GUJARAT TECHNOLOGICAL UNIVERSITY
BE – SEMESTER – VIII.EXAMINATION – WINTER 2016

Subject Code: 183103   **Date: 22/10/2016**
Subject Name: Business Intelligence & Data Mining (Department Elective II)
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)**	Define data mining. Draw and explain KDD process using a neat diagram.	**07**
	(b)	Discuss the differences between data warehouse and operational database management system.	**07**
Q.2	**(a)**	What is the need of data normalization? List and discuss such normalization techniques with proper example.	**07**
	(b)	Explain generalization using attribute oriented induction with an appropriate example.	**07**
		OR	
	(b)	What is the importance of outliers? Write a note on binning to smooth the data.	**07**
Q.3	**(a)**	For given following transactional database (Refer Table 1), generate frequent itemsets using Apriori algorithm. Consider minimum support count=2.	**07**

Table 1

TID	List of Items
T1	I1, I2, I5
T2	I2, I4
T3	I2, I3
T4	I1, I2, I4
T5	I1, I3
T6	I2, I3
T7	I1, I3
T8	I1, I2, I3, I5
T9	I1, I2, I3

	(b)	Write a note on following: i) Association Rule Mining ii) Support and Confidence.	**07**
		OR	
Q.3	**(a)**	Compare FP-growth approach with Apriori and prepare FP-growth tree for transactional database given in Table 1.	**07**
	(b)	List major limitations of Apriori algorithm. Also, write and explain various alternatives to improve the efficiency of Apriori algorithm.	**07**
Q.4	**(a)**	Briefly discuss various steps of decision tree induction for classification.	**07**

(b) Write a note on following   **07**
     i) Information Gain
     ii) Tree Pruning

<div align="center">OR</div>

Q.4 **(a)** Following table (Refer Table 2) shows a set of paired data, where X is the number of years of work experience of a college graduate and Y is the corresponding salary of the graduate. Predict the salary of a college graduate who has 10 years of work experience using linear regression.   **07**

<div align="center">Table 2</div>

X (work experience in years)	Y (Salary in thousands)
3	30
8	57
9	64
13	72
3	36
6	43
11	59
21	90
1	20
16	83

Q.4 **(b)** Explain following :   **07**
     i) Bayesian Classification
     ii) k-Nearest Neighbor classifiers
     iii) Rough Set Approach

Q.5 **(a)** Define data warehouse. Draw and explain three tier architecture of data warehouse.   **07**

(b) Explain data mining issues related to performance and diversity of data types.   **07**

<div align="center">OR</div>

Q.5 **(a)** List and discuss various OLAP operations using example and diagram.   **07**

(b) Discuss the following: **07**
     i) Data Mart
     ii) Business Intelligence
     iii) Life cycle of Data

<div align="center">*************</div>