Q1 to Q11 have only one correct answer. Choose the correct option to answer your question.

1. Movie Recommendation systems are an example of:
   i) Classification
   ii) Clustering iii) Regression
   Options:
   a) 2 Only
   b) 1 and 2
   c) 1 and 3
   d) 2 and 3

Ans - b

2. Sentiment Analysis is an example of:
   i) Regression
   ii) Classification iii) Clustering iv) Reinforcement Options:
   a) 1 Only
   b) 1 and 2
   c) 1 and 3
   d) 1, 2 and 4

Ans - d

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

Ans - a

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: i) Capping and flooring of variables
   ii) Removal of outliers Options:
   a) 1 only
   b) 2 only
   c) 1 and 2
   d) None of the above

Ans - a

5. What is the minimum no. of variables/ features required to perform clustering? a) 0
   b) 1
   c) 2
   d) 3

Ans – b

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

Ans - b

7. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
   a) Yes
   b) No
   c) Can't say
   d) None of these **Ans - a**

MACHINE LEARNING

8. Which of the following can act as possible termination conditions in K-Means?
   i) For a fixed number of iterations. ii) Assignment of observations to clusters does not change between iterations. Except for cases with a bad local minimum.
   iii) Centroids do not change between successive iterations.
   iv) Terminate when RSS falls below a threshold. Options:
   a) 1, 3 and 4
   b) 1, 2 and 3
   c) 1, 2 and 4
   d) All of the above

Ans - d

9. Which of the following algorithms is most sensitive to outliers?
   a) K-means clustering algorithm
   b) K-medians clustering algorithm
   c) K-modes clustering algorithm
   d) K-medoids clustering algorithm

Ans - a

10. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):
    i) Creating different models for different cluster groups.
    ii) Creating an input feature for cluster ids as an ordinal variable. iii) Creating an input feature for cluster centroids as a continuous variable.
    iv) Creating an input feature for cluster size as a continuous variable. Options:
    a) 1 only
    b) 2 only
    c) 3 and 4
    d) All of the above

Ans - d

11. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?
    a) Proximity function used
    b) of data points used
    c) of variables used
    d) All of the above

Ans - d

Q12 to Q14 are subjective answers type questions, Answers them in their own words briefly

12. Is K sensitive to outliers?

Ans - The *k*-means algorithm is sensitive to the outliers. Because the mean, as a statistic, is generally sensitive to outliers.

The mean of 2,2,2,3,3,3,4,4,4 is 3

If we add a single 23 to that, the mean becomes 5, which is larger than any of the other values.

Since in k-means, you'll be taking the mean a lot, you wind up with a lot of outlier-sensitive calculations.

13. Why is K means better?

Ans - K-means clustering is a very famous and powerful unsupervised machine learning algorithm. It is used to solve many complex unsupervised machine learning problems. A K-means clustering algorithm tries to group similar items in the form of clusters. The number of groups is represented by K. So there are two methods by which you can select the right value of k.

1. Elbow Method.
2. Silhouette Method.

14. 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.