

Name: Jiya Sharma

Roll No: 281056

PRN: 22311741

Assignment 5

Problem Statement:

This dataset gives the data on Income and money spent by the customers visiting a shopping mall. The dataset contains Customer ID, Gender, Age, Annual Income, and Spending Score. Therefore, as a mall owner, you need to find a group of people who are profitable customers for the mall. Apply at least two clustering algorithms (based on Spending Score) to find the group of customers.

Objectives:

- To apply unsupervised machine learning algorithms for customer segmentation.
- To analyze the dataset and preprocess the data for clustering.
- To apply two different clustering algorithms.
- To visualize and interpret the resulting clusters.

Resources used:

- Software used: Visual Studio Code
- Libraries used: Pandas, Matplotlib, Seaborn, Scikit-learn

Theory:

Clustering is an unsupervised machine learning technique where the objective is to group a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups.

In this assignment, we focus on customer segmentation using the Spending Score. Two popular clustering algorithms are:

- **K-Means Clustering:**

K-Means partitions the data into K clusters where each point belongs to the cluster with the nearest mean. It minimizes the within-cluster variance.

- **Hierarchical Clustering:**

Hierarchical clustering builds a hierarchy of clusters either using a bottom-up (agglomerative) or top-down (divisive) approach. It does not require specifying the number of clusters beforehand.

Methodology:

1. Data Preprocessing:

- Load the dataset using Pandas.
- Handle missing values (if any).
- Select relevant features (Spending Score) for clustering.
- Normalize the data if required.

2. Applying Clustering Algorithms:

- K-Means Clustering:
- Use the Elbow Method to determine the optimal number of clusters.
- Apply K-Means to the dataset.
- Hierarchical Clustering:
- Create a dendrogram to visualize possible cluster separations.
- Apply Agglomerative Clustering based on dendrogram analysis.

3. Visualization and Interpretation:

- Visualize the clusters formed by each algorithm using scatter plots.
- Analyze the nature of different clusters to identify profitable customer groups.

Conclusion:

- Using K-Means and Hierarchical Clustering, distinct groups of customers were identified based on their Spending Scores.
- Profitable customer groups (high income and high spending) were successfully separated.
- Proper customer segmentation enables targeted marketing strategies and improved business decisions.

