GUJARAT TECHNOLOGICAL UNIVERSITY
BE - SEMESTER–VII (NEW) - EXAMINATION – SUMMER 2017

Subject Code: 2170715   **Date: 09/05/2017**
Subject Name: Data Mining and Business Intelligence
Time: 02.30 PM to 05.00 PM   **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)**	1. What is Cuboid? Explain any three OLAP Operations on Data Cube with example.	**04**
		2. Compare OLTP & OLAP systems.	**03**
	(b)	Explain Star, Snowflake, and Fact Constellation Schema for Multidimensional Database.	**07**
Q.2	**(a)**	What is Data Mining? Why is it called data mininig rathar knowledge mininig? Explain KDD process.	**07**
	(b)	1. Minimum salry is 20,000Rs and Maximum salary is 1,70,000Rs. Map the salry 1,00,000Rs in new Range of (60,000 , 2,60,000) Rs using min-max normalization method.	**04**
		2. If Mean salary is 54,000Rs and standard deviation is 16,000 Rs then find z score value of 73,600 Rs salry.	**03**

OR

	(b)	Explain Mean, Median, Mode,Variance, Standard Deviation & five number summay with suitable database example.	**07**
Q.3	**(a)**	State the Apriori Property. Generate candidate itemsets, frequent itemsets and association rules using Apriori algorithm on the following data set with minimum support count is 2.	**07**

	TID	List of *items_IDs*
1.	T100	I1,I2,I5
2.	T200	I2,I4
3.	T300	I2,I3
4.	T400	I1,I2,I4
5.	T500	I1,I3
6.	T600	I2,I3
7.	T700	I1,I3
8.	T800	I1,I2,I3,I5
9.	T900	I1,I2,I3

	(b)	What is Concept Hierarchy? List and explain types of Concept Hierarchy.	**07**

OR

Q.3	**(a)**	1.What is Market Basket Analysis? Explain Association Rules with Confidence & Support.	**04**
			03

2. List and describe methods for handling missing values in data cleaning.

(b) What is noise? Explain data smoothing methods as noise removal technique to divide given data into bins of size 3 by bin partition (equal frequency), by bin means, by bin medians and by bin boundaries. Consider the data: **07**

10, 2, 19, 18, 20, 18, 25, 28, 22

Q.4 **(a)** Explain mining in following Databases with example. **07**
1. Temporal Databases
2. Sequence Databases
3. Spatial Databases
4. Spatiotemporal Databases.

(b) What is the importance of visualization of discovered patterns? Explain the role of presentation in pattern visualization. Discuss various visualization techniques in KDD. **07**

OR

Q.4 **(a)** Define following terms & differenciate them: **07**

Data Mart , Enterprise Warehouse & Virtual Warehouse

(b) Explain role of Business intelligence in any one of following domain: **07**
Fraud Detection,Market Segmentation,retail industry, telecommunications industry. Explain how data mininig can be helpful in any of these cases.

Q.5 **(a)** What is outlier analysis? Why outlier mining is important? Briefly describe the different approaches : statistical-based outlier detection, distance-based outlier detection and deviation-based outlier detection. **07**

(b) Explain Hadoop Architecture. **07**

OR

Q.5 **(a)** What is Decision Tree? Explain how classification is done using decision tree induction. **07**

(b) What is Big Data? What is big data analytic ? Explain the big data- distributed file system. **07**

****