### **ML ASSIGNMENT 4: K-MEANS CLUSTERING**

**ID: ASGN2606**

**K-Means Clustering on Health and Fitness Dataset**

#### **1) Data Preprocessing**

* Dropped unnecessary column: ID
* Filled missing values in numerical columns (Height, Weight, Exercise Hours) using mean imputation
* Encoded categorical columns (Gender, Chronic Conditions) using One-Hot Encoding
* Scaled all numerical features using StandardScaler to normalize the data for clustering

#### **2) Elbow Method**

* The Elbow Method was applied by plotting the inertia (within-cluster sum of squares) for different values of k
* The curve showed a clear elbow at k = 4, which was selected as the optimal number of clusters

#### **3) K-Means Clustering**

* K-Means algorithm was applied with n\_clusters = 4
* Cluster labels were added to the preprocessed dataset
* PCA was used to reduce dimensions for visualization of clusters in 2D

#### **4) Cluster Interpretation**

* Cluster 0  
   Active individuals with healthy BMI and good fitness levels  
   Diet and activity scores are average to above-average  
   Represent generally fit, moderate lifestyle group
* Cluster 1  
   Least physically active with low exercise and high BMI  
   Most members have chronic health issues  
   Sedentary lifestyle, higher health risk group
* Cluster 2  
   Most active group with highest step count and exercise hours  
   High diet scores and no chronic conditions  
   Health-conscious, energetic individuals
* Cluster 3  
   Moderate activity, some chronic conditions, and average diet  
   Transitional group with mixed lifestyle traits

#### **5) Visualization**

* PCA was used to reduce features to 2 dimensions
* Clusters were visualized using matplotlib scatter plots
* Each cluster was clearly distinguishable in the 2D plot

#### **Conclusion**

* K-Means successfully grouped individuals into 4 unique health profiles
* The results can be used to understand fitness behaviors and target wellness programs
* Clustering helped reveal hidden patterns in lifestyle habits based on numerical and categorical data