Enrollment No: _____ Exam Seat No:_____

C.U.SHAH UNIVERSITY

Winter Examination-2018

Subject Name: Data Warehousing and Data Mining

Subject Code: 5CS05MDW1 Branch: MCA

Semester: 5 Date: 03/12/2018 Time: 10:30 To 01:30 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.	Define Data Cleaning.	**1**
	b.	What is smoothing?	**1**
	c.	What is do you mean by concept hierarchies?	**1**
	d.	Define support.	**1**
	e.	Name some Conventional Visualization techniques.	**1**
	f.	What is a metadata?	**1**
	g.	What is data mart?	**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? Briefly explain the key words used in the definition.	**4**
		OR	
Q-2		**Attempt all questions**	**(14)**
	a.	Explain the KDD process with suitable figure.	**5**
	b.	Explain Snowflake Schema with suitable Examples	**5**
	c.	Explain with figure: 3-Tire Data Warehouse Architecture.	**4**
Q-3		**Attempt all questions**	**(14)**
	a.	What is Indexing Data? Explain Different type of Indexing 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.	Explain Data Mining Task Primitives.	**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 is linear Regression? **1**

e. Define Data Classification. **1**

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

g. What are Facts? **1**

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

a. Explain Data Integration and Data Cleaning. **5**

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

TID	Mobile Colors Purchased
1	red, white, green
2	white, orange
3	red, white, orange
4	white, blue
5	red, blue
6	red, white, blue, green
7	red, white, blue, orange

{red, white} → green

{white } → {blue, green}

{red} → {white, orange}

{white } → { orange }

{red} → {white, blue, green}

c. Explain different types of Normalization in Data Mining. **4**

OR

Q-5 **Attempt all questions**

a. Explain Information Gain and Gain Ratio with suitable examples. **5**

b. Apply Apriori on following data-set and find out frequent combination/Item-set for 3 products where minimum support =2. **5**

Transaction Number	Items Purchased
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B,E

c. What is Rule-based Classification? Explain Using IF-THEN Rules for Classification? **4**

Q-6 **Attempt all questions** **(14)**

a. Explain the Application of Financial Data analysis. **7**

b. 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),A4(5,8), A5(7, 5), A6(6,4), A7(1, 2), A8(4, 9):The distance function is Euclidean distance. Suppose initially we assign A1 (2, 10), A4 (5, 8) and A7 (1, 2)as the center of each cluster, respectively. Use K-means algorithms to find the three clusters after the first iteration. **7**

OR

Q-6 **Attempt all Questions**

a. Write a note on types of clusters. **7**

b. Explain the application of Retail industry. **7**