

# Sachin Boora (DS2302)

<https://github.com/sachinboora/Internship>

## Worksheet\_set\_2

### Machine Learning Worksheet

**Ans 1 :** Movie Recommendation System is a clustering example(Option A).

**Ans 2 :** Option B - Regression and Classification

**Ans 3 :** False, Decision Trees cant be used for clustering.

**Ans 4 :** Option D is correct.

**Ans 5 :** Option A is correct - minimum 1 is required.

**Ans 6 :** Option B is correct.

**Ans 7 :** Option A is correct.

**Ans 8 :** Option D - All of the above.

**Ans 9 :** K-Means Clustering - Option A

**Ans 10 :** Option D - All of the above.

**Ans 11 :** Option D - All of the above.

## **Subjective Questions**

**Ans 12 :** Yes, K-Means clustering is sensitive to outliers. It is a popular unsupervised machine learning algorithm that groups similar data points together into clusters so even if a single data point is way different from other data points in a group it can significantly affects the result of the model. There are several

ways to deal with outliers including removal of that data point from dataset or using a different model like DBSCAN.

**Ans 13 :** K-means is a popular clustering algorithm that is widely used in many fields, including machine learning, data science, and business analytics. Here are some reasons why K-means is a good choice for clustering:

**Scalability:** K-means is a scalable algorithm that can be used on large data sets with a high number of variables. It can quickly cluster large datasets and is often faster than other clustering algorithms.

**Simplicity:** K-means is a simple and easy-to-understand algorithm that can be implemented quickly. It is often the first algorithm that data scientists use when they are new to clustering.

**Flexibility:** K-means is a flexible algorithm that can be used for a variety of clustering tasks. It can be used with different distance metrics and can be modified to handle categorical data or other special data types.

**Interpretability:** The results of K-means are easy to interpret. Each data point is assigned to a single cluster, making it easy to understand which points are similar to each other.

**Ans 14 :** Yes, K-means is a deterministic algorithm. This means that for a given data set and a fixed value of K (the number of clusters), the algorithm will always produce the same clustering result, regardless of how many times it is run.

The K-means algorithm works by iteratively updating the cluster centroids until they converge to a stable solution. The convergence is guaranteed because the algorithm is designed to minimize the within-cluster sum of squares (WCSS) objective function. The WCSS measures the sum of the squared distances between each data point and its assigned cluster centroid, and the algorithm aims to minimize this value by iteratively updating the centroids.

Because the algorithm always follows the same steps and uses the same objective function, it will always produce the same result for a given data set and value of K. However, it is worth noting that the algorithm is sensitive to the initial placement of the centroids, which can affect the final clustering result.

Therefore, it is common practice to run the algorithm multiple times with different initial centroids and choose the result with the lowest WCSS value.

