

**GUJARAT TECHNOLOGICAL UNIVERSITY****BE - SEMESTER-VII(NEW) • EXAMINATION – WINTER 2016****Subject Code:2170715****Date:29/11/2016****Subject Name:Data Mining and Business Intelligence(Departmental Elective - II)****Time:10.30 AM to 1.00 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 why data warehouses are needed for developing business solutions from today's perspective. Discuss the role of data marts. **07**  
 (b) Give differences between OLAP and OLTP. **07**
- Q.2** (a) Define the term "data mining". Discuss the major issues in data mining. **07**  
 (b) With the help of a suitable example, illustrate the OLAP operations: 'drill-down', 'roll-up', 'slice' and 'dice'. **07**
- OR**
- (b) Explain the following data normalization techniques: (i) min-max normalization and (ii) decimal scaling. **07**
- Q.3** (a) What are the limitations of the Apriori approach for mining? Briefly describe the techniques to improve the efficiency of Apriori algorithm. **07**  
 (b) In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem. **07**
- OR**
- Q.3** (a) What is market basket analysis? Explain the two measures of rule interestingness: *support* and *confidence* with suitable example. **07**  
 (b) Explain the need for data smoothing during pre-processing and discuss data smoothing by Binning. **07**
- Q.4** (a) With the help of a neat diagram explain the topology of a multilayer, feed-forward Neural Network. Also explain the terms: "activation function" and "epoch". **07**  
 (b) Briefly explain Linear and Non-linear regression. **07**
- OR**
- Q.4** (a) Explain the steps of the ID3 algorithm for generating Decision trees. **07**  
 (b) Explain the methods to measure the accuracy of a classifier or predictor. **07**
- Q.5** (a) Briefly explain the life-cycle of Data Analytics and discuss the role of data scientists. **07**  
 (b) What is meant by "clustering"? Explain why clustering is called unsupervised learning. Mention any two applications of clustering. **07**
- OR**
- Q.5** (a) Discuss applications of data mining in Banking and Finance. **07**  
 (b) Discuss the main features of Hadoop Distributed File System. **07**

\*\*\*\*\*