Abstract

This research is built around the application of Machine Learning technologies to the supermarket retail sector in Sri Lanka. Two areas identified for the study: Customer Segmentation for the application of Unsupervised Clustering algorithms and Market Basket Analysis for the application of Association Rule Mining.

The main aim of the research was to identify the different clusters of customers found within the supermarket retail domain of Sri Lanka. To facilitate this, it first required the collection and analysis of the POS (Point of Sale) sale data in combination with the customer information. The access to this information was provided by the Keells Supermarket chain, owned, and maintained by JayKay Marketing Pvt. Ltd which is a Part of the John Keells Group of Companies. They provided limited access to the relevant information as well as the Nexus Customer Loyalty Program which contained most of their customer data. The data was analyzed both in its entirety and in various derivative forms that yielded diverse results.

In the clustering process several clustering algorithms were applied, K-Means, K-Modes, K-Prototypes, DBSCAN and Mean Shift algorithms were some of the successfully tested algorithms. They provided diverse outcomes, some with very clear clusters while other clusters without any coherent meaning. There were also instances where an algorithm would not be able to deliver a clear and coherent outcome with the main dataset but would give a viable result for one of the derived datasets.

The Association Rule Mining (ARM) process considered the Apriori and Frequent Pattern Growth (FP Growth) algorithms which are two of the most popular ARM algorithms in use today. The outcomes of these algorithms were able to clearly provide association rules between products which were consistent through tests on different samples of data.

Based on the finding it was successfully concluded that it is indeed possible to apply Clustering to the retail industry in a customer segmentation capacity, albeit the viability of the outcomes may differ based on the requirement and mode of application.

Great potential can be found in the application of the findings of both Clustering and ARM in customer attraction and retention. It opens a new frontier for building customer value.

Identify customer segments in the supermarket retail sector of Sri Lanka using Machine Learning techniques

By

Dexter Martel

Robert Gordon University

Module Code :   **CMM799**

RGU ID   :   **1915376**

IIT ID   :   **20200058**