Week 2 – Task 3: Customer Segmentation using K-Means

This report summarizes Week 2 Task 3 of the Edulumos Machine Learning Internship. The goal of this task was to perform customer segmentation using K-Means clustering on a mall customers dataset and interpret the resulting customer groups.

Dataset Overview

The dataset consists of 200 customers with features such as Gender, Age, Annual Income, and Spending Score. Each row represents a unique customer with behavioral and demographic details. No missing values were found in the dataset.

Preprocessing Steps

- Dropped the 'CustomerID' column as it does not contribute to clustering. - Encoded the 'Gender' feature numerically (Male=0, Female=1). - Applied StandardScaler to normalize numerical features so that all variables contribute equally to the distance-based K-Means clustering algorithm.

K-Means Clustering and Elbow Method

K-Means clustering was applied to the scaled features. To determine an appropriate number of clusters (k), the Elbow Method was used by plotting Within-Cluster Sum of Squares (WCSS) for k values ranging from 2 to 10. The 'elbow' in the curve suggested $k \approx 5$ as a good trade-off between compact clusters and model simplicity.

Cluster Interpretation

The final K-Means model was trained with k = 5 clusters. Each customer was assigned a cluster label. By analyzing the average Age, Annual Income, and Spending Score per cluster, the following insights can be drawn (example interpretation):

Cluster	Description (Example)
0	Low income, low spending – budget-conscious customers
1	High income, high spending – premium/VIP customers
2	High income, low spending – potential customers with untapped value
3	Young, high spending – brand and trend-focused customers
4	Moderate income & spending – regular average customers

Conclusion

This task demonstrates the power of unsupervised learning in discovering hidden patterns in customer behavior. K-Means clustering helped segment mall customers into meaningful groups based on their income and spending patterns. These segments can be used by businesses for targeted marketing, personalized offers, and better customer relationship strategies.