

Homework 4

Introduction

In the evolving landscape of digital media, understanding customer behavior and preferences is vital. This report presents an analysis of customer segmentation for a digital magazine company, utilizing two datasets: Behavioral Data and Article Data. The former includes metrics like browsing time and income, while the latter comprises counts of articles read across various topics. Effective segmentation can lead to personalized customer experiences and increased engagement, ultimately enhancing the company's success.

Methods

The analysis employs clustering models to segment customers based on their behavior and reading preferences. Appropriate pre-processing was applied to each dataset to ensure optimal clustering results.

Behavioral Clustering Model

Pros and Cons

K-Means Clustering, known for its efficiency with large datasets, was chosen for the Behavioral Data. Its major advantage lies in its simplicity and speed, especially suitable for continuous data. However, it assumes cluster symmetry and struggles with varying cluster sizes. The choice was driven by the data's nature, primarily continuous variables.

Chosen Model Details

The K-Means model required standardizing the features using Z-score normalization. The number of clusters was chosen based on the Elbow Method, ensuring the most distinct segmentation.

Article Clustering Model

Hierarchical Clustering with average linkage and cosine similarity was employed for Article Data. This model excels in revealing the hierarchical structure of data and is particularly effective with non-uniform cluster sizes. Cosine similarity was chosen due to the count nature of the data, focusing on the pattern of reading rather than the volume.

Results

Behavioral Clustering Model

The K-Means model effectively segmented customers into distinct groups. Each cluster represented unique behavioral patterns, potentially guiding targeted marketing strategies.

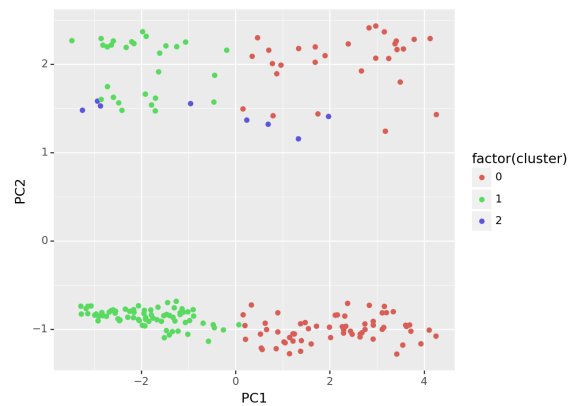


Figure 1: PCA Plot of Customer Segments

Article Clustering Model

Hierarchical Clustering revealed distinct content preferences among customers. This segmentation can inform content curation and recommendation algorithms.

Discussion/Reflection

This analysis highlighted the diversity in customer interactions and preferences. A key learning was the importance of choosing the right clustering approach for different types of data.

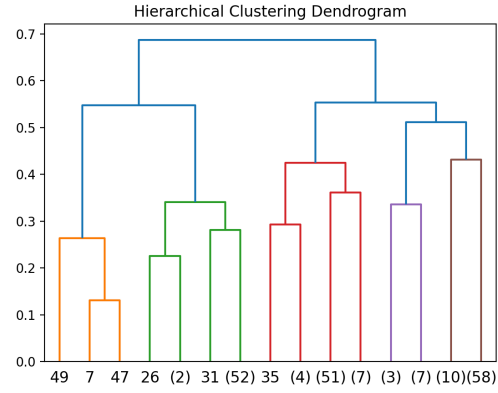


Figure 2: Dendrogram of Article Clusters

In future analyses, integrating both datasets could provide a more comprehensive view of customer behavior. Additionally, exploring dynamic clustering techniques could offer real-time insights into evolving customer preferences.

Pelin Nisa Top