Assignment 02: Clustering

Let's imagine you're owning a supermarket mall and through membership cards, you have some basic data about your customers like **Customer ID, age, gender, annual income and spending score,** which is something you assign to the customer based on your defined parameters like customer behavior and purchasing data.

The main aim of this problem is learning the purpose of the customer segmentation concepts, also known as market basket analysis, trying to understand customers and separate them in different groups according to their preferences, and once the division is done, this information can be given to marketing team so they can plan the strategy accordingly.

This **Mall_Customer** dataset that has been provided to you is composed by the following five features:

- ***CustomerID***: Unique ID assigned to the customer
- ***Gender***: Gender of the customer
- ***Age***: Age of the customer
- ***Annual Income (k$)***: Annual Income of the customer
- ***Spending Score (1-100)***: Score assigned by the mall based on customer behavior and spending nature.

In this particular dataset we have **200** samples to study.

PART (A) [10 points]

K-means Clustering: In this part, you will be utilizing K-means clustering algorithm to identify the appropriate number of clusters. You may use any language and libraries to implement K-mean clustering algorithm. Your K-mean clustering algorithm should look for appropriate values of K at least in the range of 0 to 15 and show their corresponding sum-of-squared errors (SSE).

PART (B) [10 points]

Hierarchical Clustering: In this part, you will apply hierarchical clustering algorithm (agglomerative or divisive) to the provided mall dataset.

PART (C) [10 points]

Density-based Clustering: In this part, you will apply density-based clustering algorithm to the provided dataset.

PART (D)

Submission: You will need to submit two files:

(a) Programming code files, and

(b) A report file consisting **a detail explanation of your approaches and the results obtained for all of the clustering algorithms required in Parts A, B, and C.**

Put all the files in a folder and your compressed file should look like **DS_Task2_YourID.zip**. You have to upload both files in your **personal GitHub profile** and provide a link of the file in your report.

Deadline: 14th October, 2023.

Submission Type: Individual.