Time: 3 hours **Marks: 80**

N.B. 1. Question 1 is compulsory

2. Attempt any three questions out of the remaining five questions

Q.1 (a) Demonstrate with a diagram the process of KDD. **(5)**

(b) Describe the different types of attributes one may come across in data mining with two examples of each. **(5)**

(c) Use k means clustering to cluster the following data into 2 clusters. 2,3,4,10,11,12,20,25,30. **(5)**

(d) Find Mean, median, mode for a given data. Show box plot. 11,13,13,15,15,16,19,20,20,21,21,22,23,24,30,40,45,45,45 **(5)**

Q.2 (a) Illustrate any one classification technique for the following dataset. Show how we can classify new tuple(HOMEOWNER=Yes, Status= Employed, Income=Average) **(10)**

ID	Homeowner	Status	Income	Defaulted
1	Yes	Employed	High	No
2	No	Business	Average	No
3	No	Employed	Low	No
4	Yes	Business	High	No
5	No	UnEmployed	Average	Yes
6	No	Business	Low	No
7	Yes	UnEmployed	High	No
8	No	Employed	Average	Yes
9	No	Business	Low	No
10	No	Employed	Average	Yes

(b) Explain different methods that can be used to evaluate and compare the accuracy of different classification algorithms. **(10)**

0C2D777098F63A6EE2ADA0FE586EEB14

Q.3 (a) Explain multilevel and multi dimensional association rules with examples. **(10)**

(b) What is market basket analysis? Give Apriori algorithm **(10)**

Q.4 (a) Discuss Supervised, Semi supervised and Unsupervised outlier detection methods. **(10)**

(b) What is the need of pre-processing. Explain the different steps involved in data pre-processing. **(10)**

Q.5 (a) Explain simple linear regression with example **(10)**

(b) Define BI and give its architecture. Explain any business application where data mining can be used. **(10)**

Q.6 (a) Use any hierarchical clustering algorithm to cluster the following into 3 clusters. a1=(2,10), a2=(2,5),a3=(8,4),a4=(5,8),a5=(7,5),a6=(6,4),a7=(1,2),a8=(4,9) **(10)**

(b) Explain DBSCAN algorithm with example. **(10)**

****

0C2D777098F63A6EE2ADA0FE586EEB14