ASSIGNMENT – 2 **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: i) Classification
ii) Clustering
iii) Regression Options:
d) 2 and 3

2. Sentiment Analysis is an example of: i) Regression
ii) Classification
iii) Clustering
iv) Reinforcement Options
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: i) Capping and flooring of variables
ii) Removal of outliers Options:
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

SSIGNMENT – 2 **MACHINE LEARNING**

8. Which of the following can act as possible termination conditions in K-Means? i) For a fixed number of erations.

ii) Assignment of observations to clusters does not change between iterations. Except for cases witha bad cal minimum.

iii) Centroids do not change between successive iterations.

iv) Terminate when RSS falls below a threshold. Options:

All of the above

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

a) K-means clustering algorithm

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

d) All of the above

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

a) Proximity function used

b) of data points used

c) of variables used

Answer d) All of the above

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

2. Is K sensitive to outliers?

K is sensitive to outliers because these is great effect on mean by the extreme far values from the center. Instead of ing the mean point as the center of a cluster, K-medoids uses an actual point in the cluster to represent it. Medoid is the ost centrally located object of the cluster, with minimum sum of distances to other points

3. Why is K means better?

is very simple to implement and run.

4. Is K means a deterministic algorithm?

means is non- deterministic as every time when we run algorithms on data it gives us different value,but gives us nsistent results.