

Clustering Assignment

Problem Statement:

You are a data scientist working for a marketing research firm. Your company has collected data on customer purchasing behavior from an online retail store. The goal is to segment customers based on their purchasing habits to better understand their preferences and tailor marketing strategies accordingly. Your task is to perform clustering analysis on the dataset provided and interpret the results to derive actionable insights for the marketing team.

Tasks:

1. Data Exploration:
 - Load the dataset and perform preliminary data exploration.
 - Examine the structure of the dataset, including the number of features and instances.
 - Check for missing values and handle them appropriately.
 - Visualize the distribution of key variables to gain insights into the data.
2. Data Preprocessing:
 - Normalize or scale the features if necessary.
 - Encode categorical variables if applicable.
 - Split the dataset into training and testing sets (if required).
3. Clustering Analysis:
 - Apply K-means clustering algorithm to the preprocessed dataset.
 - Determine the optimal number of clusters using methods like the Elbow Method.
 - Visualize the clusters in a 2D.
4. Evaluation:
 - Evaluate the quality of the clustering solution using appropriate metrics (e.g., silhouette score, Davies–Bouldin index).
 - Discuss the strengths and limitations of the chosen clustering algorithm.
5. Insights and Recommendations:
 - Provide actionable insights based on the clustering results.
 - Suggest marketing strategies tailored to each customer segment.
 - Discuss how the company can use these insights to improve customer engagement and increase sales.

Dataset:

You are provided with a dataset ([customers.csv](#)) containing the following features:

- Customer ID
- Age
- Gender
- Annual Income
- Spending Score

Submission Guidelines:

- Submit a report documenting your analysis, including code snippets, visualizations, and interpretations.
- Submit your report and code files by jupyter notebook files