Customer Segmentation In E-Commerce Using Rfm Analysis And Machine Learning

Mohammad Uzair Chanda[1], Shreyas Shrotri[2], Swapnil Bangar[3]
[1, 2, 3] Dept of Computer Department
[1, 2] V.I.I.T, Pune.

Abstract- *E-commerce companies need to understand and interpret their customer's data in all perspective detecting correlation and similarity among customer, foreseeing their behavior recommending better choices and opportunities to customer becomes very important for customer relationship management. RFM analysis is an efficient method to identify or to cluster customers based in their interaction with the company. RFM values have been widely used in order to identify the potential or valuable customer to the company and which customers need promotional activities.*

Keywords- RFM, Customer segmentation, e-commerce, k-means, data mining, CRM.

I. INTRODUCTION

In today's world where a vast amount of data is created and collected each day, analyzing such data is an important aspect. Data mining techniques can be implemented on existing software and hardware platform to up skill the value present information resources and can be integrated with new products and systems as they are bought online. The market should be able to understand what the customer wants to buy according to the certain product that is not seen in the market. Data mining techniques are known to have great potential to assist organizations to take important decision in order to increase the business. The relation between the customer and the companies has become an undeniable aspect of the business and thus the continuation of a structure to manage this relation is mandatory. The controlling process of the interaction between the organizations and the customer is called Customer Relationship Management (CRM). The CRM is further divided into four categories:(1) Customer Identification, (2) Customer Attraction, (3) Customer Retention, (4) Customer development. The RFM analysis is known as the behavioral based data mining technique which extracts customer profile by using their Recency, Frequency and Monetary values. These RFM matrices are important indicators of a customer's behavior because frequency and monetary impact the customer's lifetime value, and recency affects retention.

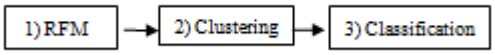
RFM matrices:-

1. Recency:- Recency stands for the freshness of the customer activity be it purchases or visits.
2. Frequency:- The frequency of the customer transaction or visits.
3. Monetary:- The intention of the customer to spend or the purchasing power of the customer.

RFM matrices help us to understand the following factors

a. The more recent purchases, the more acceptance of the customer.
b. The more frequently the customer buys, the more engaged and satisfied they are.
c. Monetary value differentiates heavy spenders from low value purchasers.

II. IMPLEMENTATION



There are three basic steps to implement RFM analysis:

1. Sort all customers in ascending order based on Recency, Frequency and Monetary Value.
2. Split customers into quartiles for each factor.
3. Combine factors to group customers into RFM segments for targeted marketing.

The first goal is to identify and sort customers. To do this, you need to download a dataset or a .csv file with customer purchase history. The file should include the date of the most recent order, the number of orders placed over your selected time period, the total value of all purchases made in that time period, and customer ID.

Now your dataset or.csv is ready to be segmented. Resolve each column in ascending order based on its RFM factor. For example, select the column for Recency, and sort so that the most recent orders are first, and the oldest orders are last.

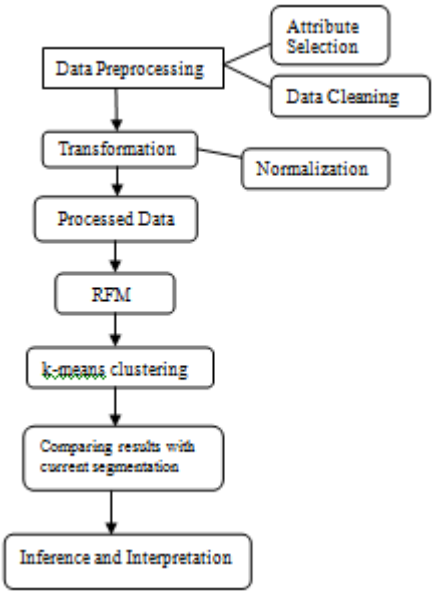
The next step is to score your customers. For this purpose we use Basic Ranking, as it will serve the needs of most online e-commerce companies. With customers now organized in ascending order, divide them into quartiles, or four equal groups, for each RFM factor. The customers in the top quartile represent your best customers for each factor. For example, the top quartile for Monetary Value will have the 25% of your customers who have spent the most at your store. Each quartile also has a name: the top quartile for Recency is called R-1, the second quartile is called R-2, and so on.

Dividing into quartiles will create 64 RFM segments: 4 Recency groups x 4 Frequency groups x 4 Monetary Value groups.

A. RFM Segments

Name	Segment
Best Customer	R1,F1,M1
Loyal Customer	F=1
Big Spenders	M=1
New Spenders	141
Lost Customers	411
Deadbeats	444

B. RFM Model Flow Chart



III. K-MEANS

K-means is the simplest and most widely used clustering algorithm. This algorithm uses as input a predefined number of clusters that is the k from its name. Mean stands for an average, an average location of all the members of a particular cluster. When dealing with clustering techniques, a notion of a high dimensional space must be adopted, or space in which orthogonal dimensions are all attributes from the table of analysed data. The value of each attribute of an example represents a distance of the example from the origin along the attribute axes. Of course, in order to use this geometry efficiently, the values in the data set must all be numeric and should be normalized in order to allow fair computation of the overall distances in a multi-attribute space. K-means algorithm is a simple, iterative procedure, in which a crucial concept is the one of centroid. Centroid is an artificial point in the space of records that represents an average location of the particular cluster.

The steps of the K-means algorithm are given as follows

1. Select randomly k points to be the seeds for the centroids of k clusters.
2. Assign each example to the centroid closest to the example, forming in this way k exclusive clusters of examples.
3. Calculate new centroids of the clusters. For that purpose average all attribute values of the examples belonging to the same cluster (centroid).
4. Check if the cluster centroids have changed their "coordinates". If yes, start again form the step 2). If not, cluster detection is finished and all examples have their cluster memberships defined.

IV. CONCLUSION

Organizations should develop better understanding of their components. Especially, it is important for businesses, they should have detailed understanding about their customers' characteristics, behaviors, demographics, etc. In this context, techniques have been developed. RFM models and k-means algorithms have been used to classify the customers. With the aid of these models and algorithms, businesses have pure knowledge about their customers. Businesses could develop appropriate and special strategies about their customers easily by grouping customers according to their data. It can be noted that the RFM-analysis implementation as a segmentation tool is an effective way to amend the work with client groups and, as a consequence, to increase sales and profits. Adaptation of enhanced RFM-analysis on the example of e-commerce and retail trade of high-tech products has identified three target groups of consumers. According to this it is recommended to improve the marketing and promotion policy.

REFERENCES

[1] Li, B., & Chen, Y.-C. (2018). *The Application of Clustering Mining Technology in E-Commerce Website. 2018 9th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP).* doi:10.1109/paap.2018.00025

[2] CHA, N., CHO, H., LEE, S., & HWANG, J. (2019). *Effect of AI Recommendation System on the Consumer Preference Structure in e-Commerce: Based on Two types of Preference. 2019 21st International Conference on Advanced Communication Technology (ICACT).* doi:10.23919/icact.2019.8701967.

[3] Rao, H., Zeng, Z., & Liu, A. (2018). *Research on personalized referral service and big data mining for e-commerce with machine learning. 2018 4th International Conference on Computer and Technology Applications (ICCTA).* doi:10.1109/cata.2018.8398652

[4] Dongre, J., Prajapati, G. L., & Tokekar, S. V. (2014). *The role of Apriori algorithm for finding the association rules in Data mining. 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT).* doi:10.1109/icicict.2014.6781357

[5] Chauhan, N., & Singh, P. (2017). *Feature based opinion summarization of online product reviews. 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM).* doi:10.1109/iconstem.2017.8261247

[6] Bansal, S., Gupta, C., & Sinha, A. (2017). *Clickstream & behavioral analysis with context awareness for e-commercial applications. 2017 Tenth International Conference on Contemporary Computing (IC3).* doi:10.1109/ic3.2017.8284328

[7] Tavakoli, M., Molavi, M., Masoumi, V., Mobini, M., Etemad, S., & Rahmani, R. (2018). *Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study. 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE).* doi:10.1109/icebe.2018.0002.