MACHINE LEARNING

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

1) Movie Recommendation systems are an example of:

Ans -**a) 2 Only**

2) Sentiment Analysis is an example of:

Ans -**d) 1, 2 and 4**

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**

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**

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-**b) 2 only**

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**

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 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.

13)

Why is K means better?

Ans-

The k-means cluster analysis command is efficient primarily because it does not compute the distances between all pairs of cases, as do many clustering algorithms, including the algorithm that is used by the hierarchical clustering command

14)

Is K means a deterministic algorithm?

Ans-

Clustering algorithms with steps involving randomness usually give different results on different executions for the same dataset. This non-deterministic nature of algorithms such as the K-Means clustering algorithm limits their applicability in areas such as cancer subtype prediction using gene expression data. It is hard to sensibly compare the results of such algorithms with those of other algorithms. The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids.