RFM model for customer purchase behavior using K-Means algorithm

P. Anitha, Malini M. Patil, Abhishek Puppala

October 23, 2022

1 The main idea

The motive of this paper is identifying the different potential customers by providing relevant data in Retail industry by applying business intelligence. It is analyzed based on predicting consumer purchasing behaviour. This study is based on RFM model(i.e Recency, Frequency, Monetary) and deploys the segmentation principles with help of k-means clustering algorithm.

2 The methodology

Here, they used the online-retail dataset which can be found at https://archive.ics.uci.edu/ml/datasets/Online+Retail. It is a UK based online retail company with total 8492 instances. Here the attributes used are Stock Code, Description, Quantity, Unit Price, Customer Id.

After the data is preprocessed, it checks for recent transaction frequency, which creates recency variable. Then partitioning the customers into clusters, using euclidean distance with help of k means clustering(applied twice for analyzing amount obtained for recent and frequent transactions). Now, Silhouette Score(A metric used to calculate the goodness of a clustering technique) is calculated and evaluating the clusters, by comparing the K-values.

3 The results

It is observed that, the results of silhouette matrix for k=3 is more accurate than k=5. The value of silhouette nearer to +1 represents optimal one comparatively to other clusters. Based on the Silhouette Score, the Sales Recency, Sales Frequency and Sales Monetary can be analyzed and an optimal solution is found.

4 Recommendation

The future work can be done through examination of various items, that includes laptop, projectors. For the purpose of getting efficient business profit, can be researched on threshold parameters in various locations. These type of developments and discussions in this field will assist the firms in advancing their operations by developing different strategies.