

Name: Snehal Laxmikant Yelwande

Roll No: 281063

Batch: A3

Assignment 5

Problem Statement:

The goal of this project is to segment customers visiting a shopping mall based on their spending patterns. The dataset consists of fields like Customer ID, Gender, Age, Annual Income, and Spending Score. As the mall owner, we aim to identify clusters of customers who contribute significantly to the business and could be potential targets for marketing campaigns. To achieve this, we will apply unsupervised machine learning techniques to find meaningful customer segments based on their Spending Score.

Objectives:

1. To understand and implement unsupervised learning techniques for customer grouping.
2. To clean and prepare the dataset for clustering.
3. To use two clustering approaches: K-Means and Hierarchical Clustering.
4. To visualize the results and gain insights into customer behaviour patterns.

Resources used:

1. **Platform Used:** Visual Studio Code
2. **Python Libraries:** Pandas, Seaborn, Matplotlib, Scikit-learn

Theory:

Clustering is a type of unsupervised learning where we divide a dataset into groups, such that data points in the same group are more similar to each other than to those in other groups. In this project, we will use two commonly used clustering methods:

- **K-Means Clustering:**

K-Means clustering divides the data into K distinct clusters. It works by assigning data points to the nearest centroid and minimizing intra-cluster variance. The ideal number of clusters is usually found using the Elbow Method.

- **Hierarchical Clustering:**

Hierarchical clustering builds nested clusters in a tree-like structure (dendrogram). It can be either agglomerative (bottom-up) or divisive (top-down). It doesn't require specifying the number of clusters in advance and gives a visual representation through the dendrogram.

Methodology:

1. Data Preparation

- Load the dataset using Pandas.
- Check and handle missing or inconsistent data.
- Select the "Spending Score" and other relevant features if necessary.
- Normalize the data (if the scale of features varies significantly).

2) Applying Clustering Algorithms

a. K-Means Clustering

- Use the Elbow Method to choose the appropriate number of clusters (K).
- Apply the K-Means algorithm to form clusters based on Spending Score.
- Assign cluster labels to each data point.

b. Hierarchical Clustering

- Generate a dendrogram using the linkage matrix.
- Determine the optimal number of clusters by analyzing the dendrogram.
- Apply Agglomerative Clustering and assign cluster labels.

3) Visualization and Interpretation

- Use scatter plots to display clusters formed by both algorithms.
- Colour-code points by cluster and compare patterns visually.
- Discuss the characteristics of each cluster – for example, high spenders, low-income groups, or moderate customers.

- Identify which group represents the most profitable customers (typically high income + high spending)

Conclusion:

- The use of K-Means and Hierarchical Clustering effectively segmented the mall customers.
- Profitable segments were clearly identifiable, which can be used for focused marketing strategies.
- With better customer targeting, the mall can optimize its promotions, improve customer satisfaction, and enhance profitability.