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

1. Movie Recommendation systems are an example of:
   1. Classification
   2. Clustering
   3. Regression Options:
      1. 2 Only
      2. 1 and 2
      3. 1 and 3
      4. 2 and 3
2. Sentiment Analysis is an example of:
   1. Regression
   2. Classification
   3. Clustering
   4. Reinforcement Options:
3. 1 Only
4. 1 and 2
5. 1 and 3
6. 1, 2 and 4
7. Can decision trees be used for performing clustering?
8. True
9. False
10. Which of the following is the most appropriate strategy for data cleaning before performing clustering analysis, given less than desirable number of data points:
    1. Capping and flooring of variables
    2. Removal of outliers Options:
       1. 1 only
       2. 2 only
       3. 1 and 2
       4. None of the above
11. What is the minimum no. of variables/ features required to perform clustering?
12. 0
13. 1
14. 2
15. 3
16. For two runs of K-Mean clustering is it expected to get same clustering results?
17. Yes
18. No
19. Is it possible that Assignment of observations to clusters does not change between successive iterations in K-Means?
20. Yes
21. No
22. Can't say
23. None of these
24. Which of the following can act as possible termination conditions in K-Means?
    1. For a fixed number of iterations.
    2. Assignment of observations to clusters does not change between iterations. Except for cases witha bad local minimum.
    3. Centroids do not change between successive iterations.
    4. Terminate when RSS falls below a threshold. Options:
       1. 1, 3 and 4
       2. 1, 2 and 3
       3. 1, 2 and 4
       4. All of the above
25. Which of the following algorithms is most sensitive to outliers?
26. K-means clustering algorithm
27. K-medians clustering algorithm
28. K-modes clustering algorithm
29. K-medoids clustering algorithm
30. How can Clustering (Unsupervised Learning) be used to improve the accuracy of Linear Regression model (Supervised Learning):
    1. Creating different models for different cluster groups.
    2. Creating an input feature for cluster ids as an ordinal variable.
    3. Creating an input feature for cluster centroids as a continuous variable.
    4. Creating an input feature for cluster size as a continuous variable. Options:
31. 1 only
32. 2 only
33. 3 and 4
34. All of the above
35. What could be the possible reason(s) for producing two different dendrograms using agglomerative clustering algorithms for the same dataset?
36. Proximity function used
37. of data points used
38. of variables used
39. All of the above

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

1. Is K sensitive to outliers?

YES

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.

1. Why is K means better?

Advantages of k-means:-

Guarantees convergence. Can warm-start the positions of centroids. Easily adapts to new examples. Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

1. Is K means a deterministic algorithm?

NO

The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids



# Only two pages were converted.

Please **Sign Up** to convert the full document.

[**www.freepdfconvert.com/membership**](https://www.freepdfconvert.com/membership)