MACHINE LEARNING WORKSHEET 2

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 34

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. d) All of the above

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. K-Means clustering algorithm is most sensitive to outliers as it uses the mean of cluster data points to find the cluster center.

13. Why is K means better?

Ans. Relatively simple to implement. Scales to large data sets. 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.

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.