Course-6 Assignment-2

Assignment: Application and Challenges of K-Means Clustering

Part-1 Real-World Applications of K-Means

Task 1: Select a Real-World Scenario

1. Choose one real-world application where K-Means clustering can be used (e.g., customer segmentation, image compression, anomaly detection, or market segmentation).
2. Provide an explanation of how K-Means clustering works in this scenario and why it is useful.

Ans 1:

Topic: Customer Segmentation

Business often deal with large amount of data from its large and diverse customer base.

K-Means is used to divide customers into meaningful segments.

Everyone has different purchasing behaviour, preference.

here K-means can be applied for segmenting the customers into different segments based on different variables like, age, gender, frequency of purchases, categories/types of product bought.

Ans 2:

K-means is a part of unsupervised learning algorithm, it is used to divide given input into similar features clusters, where features sharing similar characteristics mapped closer together while different features have their own separate mapped cluster.

In K-Means, K is also representing the cluster numbers. Every cluster has a centroid, which is centre point of its cluster.

Every customer is represented as data point in multi-dimensional space.

K-Means identifies the group of similar behaviours.

From data we can gather information on high-value customers like frequent buyers, type of product sold in region, most bought product, etc. which can enable personalized recommendations based on customer interest or need.

Task 2: Benefits of Using K-Means

- Discuss two main benefits of using K-Means in your chosen scenario. For example, how it improves decision-making, reduces complexity, or enhances predictions.

Ans :

Efficient in large volume:

It very efficient in case of large number of customers. K-Means is computationally efficient can handle large data. Making is suitable for on the go recommendation based on customer purchase history.

Improved Decision Making:

Businesses can get insights of customer behavior which otherwise remain hidden in plain sight. By understanding different segments, companies can make informed decision, recommend and make better products.

Part 2: Challenges and Alternatives

Task 1: Limitations of K-Means Clustering
- List and explain two limitations of K-Means clustering (e.g., sensitivity to initial centroids, difficulty handling non-spherical clusters, or issues with clusters of varying sizes).

Ans:

Dependence on the number of clusters(K): It require, the value of K in advance. Choosing the wrong K leads to inaccurate clusters, which give meaningless information.

Sensitivity to outliers: outliers can pull cluster away from the required true centroid position, which is real centre of cluster. So outlier can affect the centroid calculation and distort resulting clusters.

Task 2: When Not to Use K-Means
- Describe a situation where K-Means clustering is not the best choice and explain why. Suggest a more suitable algorithm for that scenario.

Ans :

When there are significant outliers in the data:

If the data contain some customers with very high purchase value orders then normal like 4 times, 10 times or 20 times higher volume of purchase.

Outliers are important too but can distort the result.

K-means are not robust to outliers.

It can easily skew the centroid positions, and result in sub optimal model if treated carefully.

more suitable algorithm:

techniques like outlier detection and removal.

Or using algorithms that are robust to outliers can perform better like GMM(Gaussian Mixture Models)