Data Mining & Warehousing (CT#1)

1. Describe Data mining as a process of knowledge discovery. (4)
2. Differentiate between incomplete, noisy and inconsistent data? Describe the process of recovery. (6)
3. Suppose a group of 12 sales price records has been sorted as follows: 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215. Partition them into three bins by each of the following methods. (a) equal-frequency (equidepth) partitioning (b) equal-width partitioning (c) clustering. (6)
4. Explain the data analytical tools: Boxplot, Quantile plot, Scatter plot and Loess curve. (4)

Data Mining and Warehousing (CT#2)

1. LDA mainly projects the features in higher dimension space to lower dimensions- Explain how? (5)
2. Is LDA is linear classifier or dimensionality reduction technique? Place your arguments. (4)
3. The major challenge in Decision tree is the identification of the attribute for the root node in each level. This process is known as *attribute selection*. Explain one of the popular attribute selection measures. (7)
4. Define "Naïve Bayes Classifier". (4)

Data Mining and Warehousing (CT#3)

1. Consider the 9 two-dimensional data points: x1(0,0), x2(1,0), x3(1,1), x4(2,2), x5(3,1), x6(3,0), x7(0,1), x8(3,2), x9(6,3). Use the Euclidean Distance with Eps =1 and MinPts = 3. Find all core points, border points and noise points, and show the final clusters using DBCSAN algorithm. (8)
2. Find the differences: i) Soft clustering and Hard Clustering, and ii) K-means and Expectation Maximization. (4)
3. Consider the six points: a(3,3), b(4,10), c(9,6), d(14,8), e(18,11) and f(21,7), show 2 steps of fuzzy clustering using EM. (8)

8.47, 5.
10.42,

Data Mining and Warehousing (CT#4)

1. What are the steps of Apriori algorithm? Why is it called Apriori algorithm? (6)
2. For the following transactional dataset, identify frequent itemset using Dynamic Itemset Counting consider support = 50% and M= 2. (10)

Transactional Data	Item Purchased
T1	A,B,C
T2	B,C,D
T3	D,E
T4	A,B,D
T5	A,B,D
T6	A,B,C,D

3. Which procedure you will prefer in prediction i.e., linear or non linear regression? Explain the reason behind. (4)