MACHINE LEARNING ASSIGNMENT – 2
Q1 to Q12 have only one correct answer.
Choose the correct option to answer your question.

1. Movie Recommendation systems are an example of:
a) 2 Only

2. Sentiment Analysis is an example of:
d) 1, 2 and 4

3. Can decision trees be used for performing clustering?
a) True

4. Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:
a) 1 only

5. What is the minimum no. of variables/ features required to perform clustering?
b) 1

6. For two runs of K-Mean clustering is it expected to get same clustering results?
b) No

7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
a) Yes

8. Which of the following can act as possible termination conditions in K-Means?
d) All of the above

9. Which of the following can act as possible termination conditions in K-Means?
a) 1 only

10. Which of the following algorithms is most sensitive to outliers?
a) K-means clustering algorithm

11. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):
d) All of the above

12. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?
d) All of the above

Q13 to Q15 are subjective answers type questions, Answers them in their own words briefly
13. Is K sensitive to outliers?
Ans: The *K*-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values. *K*-medoids clustering is a variant of *K*-means that is more robust to noises and outliers. Instead of using the mean point as the centre of a cluster, *K*-medoids uses an actual point in the cluster to represent it. Medoid is the most centrally located object of the cluster, with minimum sum of distances to other points

14. Why is K means better?
Ans: *K*-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable *K*. The algorithm works iteratively to assign each data point to one of *K* groups based on the features that are provided. Data points are clustered based on feature similarity.

15. Is K means a deterministic algorithm?
Ans: The basic k-means clustering is based on a non-deterministic algorithm. This means that running the algorithm several times on the same data, could give different results. However, to ensure consistent results, FCS Express performs k-means clustering using a deterministic method.