

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER-VII (OLD) - EXAMINATION – SUMMER 2018

Subject Code:171601

Date:08/05/2018

Subject Name:Data warehousing and Data Mining

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) Explain: Star”, “Snowflake” and “Fact constellation” schemas of data warehouse. **07**

(b) Clearly state the differences between “Data Warehouses” and “Operational Database Systems”. **07**

Q.2 (a) Define the term “Data Mining”. With the help of a suitable diagram explain the process of knowledge discovery from databases **07**

(b) What is data cleaning? Describe the different methods of handling missing values during data cleaning **07**

OR

(b) What is data transformation? Explain the different data transformation approaches for transforming data. **07**

Q.3 (a) What is Concept Hierarchy? List and explain different types of Concept Hierarchy. **07**

(b) What is an ‘outlier’? How do outliers impact the results of mining? Explain any one method to detect outliers **07**

OR

Q.3 (a) State the Apriori Property. Generate large itemsets and association rules using Apriori algorithm on the following data set having minimum support value as 2 and minimum confidence value as 75%. **07**

| TID | Items Purchased |
|------|-----------------------------|
| T101 | Cheese, milk, cookie |
| T102 | Butter, milk, bread |
| T103 | Cheese, butter, milk, bread |
| T104 | Butter, bread |

(b) What is noise? Explain data smoothing methods as noise removal technique to divide given data into bins of size 3 by equal frequency bin partitioning, by bin means, by bin medians and by bin boundaries. Consider the data: 10, 2, 19, 18, 20, 18, 25, 28, 22 **07**

Q.4 (a) What is supervised and unsupervised learning? What is Cluster Analysis? List and explain requirements of clustering in data mining. **07**

(b) What are the methods to evaluate accuracy of classifier/predictor? **07**

OR

Q.4 (a) Write steps of K-Means clustering algorithm with its pros and cons. How K-Mean clustering method differs from K-Medoid clustering method? **07**

(b) State and explain attribute selection measure with suitable example. **07**

Q.5 (a) List and explain various OLAP operations. **07**

(b) What is time series database? How to characterize the time series data using trend analysis? **07**

OR

- Q.5** (a) What is web log? Explain web structure mining and web usage mining in detail. **07**
(b) Discuss the characteristics and limitations of Neural Networks. Mention any two suitable applications of neural networks. **07**
