

Enrollment No: \_\_\_\_\_

Exam Seat No: \_\_\_\_\_

# C.U.SHAH UNIVERSITY

## Summer Examination-2020

Subject Name: Data Warehousing & Data Mining

Subject Code: 5CS05DWD1

Branch: MCA

Semester : 5

Date : 04/03/2020

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. What is Dimensional Tables 1
  - b. What are the characteristics of data warehouse? 1
  - c. What are Time-Series databases? 1
  - d. What are the steps in the data Mining process? 1
  - e. What is Linear Regression? 1
  - f. What is base cube ID? 1
  - g. What is metadata repository? 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. Define KDD process with its steps to generate knowledge. 5
  - b. Explain Snowflake Schema with suitable Examples. 5
  - c. Explain Attribute Subset Selection Method. 4
- Q-3 Attempt all questions (14)**
- a. Explain OLAP operations with suitable examples. 7
  - b. Explain Data Mining Task Primitives. 7
- OR**
- Q-3**
- a. What is Indexing Data? Explain Different type of Indexing with examples. 7
  - b. Explain types of OLAP servers with suitable example. 7

### SECTION – II

- Q-4 Attempt the Following questions (07)**
- a. Define Predictive model. 1
  - b. What is Cluster analysis? 1
  - c. What is Hierarchical method? 1



- d. What is Association rule? 1
  - e. What is Bayesian classification? 1
  - f. What is Market Basket Analysis? 1
  - g. What is Accuracy in classification? 1
- Q-5 Attempt all questions (14)**
- a. Explain Data Integration and Transformation. 5
  - b. A 2X2 contingency table summarizing the transactions with respect to game and video purchases. 5

	Game	Game Not	Sum of Row
Video	3000	2500	5500
Not Like Video	2000	2500	4500
Sum of Column	5000	5000	10000

Find Correlation analysis using  $\chi^2$ .

- c. Explain Information Gain and Gain Ratio with suitable examples. 4

**OR**

- Q-5 Attempt all questions**
- a. What is Rule-based Classification? Explain Using IF-THEN Rules for Classification? 5
  - b. Correlation with Lift (A, B). Define the Lift (A, B) measurement between 2 products TV and GAME 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 GAME together sold in 2000 transactions. Find out the Lift (TV, GAME) correlation. 5
  - c. Explain Naïve Bayesian classification with examples. 4

- Q-6 Attempt all questions (14)**
- a. Explain the Application of Financial Data analysis. 7
  - b. Cluster the following eight points (with (x, y) representing locations) into three clusters: 7  
A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9).  
Assume that initial cluster centers are: A2 (2, 5), A5 (5, 8) and A8 (1, 2) Use Manhattan distance for the distance function between two points after the first iteration.

**OR**

- Q-6 Attempt all Questions**
- a. Explain Clustering K-Means method with suitable examples. 7
  - b. Discuss the Application of Telecommunication Industries. 7

