

GUJARAT TECHNOLOGICAL UNIVERSITY**BE - SEMESTER-VII (NEW) EXAMINATION – WINTER 2018****Subject Code: 2170715****Date: 03/12/2018****Subject Name: Data Mining and Business Intelligence****Time: 10:30 AM TO 01:00 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) What is Data Mining? Why is it called data mining rather knowledge mining? **03**

(b) Explain various features of Data Warehouse? **04**

(c) Differentiate between Operational Database System and Data Warehouse **07**

Q.2 (a) What is the difference between KDD and Data Mining? **03**

(b) What is Concept Hierarchy? List and briefly explain types of Concept Hierarchy **04**

(c) Explain Mean, Median, Mode, Variance, Standard Deviation & five number summary with suitable database example. **07**

OR

(c) 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. **07**

Consider the data: 10, 2, 19, 18, 20, 18, 25, 28, 22

Q.3 (a) Differentiate Fact table vs. Dimension table **03**

(b) Suppose that the data for analysis includes the attribute *age*. **04**

The *age* values for the data tuples are (in increasing order):
13, 15, 16, 16, 19, 20, 23, 29, 35, 41, 44, 53, 62, 69, 72

Use min-max normalization to transform the value 45 for age onto the range [0:0, 1:0]

(c) Explain mining in following Databases with example. **07**

1. Temporal Databases
2. Sequence Databases
3. Spatial Databases
4. Spatiotemporal Databases.

OR

Q.3 (a) List and describe methods for handling missing values in data cleaning. **03**

(b) Explain the following as attribute selection measure: **04**

- (i) Information Gain
- (ii) Gain Ratio

(c) Explain three tier data warehouse Architecture in details. **07**

- Q.4** (a) How K-Mean clustering method differs from K-Medoid clustering method? **03**
- (b) Define data cube and explain 3 operations on it. **04**
- (c) State the Apriori Property. Generate large itemsets and association rules using Apriori algorithm on the following data set with minimum support value and minimum confidence value set as 50% and 75% respectively **07**

<i>TID</i>	<i>Items Purchased</i>
T101	Cheese, Milk, Cookies
T102	Butter, Milk, Bread
T103	Cheese, Butter, Milk, Bread
T104	Butter, Bread

OR

- Q.4** (a) Define following terms : **03**
Data Mart , Enterprise Warehouse & Virtual Warehouse
- (b) Discuss the application of data warehousing and data mining **04**
- (c) What is web log? Explain web structure mining and web usage mining in detail **07**

- Q.5** (a) Draw the topology of a multilayer, feed-forward Neural Network. **03**
- (b) Explain Linear regression with example. **04**
- (c) Explain the major issues in data mining **07**

OR

- Q.5** (a) Briefly explain text mining **03**
- (b) What is market basket analysis? Explain the two measures of rule interestingness: *support* and *confidence* **04**
- (c) What is Big Data? What is big data analytic? Explain the big data-distributed file system. **07**
