BABES BOLYAI UNIVERSITY, CLUJ NAPOCA, ROMÂNIA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Customer Segmentation

Rus Alina Petruta, 258-2 – Report – Front End – alina.rus999@gmail.com

1. Introduction
   1.1. What? Why? How?

As there are hundreds of thousands of online stores, hundreds targeting the same customers and trying to compete with their competitors, the world of e-commerce is a challenging one. Having loyal customers is one of the best ways to do this, and what is a better way to keep a consumer purchasing goods from you than having tailored offers that are exactly what the individual wants?

But how do businesses understand what advertisements/recommendations to give to their clients? This method of getting to know the consumer is called customer segmentation.

For a stronger recommendation of different items, the project is used to segment clients into similar categories.

The scientific problem was to find a good algorithm for users to be separated into classes. This can be achieved through a variety of algorithms, but K Means Clustering is currently the best fit for this after researching some eligible articles.

   1.2. Paper structure

The study discussed in this paper advances the application of K Means Clustering to the issue of consumer segmentation. A real-world dataset was used to h validate the findings using different parameters, as this is a real-world challenge.

The algorithm was implemented into a small application to suggest to different users the right kind of items.

A brief introduction to the issue at hand and a slight idea of how to get through has been presented in the first chapter.

The second chapter explains in more depth the question and a scientific approach that is potentially feasible.

The third section outlines the latest state of the art in the field of e-commerce businesses.

The fourth section explains the algorithm and the various criteria we used to execute it.

The fifth chapter addresses the program, methods and dataset used, and the effects of using it on the UCI dataset described above.

The sixth and final chapter shows the end of the experiment and the scope for future work.

2. Scientific Problem
   2.1. Problem definition

Just as mentioned before, the best way a retail shop manages to keep and expand their clients base is by ‘buying’ their loyalty. How to get loyal customers? By assuring them that you have their needed items at a price they can afford. Now the problem is, finding out what kind of products your client base is looking for. This can be accomplished by observing their purchasing behavior and comparing it to other clients. After that, based on common interests you can group your clients into classes and make appropriate suggestions to each class.

Since solving this problem would take an enormous amount of time and effort on the human part, this can be a great job for a computing machine. The only thing remaining is, teaching a machine how to do this classification.

K Means is an algorithm that aims to partition n observations into k clusters in which observation belongs to the cluster with the nearest mean. Using this algorithm requires minimum assistance from the programmer side, as he can just analyze the results and determine how accurate are they and then adjust the parameters to obtain

more appropriate results. The downside of this method is that with a poorly chosen starting dataset the results may not be accurate even after adjusting the parameters.

In this application we intend to use the K Means Clustering, which is an unsupervised learning algorithm. Using this method, the data doesn't need to be trained, it will just find groups based on similar entities. A small amount of attention it is indeed required so that the proper number of clusters is determined. After that, most of the work is done by already dedicated tools that implement most of the hard calculations.

An advantage of using this method is that the data doesn't need to be trained and that the starting dataset is big enough so that it gives accurate results.

3. State of the art/ Related work

Customer segmentation on e-commerce

The paper studied presents the importance of customer segmentation in keeping, gaining and satisfying the customers in order to obtain more revenue. In addition, multiple variables are addressed to execute various consumer segmentation strategies, such as: commodity, purchase, regional, hobbies and variables viewed on the list. Moreover, the approaches used for these variables are illustrated as follows: Market Law, Quantile Membership, Supervised Clustering, Unsupervised Clustering, Consumer Profiling, Grouping of RFM Cell Classification, Customer Likeness Clustering and Clustering of Purchasing Affinity.

Customer Segmentation Intelligence is brought into light by pointing out the improvements added to the market in offering products or services that meet the need of each group of customers. Moreover, the process of classifying an item into a group with similar characteristics is used to determine the similarities between the customers by segmenting the records from the customer database. In the process of segmenting customers there were categorized 2 type of data, internal, which are customer registration, profile, purchase history and external are media browsing, surveys and market search, cookies, web and social media analysis.

4. Investigated approach

The K-Means clustering beams at partitioning the 'n' number of observations into a mentioned number of 'k' clusters (produces sphere-like clusters). The K-Means is an unsupervised learning algorithm and one of the simplest algorithms used for clustering tasks. The K-Means divides the data into non-overlapping subsets without any cluster-internal structure. The values which are within a cluster are very similar to each other but, the values across different clusters vary enormously. K-Means clustering works really well with medium and large-sized data.

Imagine that you have a customer dataset, and you need to apply customer segmentation on this historical data. Customer segmentation is the practice of partitioning a customer base into groups of individuals that have similar characteristics. It is a significant strategy as a business can target these specific groups of customers and effectively allocate marketing resources. For example, one group might contain customers who are high-profit and low-risk, that is, more likely to purchase products, or subscribe for a service. A business task is to retain those customers. Another group might include customers from non-profit organizations.

5. Application
6. Conclusion