

RFM ANALYSIS FOR CUSTOMER SEGMENTATION USING MACHINE LEARNING: A SURVEY OF A DECADE OF RESEARCH

Sushilkumar Chavhan

Assistant Professor, Department of Information Technology Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, (India).

R. C. Dharmik

Assistant Professor, Department of Information Technology Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra, (India).

Sachin Jain

Assistant Professor, Department of Computer Science Oklahoma State University Stillwater, (United States).

Ketan Kamble

Student, Department of Information Technology YCCE, Nagpur, Maharashtra, (India).

Reception: 07/11/2022 **Acceptance:** 22/11/2022 **Publication:** 29/12/2022

Suggested citation:

Chavhan, S., Dharmik, R. C., Jain, S., y Kamble, K. (2022). RFM analysis for customer segmentation using machine learning: a survey of a decade of research. *3C TIC. Cuadernos de desarrollo aplicados a las TIC*, 11(2), 166-173. <https://doi.org/10.17993/3ctic.2022.112.166-173>



<https://doi.org/10.17993/3ctic.2022.112.166-173>

ABSTRACT

Customer segmentation is a method of categorizing corporate clients into groups based on shared characteristics. In this study, we looked at the different customer segmentation methods and execute RFM analysis by using various clustering algorithms. Based on RFM values (Recent, Frequency, and Cost) of customers, the successful classification of company customers is divided into groups with comparable behaviors. Customer retention is thought to be more significant than acquiring new clients are analyzed on two different databases. Results show the significance of each method. Comparison is helps for selection of better customer segmentation.

KEYWORDS

Clustering, Classification, RFM, Customer segmentation.

1. INTRODUCTION

Customer Segmentation is way of organization of customers with respect to the various features. In recent years there has been a huge boom in opposition between companies to stay in the field. The income of the organization may be stepped forward through a patron segmentation model. According to the Pareto principle (Srivastava, 2016), 20% of the clients make a contribution greater to the sales of the organization than the relaxation Customer segmentation is the exercise of dividing an organization's clients into agencies that mirror similarity amongst clients in every group. The intention of segmenting clients is to determine how to narrate to clients in every section that allows you to maximize the fee of every patron to the business. Customer segmentation has the ability to permit entrepreneurs to cope with every patron withinside the simplest way. Using the massive quantity of statistics to be had on clients (and ability clients), a patron segmentation evaluation lets in entrepreneurs to identify discrete agencies of clients with an excessive diploma of accuracy primarily based totally on demographic, behavioral and different indicators.

Evaluation of RFM (Recency, Frequency, and Monetary) is a famend approach is worn for comparing the clients primarily based totally on their shopping for behavior. Scoring method was developed to test Recent, Frequency, and Finance ratings. Finally, ratings of all three variables are strengthened as RFM ratings from different ranges (Haiying and Yu, 2010) which are compiled to anticipate recants trends for studying existing and higher sponsor transactions history. Next step is defined as the remaining time the consumer buys. The latest currency is the type of days the sponsor takes between purchases. The latest small payment means that the sponsor visits the organization frequently in a timely manner. Similarly, extra money means that the sponsor is less likely to go to the organization soon. Frequency is described because the variety of transaction a patron makes in a selected period. The better the fee of frequency the greater unswerving are the clients of the organization.

Cash is defined as the amount spent by the investor over a period of time in a favorable period. The improvement in the amount of money spent by the large sales they provide to the organization. Each sponsor is given 3 different ratings of the latest, frequency, and economic volatility. Score points are used within a range from five to 1. The core quintile is given a five-point scale, while the others are given 4, 3, 2 and 1.

In recent years, there has been a significant increase in the number of opposition groups among companies in care within the arena. Customer retention is more important than purchasing the latest customers. Customer segregation allows people's messages to speak more to target audiences.

2. LITERATURE REVIEW

Segmentation is middle of the advertising and marketing approach due to the fact exclusive consumer organizations mean the want for exclusive advertising and marketing mixes primarily based totally on consumer conduct and its needs. Many authors give the segamentaion methods to increase the profit and sustain the company position. (Jiang and Tuzhilin, 2009) proposed K-Classifiers Segmentation algorithm which recognized that each client segmentation and consumer focused on are important to enhance the marketing performances. K-Classifiers Works as optimizer who have two tasks. Above method more resources to the ones clients who supply greater returns to the company. (He and Li, 2016) proposed a 3-dimensional approach to improving consumer health (CLV), customer pride and customer behavior. The authors conclude about the customers and the requirements for a better service. A segment used to meet customer expectations and suggest better service. (Sheshasaayee and Logeshwari, 2017) used RFM Analysis which provides the usage of CRM (Customer Relationship Management). Authors analyzed the customers by segmenting them which helps to increase company profits. Further they enhanced the segmentation by using Fuzzy Clustering Method which classified them into the appropriate scoring strategies based on their needs.

(Shah and Singh, 2012) provides the K-means algorithm and K-medoids algorithms for clustering.

The presented techniques do not always yield the best answer, but they do minimise the cluster error criterion. They came to the conclusion that when the number of clusters grows, the new method takes

less time to run than existing methods. (Sheshasaayee and Logeshwari, 2017) developed hybrid method which combine RFM and LTV methods. Authors used K-means and Neural Network algorithms for segmentation with two phase models. They suggested having better optimizer for customer categorization. Using logistic regression, (Liu, Chu, Chan, and Yu, 2014) proposed predicting customer attrition. Individual marketing methods can be used to identify customers with similar churn value and to keep them. Benefit customer segmentation using various methodologies allows customers to be classified based on their relationships, allowing marketers to focus their marketing efforts on their strengths and target benefit categories accordingly.

3. TYPES OF CUSTOMER SEGMENTATION

Customer segmentation models come in a range of shapes and sizes, ranging from simple to complex, and they can be used for a variety of purposes. Demographic, Recency, Frequency Monetary (RFM), High-Value Customers, Customer Status, Behavioral, and Psychographic models are some of the most common models.

3.1 DEMOGRAPHIC

It is a method of segmenting customers based on characteristics such as age, gender, ethnicity, income, education, religion, and career (Lu, Lin, Lu, and Zhang, 2014).

3.2 RFM

It is a direct segmentation strategy whose main goal is to categories clients based on the time since their previous purchase, the total number of purchases they've made (frequency), and the amount they've spent (monetary) (Sheshasaayee and Logeshwari, 2017).

3.3 HVCS (HIGH-VALUE CUSTOMER)

It's an extended RFM segmentation for any firm, focusing on what traits they have in common so you can get more of them.

3.4 CUSTOMER STATUS

It is a mechanism which check the status of customer which categories as active and lapsed. The focus of this method is to how the customer engaged by the company on the time period as a status.

3.5 BEHAVIORAL SEGMENTATION

It is a mechanism which check the status of customer which categories as active and lapsed. The focus of this method is to how the customer engaged by the company on the time period as a status.

3.6 PSYCHOGRAPHIC SEGMENTATION

It is allows grouping the customers based on attitudes, beliefs, or even personality traits. For this we require good data analysis method. Analysis done on all above attributes.

3.7 GEOGRAPHIC SEGMENTATION

It is allows grouping the customers based on geographical location i.e region, city, country etc. It is used when target area is location wise improvement of services and increased the profit.

4. MAJOR CLUSTERING TECHNOLOGIES FOR CUSTOMER SEGMENTATION

4.1 K –MEANS

It is a popular unsupervised method that accepts parameters and k value as number of clusters as inserting and separating data into clusters with high intra-cluster similarities. K-Means is a method that repeats itself, adding the number of centroids before each multiplication. Depending on the inches calculated for each multiplication, data points are allocated among distinct sets. Using min max normalization, the RFM values are normalized (Lee and Memon, 2016).

4.2 FUZZY C-MEANS

It is a Method (?, ?) that allows a specific piece of data to appear to numerous clusters. It no longer determines a cluster's club records for a given information factor. Rather, the probabilities of a specific information factor with similarities are determined. The advantage of this method over previously discussed K-Means is that the final result obtained of a large and comparable database is most suitable than a set of K-method rules, because in the KMeans method, as cluster formation of based on data element. The normalization in this approach is done using min max normalization. Cluster RFM value is based on cluster (Zahrotun, 2017) value.

4.3 REPETITIVE MEDIAN K-MEANS

It's a novel approach to determining the initial centroids for the K-Means method. The traditional K-Means algorithm's range of iterations and computational time is reduced by choosing preliminary centroids with it's proposed distribution. RFM values will be combined and sorted into three vectors, R', F', and M', respectively. The initial centroids are calculated using the median value of each vector. The median values are derived k times iteratively from the R', F', and M' values, depending on the value of k. (number of segments).

5. RFM ANALYSIS USING ON ONLINE SHOPPING DATA

RFM: according to following definitions. The RFM method divides clients into segments. It categorises clients based on their previous purchase transactions, taking into account criteria like as Recency (R): Last purchase date in specified session. Frequency (F):Purchase count in the specified session. Monetary (M):Value of Purchase in the specified session Based on forms needs, we define a season with different intervals for this model and we calculate RFM values for each customer. Working with a set of customer activity data in an online retail store year-round from the University of California Irwin (UCI) repository was used to evaluate system performance.

The following is a sample customer separation process. STEP 1: Sort the customer by recency.

STEP 2 AND 3: Sort the customer from most to least frequent customer and summarized the F, M score. STEP 4: Rank customers by combining R,F, and M ranking

CUSTOMER ID	REGENCY (RANK)	FREQUENCY (NUMBER)	MONETARY (TOTAL)
1	4	6	640
2	6	11	940
3	46	1	35
4	23	3	65
5	15	4	179
6	32	2	56
7	7	3	140
8	50	1	950
9	34	15	2430
10	10	5	191
11	3	8	845
12	1	10	1630
13	27	3	54
14	18	2	40
15	5	1	25

Fig1: Sample Data.

CUSTOMER ID	REGENCY	RANK	R SCORE
12	1	1	5
11	3	2	5
1	4	3	5
15	5	5	4
2	6	5	4
7	7	6	4
10	10	7	3
5	15	9	3
14	18	9	3
4	23	10	2
13	27	11	2
6	32	12	2
9	34	13	1
3	46	14	1
8	50	15	1

Fig2: Calculation of Recency.

CUSTOMER ID	FREQUENCY	F SCORE	CUSTOMER ID	MONETARY	M SCORE
9	15	5	9	2500	5
2	8	5	12	1500	5
12	10	5	8	950	5
11	8	4	2	940	4
1	6	4	11	890	4
10	5	4	1	540	4
5	4	3	10	490	3
13	3	3	5	179	3
7	3	3	7	140	3
4	3	2	4	65	2
14	2	2	6	64	2
6	2	2	13	54	2
15	1	1	14	40	1
8	1	1	3	35	1
3	1	1	15	26	1

Fig3: Calculation of F and M Score.

CUSTOMER ID	RFM CELL	RFM SCORE
1	5,4,4	4.3
2	4,5,4	4.3
3	1,1,1	1.0
4	2,2,2	2.0
5	3,3,3	3.0
6	2,2,2	2.0
7	4,3,3	3.3
8	1,1,5	2.3
9	1,5,5	3.7
10	3,4,3	3.3
11	5,4,4	4.3
12	5,5,5	5.0
13	2,3,2	2.3
14	3,2,1	2.0
15	4,1,1	2.0

Fig4: Customer Segmentation.

6. ANALYSIS OF VARIOUS ALGORITHMS

For evaluation of commonly used clustering algorithm was done on two different open source databases like the transactional data set of the customers of an online retail store available at UCI repository and on e-commerce datasets which is available at UCI Machine Learning Repository.

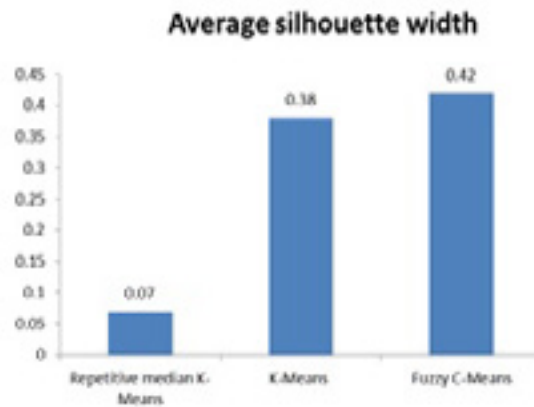


Fig5: Result Analysis of various Algorithms on online retail.

From above analysis it is observed that Fuzzy C means perform well on the basis of average Silhouette width but the time taken is more and number of iterations are also more as compared to others algorithms.

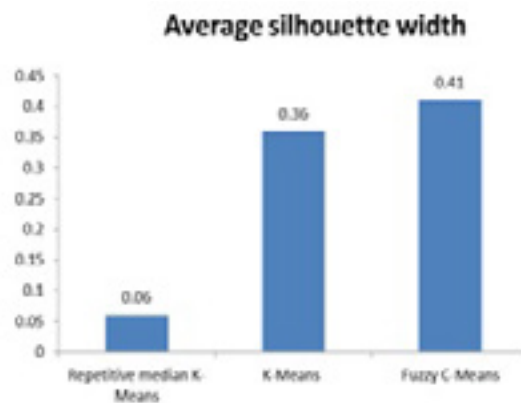


Fig6: Result Analysis of various Algorithms on E-commerce data.

7. CONCLUSION AND FUTURE SCOPE

This paper provides the overview of customer Segmentation and its different type's with types. RFM segmentation allows to group based on the requirements and target the different marketing strategies. In future more complex methods would be designed to target specific customers and the methods would also be more flexible if the company wants to target a different audience for a particular time or wants to permanently change their customers based on the needs of the company or priorities of the customer. The RFM should be made more flexible according to the needs of the different companies.

REFERENCES

- [1] Haiying, M. and Yu, G. 2010. Customer segmentation study of college students based on the rfm. 3860–3863.
- [2] He, X. and Li, C. 2016. The research and application of customer segmentation on e-commerce websites. 203–208.
- [3] Jiang, T. and Tuzhilin, A. 2009. Improving personalization solutions through optimal segmentation of customer bases. *IEEE Trans. Knowl. Data Eng.* 21, 305–320.
- [4] Lee, D.-H. and Memon, K. 2016. Generalised fuzzy c-means clustering algorithm with local information. *IET Image Processing* 11.
- [5] Liu, C., Chu, S.-W., Chan, Y.-K., and Yu, S. 2014. A modified k-means algorithm - two-layer k-means algorithm. 447–450.
- [6] Lu, N., Lin, H., Lu, J., and Zhang, G. 2014. A customer churn prediction model in telecom industry using boosting. *Industrial Informatics, IEEE Transactions on* 10, 1659–1665.

- [7] Shah, S. and Singh, M. 2012. Comparison of a time efficient modified k-mean algorithm with k-mean and k-medoid algorithm.
- [8] Sheshasaayee, A. and Logeshwari, L. 2017. An efficiency analysis on the tpa clustering methods for intelligent customer segmentation. 784–788.
- [9] Srivastava, R. 2016. Identification of customer clusters using rfm model: A case of diverse purchaser classification.
- [10] Zahrotun, L. 2017. Implementation of data mining technique for customer relationship management (crm) on online shop tokodiapers.com with fuzzy c-means clustering. 299–303.