is Meta Data Repository?	**1**
f.	Define Data Cleaning.	**1**
g.	What is Enterprise Warehouse?	**1**

Q-2 **Attempt all questions** **(14)**

a.	Differentiate between OLTP and OLAP.	**5**
b.	Explain Star Schema with suitable examples.	**5**
c.	What is Data Warehouse? Draw Architecture of Data Warehouse & Explain it.	**4**

OR

Q-2 **Attempt all questions** **(14)**

a.	What is Data Cleaning? How to Handle Missing Values in Data Cleaning?	**5**
b.	Explain fact Constellation Schema with suitable example.	**5**
c.	Explain Data Integration and Transformation.	**4**

Q-3 **Attempt all questions** **(14)**

a.	Explain Data Mining Coupling with examples.	**7**
b.	Explain types of OLAP servers with suitable example.	**7**

OR

Q-3	**a.**	Explain OLAP operations with suitable examples.	**7**
	b.	Which are the Major Issues in Data Mining? Explain it.	**7**

SECTION – II

Q-4 **Attempt the Following questions** **(07)**

a. What is Decision tree? **1**

b. Define Binary variables **1**

c. What is Prediction? **1**

d. What do you mean by Cluster Analysis? **1**

e. What is the use of Regression? **1**

f. Define DMQL **1**

g. How to find distance between two objects in Cluster? **1**

Q-5 **Attempt all questions** **(14)**

a. Apply Apriori on following data-set and find out frequent combination/Item-set for 9 products where minimum support =2. **5**

Tid	List of Item IDs
1	I1,I2,I5
2	I2,I4
3	I2,I3
4	I1,I2,I4
5	I1,I3
6	I2,I3
7	I1,I3
8	I1,I2,I3,I5
9	I1,I2,I3

b. A database has the following transactions. Find Confidence and Support **5**

TID	Items
1	Bread, Milk
2	Bread, Diaper,
3	Milk, Diaper, Beer,
4	Bread, Milk,
5	Bread, Milk,

Milk,Diaper} → {Beer}
{Milk,Beer} → {Diaper}
{Diaper,Beer} → {Milk}
{Beer} → {Milk,Diaper}
{Diaper} → {Milk,Beer}
{Milk} → {Diaper,Beer}

c. What is Classification? Explain Classification by Decision Tree. **4**

OR

Q-5 **a.** Correlation with Lift (A, B). Define the Lift (A, B) measurement between 2 products TV and DVD where there are total 10000 transactions carried out. Out of which TV is sold in 6000 transactions. DVD is sold in 7500 transactions. TV and DVD together sold in 4000 transactions. Find out the Lift (TV, DVD) correlation. **5**

b. Explain Information gain, Gain ration and Gini Index. **5**

c. What is Rule-based classification? Explain using IF-THEN Rules for classification. **4**

Q-6		**Attempt all questions**	**(14)**
	a.	Suppose that the data mining task is to cluster the following eight points (with (x, y)representing location) into three clusters:A1(4,6), A2(2,5), A3(9, 3),A4(6, 9), A5(7, 5), A6(5, 7), A7(2, 2), A8(6, 6):The distance function is Euclidean distance. Suppose initially we assign A1, A2, and A3as the center of each cluster, respectively. Use the k-means algorithm to show only (a) The three cluster centers after the first round execution	**7**
	b.	Discuss the Application of Financial Data analysis.	**7**

OR

Q-6		**Attempt all Questions**	
	a.	Suppose that the data mining task is to cluster the following eight points (with (x, y)representing location) into three clusters:A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8), B2(7, 5), B3(6, 4), C1(1, 2), C2(4, 9):The distance function is Euclidean distance. Suppose initially we assign A1, B3, and C2as the center of each cluster, respectively. Use the k-means algorithm to show only (a) The three cluster centers after the first round execution	**7**
	b.	Discuss the Application of Telecommunication Industries.	**7**