Experiment 10: Customer Segmentation Using k-Means Clustering

---

Aim:

To apply the **k-Means clustering algorithm** on the Customer Segmentation dataset to group customers based on **similar spending behavior**.

Theory:

- **k-Means** is an **unsupervised learning algorithm** used for clustering.
- It partitions data into **k clusters** based on similarity (e.g., distance metric like Euclidean).
- Each cluster has a **centroid** representing the cluster's mean.
- Useful for **market segmentation and customer analysis**.

Dataset (customer_segmentation.arff)

@relation customer_segmentation

@attribute Age numeric

@attribute AnnualIncome numeric

@attribute SpendingScore numeric

@data

25,50000,70

30,60000,60

22,35000,40

28,58000,80

35,45000,50

40,62000,30

32,52000,65

26,48000,75

...

Procedure (Using WEKA):

1. Open **WEKA → Explorer**.
2. Click **Open File** → select **customer_segmentation.arff**.
3. Go to **Cluster tab**.
4. Choose **Clusterer → SimpleKMeans**.
5. Set **number of clusters (k = 3)**.
6. Click **Start** → WEKA performs clustering.
7. Observe **cluster centroids and instance assignments**.
8. Visualize clusters using **Visualize panel**.

Result (Sample / Expected):

- Customers grouped into **3 clusters**:
    - Cluster 0: Young, high spending
    - Cluster 1: Middle-aged, medium spending
    - Cluster 2: Older, low spending
- Centroids show **average Age, AnnualIncome, SpendingScore** per cluster.

Conclusion:

- k-Means successfully grouped customers with **similar spending behavior**.
- Clustering helps in **market segmentation and targeted marketing**.
- WEKA provides an **easy interface for clustering and visualization**.