Machine Learning Course

Assignment 3

Mall Customer Segmentation Data

Introduction

Exploring customer behavior and preferences is crucial for businesses to tailor their marketing strategies and enhance customer satisfaction. In this exercise, we delve into the Mall Customer Segmentation dataset, a valuable resource containing demographic and spending information of mall customers. By applying clustering techniques to this dataset, we aim to unveil distinct customer segments based on their purchasing patterns.

Understanding customer segmentation enables businesses to personalize their marketing campaigns, optimize product offerings, and improve the overall customer experience. Through this analysis, we endeavor to extract actionable insights that can drive strategic decision-making for mall management.

Objective

1. Utilize clustering techniques to segment customers based on their demographic attributes and spending behavior.
2. Explore the Mall Customer Segmentation dataset to gain insights into customer preferences and characteristics.
3. Apply data preprocessing techniques to ensure the quality and compatibility of the dataset for clustering analysis.
4. Implement and compare different clustering algorithms to identify the most suitable approach for customer segmentation.
5. Interpret the resulting clusters to derive meaningful insights regarding customer segments and their respective marketing implications.
6. Provide recommendations for mall management based on the clustering analysis to optimize marketing strategies and enhance customer satisfaction.

Dataset Description

The dataset consists of the following features:

- **CustomerID**: Unique identifier for each customer
- **Gender**: Gender of the customer (Male/Female)
- **Age**: Age of the customer
- **Annual Income (k$)**: Annual income of the customer in thousands of dollars
- **Spending Score (1-100)**: Score assigned to the customer based on their purchasing behavior and past interactions

Tasks

1. **Data Exploration**
   a. Load the dataset and perform preliminary data exploration.
   b. Investigate the statistical summary of the dataset.
   c. Check for missing values and handle them appropriately if necessary.
   d. Visualize the distribution of features using appropriate plots (e.g., histograms, box plots).

2. **Feature Selection**
   a. Decide which features to include for clustering (e.g., age, annual income, spending score).
   b. Justify your selection based on the characteristics of the dataset and the objectives of customer segmentation.

3. **Data Preprocessing**
   a. Normalize the selected features to ensure they are on the same scale.
   b. Explain the importance of normalization in the context of clustering.

4. **Clustering Algorithms**
   a. Choose at least two clustering algorithms (e.g., K-means, Hierarchical clustering) to segment the customers.
   b. Implement the chosen algorithms using appropriate libraries (e.g., scikit-learn in Python).
   c. Tune the hyperparameters of the algorithms if necessary.
   d. Explain the intuition behind each clustering algorithm and how it works.

5. **Clustering Analysis**
   a. Apply the clustering algorithms to the preprocessed data.
   b. Visualize the clusters using scatter plots or other suitable visualization techniques.

c. Analyze and interpret the resulting clusters based on their characteristics.
d. Discuss the potential marketing strategies for each customer segment.

6. **Evaluation**
    a. Evaluate the quality of the clusters using internal validation metrics (e.g., silhouette score).
    b. Compare the performance of different clustering algorithms.
    c. Provide insights into the strengths and limitations of each algorithm in this context.

7. **Conclusion**
    a. Summarize the key findings of your analysis.
    b. Reflect on the significance of customer segmentation for business decision-making.
    c. Propose recommendations for mall management based on the clustering results.

Dataset

You can download the dataset from [this link](#)

References

● [Sklearn Clustering Documentation](#)