JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CUSTOMER ANALYTICS BASED ON SEGMENTAION, RETENTION AND APRIORI ALGORITHM

[1]Dr. Sunil Bhutada, [2]SriGeethika Nagavajyula, [3]P Tarun

[1]Head of the Department, [2]Student, [3]Student
[1]Department of Information Technology,
Sreenidhi Institute of Science And Technology, Hyderabad, India

Abstract: Customer segmentation and retention are the main components of customer analytics. The main aim of Customer Segmentation is to classify the customers of a particular market place or any retail entity into different groups using which the organization can attract the customers back with new ideas, offers and combos. The retention part of the Customer Analytics mainly focuses on bringing back the customers to the same store or organization. It deals with providing the customers with their needs and ultimately with satisfaction. We perform simple customer data analytics followed by calculating the retention rates from the taken data set. The segmentation of customers into various groups is done in 2 ways. One of them is using RFM Analysis and the other method is by using K-Means Clustering. Finally, Apriori algorithm is implemented in order to find the buying patterns of the customers. Customer retention can put any organization ahead of its competitors by increasing its own market economy and profit. Customer segmentation acts as a helping factor in customer retention, using which we can bring in customer group specific offers or policies to attract such consumer groups. It also helps in building loyalty of the customers towards the retail organization.

IndexTerms - **Customer Analytics, Segmentation, Retention, Apriori Algorithm, K-means, RFM Analysis, Basket Analysis.**

1. INRODUCTION

Customer Segmentation and Retention is understood as dividing the customer set of any retail organization or a market place into clusters based upon the similarity, whereas retention is seen as making the customers come back for more to the same store again and again. Customer Analytics is a medium through which we can perceive customer segmentation and retention in a concise and comprehensive way. Customers are the main component of any organization; thus, it is important for any store in the market place to retain its customers such that it has a stable business and no uncertainty. Hence, it can be said that customer retention should be one of the main focuses of any retail selling store.

Using Customer Analytics, we try to build a foundation for using machine learning in customer segmentation and retention. The segmentation part is an important step in coming up with strategies for customer retention. To implement the segmentation part where the customers are divided into various clusters such as degrading, decent, valuable, most valuable, lost, guests, and new customers, we would be using RFM analysis which expands to be Recency Frequency Monitoring. In the RFM analysis, we would be calculating the RFA index of each customer based upon their unique customer id. It can be said that the customers having high RFA index are more frequent buyers and are more active, whereas those customers with lesser RFA index are not actively spending. This RFA index would me based upon their frequency of purchasing and the amount they spend on purchasing.

Coming to the retention part of the analytics idea, it would mainly focus on bringing back customers to the retail store or market place. It would be done by seeing how many customers are being retained every month based upon the previous months' data. Here we would be making sure to identify the customer's first purchase month which would be later known as the cohort month. And successively we would be calculating the retention percentage.

The key step in this proposed idea is to retain customers, and we need to come up with retention strategies which would pull customers towards us, which will be done in the final step of the implementaion. Apriori algorithm will be used to perform association rule mining, using which we would come with associations between the products the customer buys and we would come up with combo offers which would mean a great deal to the customer. Thus, the Customer Analytics mainly focuses upon customer segmentation and retention using frequent pattern mining.

2. LITERATURE SURVEY

In [1] Art Weinstein mainly focused on how customer retention was important for the company and how it would boost their customer value in the long run, and emphasized on customer value/retention model; and verified how the usage of customer segmentation can accommodate in building a valuable relationship with the customer and earning profits for the organization.

In [2] A. Joy Christy, A. Umamakeswari, L. Priyatharsini, A. Neyaa focused on the importance of customer segmentation, and said that effective customer segmentation would result in the organization finding about its most potential customers. The paper says that it is more important to retain the existing customers rather than looking for new customers. Thus, once we have the cluster of customers who spend the most or are the most loyal to the organization, then the organization can deploy marketing strategies and schemes that are customized for that cluster of loyal customers which would further result in retaining that set of customers. In this paper, the idea of using the initial centroids in K- Means is proposed which has been a founding step for our proposed idea.

In [3] Tianyi Jiang and Alexander Tuzhilin proposed that it is important to segment the customers efficiently to offer customized recommendations, products and services to them. They said that conventionally the customer segmentation was done using statistics – based methods and distance – based clustering algorithms. They proposed a direct grouping – based approach to segment the customers effectively.

In [4] Guozheng Zhang said that customer segmentation was the need of the hour, in today's competitive world of commercial arena. It was proposed that mostly customer segmentation was done using only a single data mining technology from a special point, but it is more efficient to carry out customer segmentation from a systematical framework. It was proposed that one of the main purposes of customer segmentation was to achieve customer retention.

3. PROPOSED SYSTEM AND ARCHITECTURE

3.1 Proposed System:

The main aim of this Customer Analytics is customer retention, for which we would be making use of customer segmentation. Customer segmentation focuses upon dividing the customers into various cluster based upon their characteristics. The customers who are placed in the same cluster are more alike, and much different from the customers of the other clusters.
The segmentation of the customers in this proposition will be carried out by using the K-means clustering algorithm. Before that we would be making use of RFM analysis to rank the customer groups based upon recency, frequency and monetary. Here, recency factor is given by seeing how often the customer is purchasing goods from the company, whereas frequency would constitute to how many times the customer is buying the same goods, and at the end the monetary would comprise of how much the customer is spending at the store while purchasing the goods.
In the due course, when we obtain the output of customer segmentation, we can identify the customer groups which are best ranked to provide the appropriate profit to the company, and we can run customized offers for those customer groups as a part of our marketing campaign. We would be making using of Apriori Algorithm for Basket Analysis where we would be studying what the customer is purchasing. Apriori algorithm would help us in association rule mining, as it generates candidate sets which can be carefully studied to generate recommendations to organization to sell goods.
In the end, we can say that Apriori algorithm is an effective way of frequent pattern mining, which would enable us to understand the customer's buying patterns, which would help the company to run customized and targeted offers to that cluster of customers in the form of combo offers.

4.2 Proposed Architecture:

The main prerequisites to carry out the implementation of customer analytics using segmentation and retention would be data pre-processing, where in we make sure to thoroughly cleanse the data from any ambiguities or null values which might cause a concern of outliers in our clustering algorithm. The next set of processes would involve customer data analysis to study the customer behaviour and customer segmentation to group the customers based upon their similarities and differentiate the clusters from one another based upon their dissimilarities. The most crucial part of the customer analytics is customer retention, where in we try to retain the company's most valuable customers which generate the profit and result in the growth of the organization. Most commonly the organizations which apply and implement this idea would be retail organizations, shopping malls, grocery stores etc.

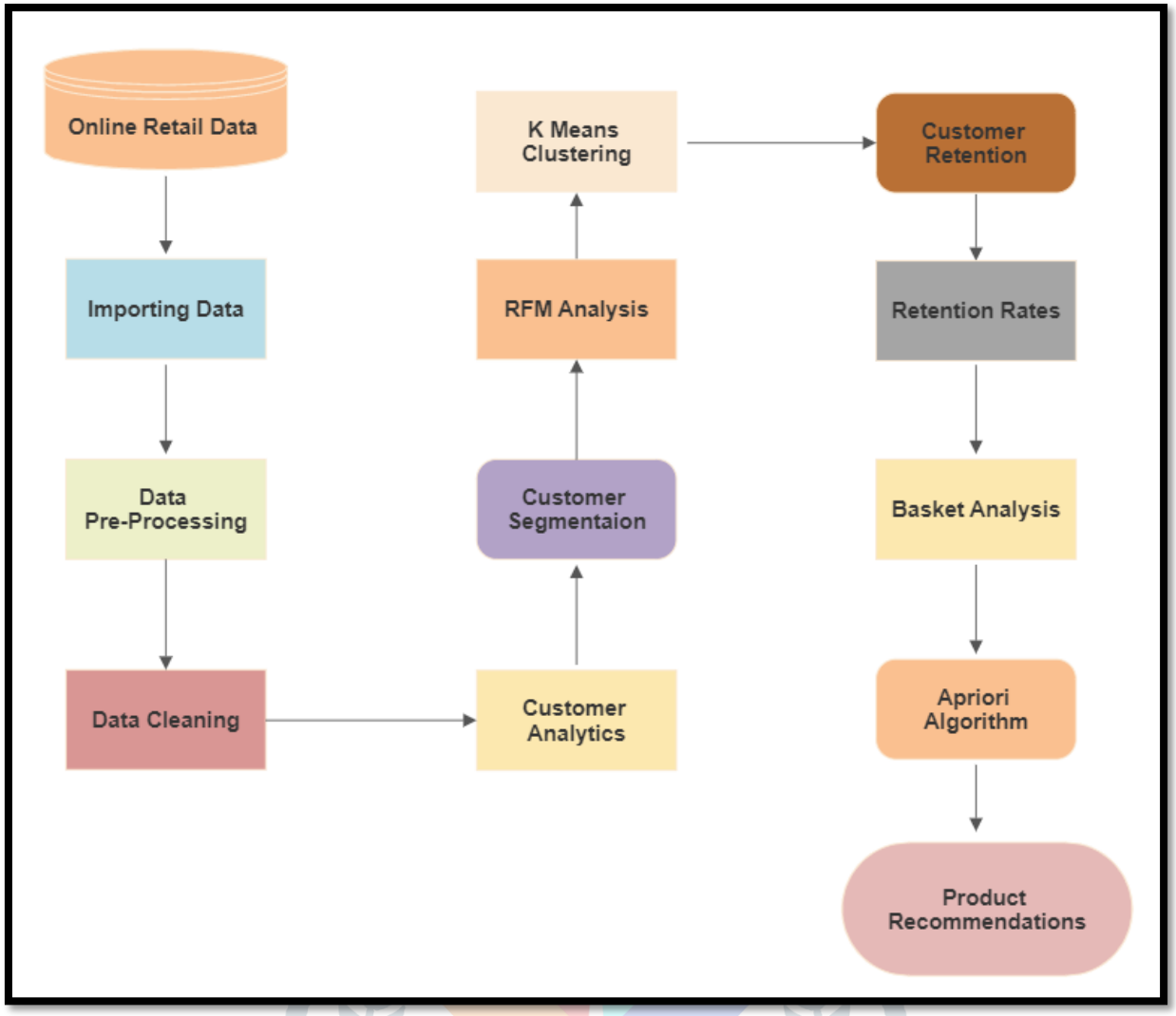

Figure 1: The Architecture of the Proposed System

5. RESULTS

The Elbow graph gives us a perspective into the data, we identify that it is ideal to divide the customers into four clusters. It would be most effective to deal with the data when we have four clusters. We target the most valuable customers and customize the offers so that they buy all the products they need in one go. Thus, we see that it is a lot easier to retain the already existing customers, rather than trying to attract newer customers.

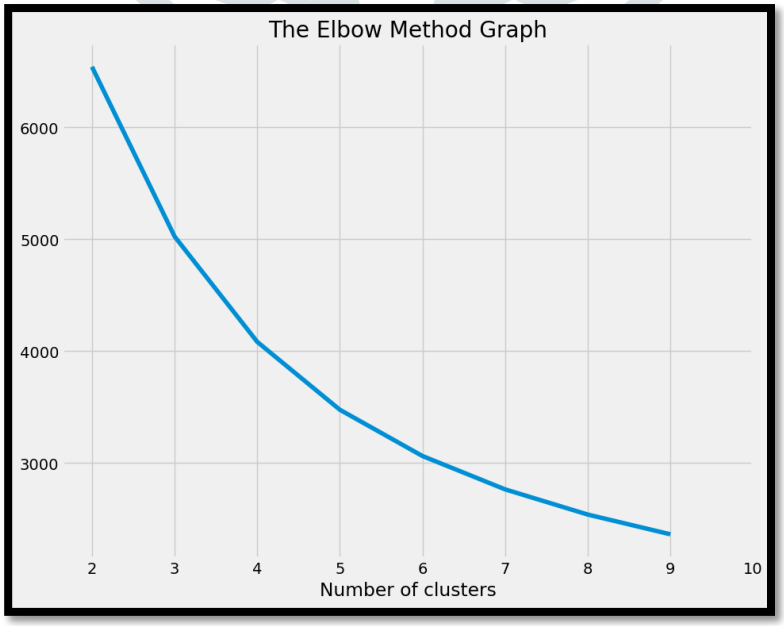

Figure 2: Elbow graph representing that it is ideal to divide customers into 4 clusters

After the data is clustered into 4 groups, we use this data as input to the Apriori Algorithm model which would give out the output of recommendations using association rule mining as shown in the figure below.

```
['6 GIFT TAGS VINTAGE CHRISTMAS']
There are no Product recommendations
['6 RIBBONS ELEGANT CHRISTMAS']
There are no Product recommendations
['6 RIBBONS EMPIRE']
There are no Product recommendations
['6 RIBBONS RUSTIC CHARM']
There are no Product recommendations
['6 RIBBONS SHIMMERING PINKS']
There are no Product recommendations
['6 ROCKET BALLOONS']
There are no Product recommendations
['60 CAKE CASES DOLLY GIRL DESIGN']
People who bought this also bought: ['PACK OF 72 RETROSPOT CAKE CASES']
['60 CAKE CASES VINTAGE CHRISTMAS']
People who bought this also bought: ['SET OF 20 VINTAGE CHRISTMAS NAPKINS']
['60 GOLD AND SILVER FAIRY CAKE CASES']
There are no Product recommendations
['60 TEATIME FAIRY CAKE CASES']
People who bought this also bought: ['72 SWEETHEART FAIRY CAKE CASES']
['6PC WOOD PLATE SET DISPOSABLE']
There are no Product recommendations
['72 SWEETHEART FAIRY CAKE CASES']
People who bought this also bought: ['PACK OF 60 DINOSAUR CAKE CASES']
['75 BLACK PETIT FOUR CASES']
There are no Product recommendations
['75 GREEN FAIRY CAKE CASES']
There are no Product recommendations
['75 GREEN PETIT FOUR CASES']
```

Figure 3: Product Recommendations given out by Apriori algorithm

6. CONCLUSION AND FUTURE ENHANCEMENT

The main focus of this idea has been customer retention. To achieve customer retention, we focused on first segmenting the customers into various clusters which helped us in categorizing them into Most Valuable Customers, Valuable Customers, Decent Customers, Degrading Customers, Lost Customers and Guests or New Customers. We have then identified the most valuable groups which spend a lot of bucks in buying goods. We recognized the profitable customers by making use of RFM analysis and the RFM index. We then tried to bring out new strategies for attracting those customers (retaining) into buying more from the store. We have implemented the Apriori algorithm for discovering the associations between the products bought by the customer, this can be also seen as frequent pattern mining, where we look into the frequent buying patterns of the customer and we try to come up with offers and combos which are most suitable for our customers to satisfy them. This appears to be very helpful for any organization in the market place as it boosts its profit by a large margin and puts it ahead in the race with its competitors. In conclusion, this is rightfully seen as the amalgamation of customer satisfaction as well as the company's satisfaction, as the customer is happy to get the best deals and the retail owner is joyous about making a greater profit.

Looking at brighter outcomes of any concept is very important, in this implementation of the idea we have used the Apriori algorithm to carry out the association rule mining, but it can be seen that there are minor disadvantages to this approach. It can be said that using FP growth algorithm can be seen as a stepping stone for this idea of implementation. As in Apriori algorithm, it requires a large amount of memory space as a greater number of candidates are generated. On the other hand, in the FP Growth algorithm, there is no candidate generation as it has a compact structure and makes use of lesser memory. The scope of this idea is exemplary, we can further develop a more visual approach to this concept which comes with a user interface, which can be easily understood and used by the company staff. We can focus more upon providing the best products to the customers, by making them happy and satisfied as this approach would be carried out among more and more organizations across the globe. The ultimate goal of any product's enhancement is the satisfaction and prosperity of its users.

7. REFERENCES

1. Weinstein, A Customer retention: A usage segmentation and customer value approach. J Target Meas Anal Mark 10, 259-268 (2002).

2. A. Joy Christy, A. Umamakeswari, L. Priyatharsini, A.Neyaa of Department of CSE, School of Computing, SASTRA Deemed to be University, Thanjavur, India, RFM ranking – An effective approach to customer segmentation, Journal of King Saud University – Computer and Information Sciences, Volume 33, Issue 10, December 2021, Pages 1251-1257.

3. Tianyi Jiang, AvePoint, Inc. Jersey, NJ, USA and Alexander Tuzhilin, Department of Information, Operations and Management Sciences, Stern School of Business, New York University, New York, NY, USA, Improving Personalization Solutions through Optimal Segmentation of Customer Bases, IEEE Transactions on Knowledge and Data Engineering (Volume 21, Issue: 3, March 2009).

4. Guozheng Zhang, College of Business, Houzhou Dianzi University, Hangzhou, China, Customer Segmentation Based on Survival Character, 2007 International Conference of Wireless Communications, Networking and Mobile Computing. Added to IEEE Xplore on 08 October 2007.