

## ASSIGNMENT – 2

1. Movie Recommendation systems are an example of:

Ans: b) 1 and 2 (Classification and Clustering)

2. Sentiment Analysis is an example of:

Ans: b) 1 and 2 (Regression and Classification)

3. Can decision trees be used for performing clustering?

Ans: 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:

Ans: a) 1 only (Capping and flooring of variables)

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

Ans: b) 1

6. For two runs of K-Mean clustering is it expected to get same clustering results?

Ans: B. no (yes if you are not changing a code but if you are changing random\_state number the results also different )

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

Ans: a) Yes

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

Ans: d) All of the above

9. Which of the following algorithms is most sensitive to outliers?

Ans: a) K-means clustering algorithm

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

Ans: a) 1 only (Creating different models for different cluster groups.)

11. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?

Ans: d) All of the above

12. Is K sensitive to outliers?

Ans: yes kmeans sensitive to outliers when the outliers present the data kmeans create separate clusters either merge the existing cluster. And this clusters is overfitting or less accuracy.

13. Why is K means better?

Ans: k means is easy to understand and adapt new example.

scale to large data set

K- means creates a specified clusters and we can focus on specified cluster to business growth.

14. Is K means a deterministic algorithm?

Ans: Yes, K-mean is deterministic algorithm. But give always give same output and same data and parameter. this algorithm is also sensitive to outliers.