**Day 12 Training Report**

**8 July 2025**

**k-Means Clustering — Working, Implementation, Hands-on Practice**

On **Day 12**, students learned **k-Means clustering**, one of the most popular unsupervised learning algorithms used to **group data points into clusters**.

---

**1. Introduction to k-Means**

- **Objective:** Partition n data points into k clusters where each point belongs to the cluster with the **nearest centroid**.
- **Centroid:** Mean of all points in a cluster.
- **Distance Metric:** Typically **Euclidean distance**.

---

**2. Algorithm Steps**

1. Choose the number of clusters k.
2. Randomly initialize k centroids.
3. Assign each data point to the nearest centroid.
4. Recalculate centroids as the mean of assigned points.
5. Repeat steps 3–4 until centroids stabilize (no significant change).

---

**3. Hands-on Implementation**

```
from sklearn.cluster import KMeans
import pandas as pd

# Sample dataset
data = pd.DataFrame({'X':[1,2,3,8,9,10], 'Y':[2,3,2,8,9,10]})

# Apply k-means
kmeans = KMeans(n_clusters=2, random_state=42)
kmeans.fit(data)

data['Cluster'] = kmeans.labels_
print(data)
```

**Outcome:** Data points are grouped into clusters automatically.