C. U. SHAH UNIVERSITY
Summer Examination 2019

Subject Name: **Data Mining and Data Warehousing**

Subject Code: **4CS06BDD1**   Branch: **B. C. A.**

Semester: **6**   Date: **30/04/2019**   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.

Q. 1	**Answer the below questions:**	**14**
a)	Write the full forms of:	**5**
	(A) DSS (B) WEKA (C) OLTP (D) ETL (E) OLAP	
b)	OLAP is used to explore the _____ knowledge.	**1**
	(A) Multidimensional (B) Shallow (C) Deep (D) Hidden	
c)	A goal of data mining includes which of the following?	**1**
	(A) To confirms that data is exists	
	(B) To analyze data for expected relationships	
	(C) To explain some observed event	
	(D) Near real-time updates	
d)	Data warehouse architecture is based on _____	**1**
	(A) SQL Server (B) DBMS (C) MySQL (D) RDBMS	
e)	Define Data Mart.	**2**
f)	List out various ETL Tools	**2**
g)	What do you mean by BI tools?	**2**

ATTEMPT ANY FOUR QUESTIONS FROM Q. 2 TO Q. 8

Q. 2	**Attempt all questions.**	
a)	Explain Data Warehouse System architecture.	**7**
b)	Explain Data Mart Structure.	**7**
Q. 3	**Attempt all questions.**	
a)	Write note on Transformation of Data.	**7**
b)	What is Data Mining? Explain Data Mining Process.	**7**
Q. 4	**Attempt all questions.**	

a)	Differentiate between Data Warehouse and Data Mart.	**7**
b)	How Correlation and Regression can be used in Data Mining Techniques?	**7**

Q. 5 **Attempt all questions.**

a)	Discuss about Mutation and Fitness Functions.	**7**
b)	What is Data Cleaning? How to missing values in Data Cleaning can be handled?	**7**

Q. 6 **Attempt all questions.**

a)	Explain Partitioning algorithm in detail.	**7**
b)	Explain characteristics of Data Warehouse	**7**

Q. 7 **Attempt all questions.**

a)	Explain HOLAP with advantages and disadvantages.	**7**
b)	Explain deployment process of Data Mining model.	**7**

Q. 8 **Attempt all questions.**

a)	Explain Comparison and contradiction of various ETL toots.	**7**
b)	Apply Apriori Algorithm on following dataset and find our frequent combination / Item Set for 10 products where minimum support = 2.	**7**

TransactionID	Item List (ItemID)	TransactionID	Item List (ItemID)
1001	C101, C103, C106	1006	C101, C103, C104
1002	C102, C103	1007	C105, C106
1003	C101, C102, C104	1008	C101, C102, C103
1004	C102, C106	1009	C102, C105, C106
1005	C103, C104, C105	1010	C103, C104, C105