Question 1: Paper Review

Title: K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data

Authors: Kayalvily Tabianan, Shubashini Velu and Vinayakumar Ravi

Motivation: The motivation of this study is to improve business performance / increase profits of E-commerce systems by optimizing the services and products offered to the consumers. The team will accomplish this goal through an increased amount of offerings to the "right" customers which will in turn improve business performance.

Summary: The COVID-19 pandemic has led to a culture where working from home is normal. Businesses are always adapting and this situation is no different. In the past two years many companies have needed to open online stores if they want to stay competitive. While brick and mortar are still able to use unsupervised clustering algorithms such as K-means, switching to ecommerce enables companies to gather a much greater amount of data. This in turn results in a much better understanding of the customer and in turn a much better result from the K-means algorithm. Performing this analysis will enable the business to find the most profitable customer segments.

The team's research focused on customer segmentation in the Malaysian ecommerce industry. They aimed to identify highly profitable customers through segmentation using K-means clustering. The questionnaire target audience is users aged between 18 and 60.

For the analysis, the team used a combination of affinity propagation and the K-means clustering algorithm. This is known as SAPK / AP + K-means.

Looking at figures 5 to 13, we can see an analysis of data on product brands, event types, product categories, and customer interest in products. Figures 5, 6, and 7 show frequency of product brands, event types, and product categories respectively. Figure 8 shows appliances and electronics are the most viewed categories. Figure 9 shows the K-means algorithm applied to the data. The illustration of clustering analysis in figure 13 shows three clusters with different levels of customer interest.

After analysis with K-means clustering and mean shift clustering, K-means was chosen as the most suitable. The results of the analysis showed 3 profitable segments based on event type and products / categories.

Approach and contributions:

- The authors used a data analysis approach to establish the results by evaluating algorithms such as hierarchical clustering, K-Means clustering, and mean shift clustering for customer segmentation.
- The main findings of this paper are that K-means is better than previously used techniques when it comes to customer segmentation.
- The important of this work is showing the improved clustering that K-means can provide for customer segmentation.
- The paper builds on the work of the K-means algorithm. It shows how K-means can be applied to customer segmentation.

Areas for improvement: One area of improvement that jumps out is the use of so few algorithms. In the future, the researchers can test more algorithms to see if there is a better option compared to K-means or the other few tested in this paper.