IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Review Paper on Customer Segmentation Approach Using RFM and K-Means Clustering Technique

[1]Ms Sarika Rathi,[2]Prof. Vijay Karwande

[1]Lecturer in School of Engineering & Technology ,[2]HOD in Everest Engineering College
[1]Department of Computer Sciecnce & Engineering,[2] Department of Computer Science & Engineering
[1]MGM University, Aurangabad , Maharashtra,India

Abstract: In recent years, the significance of behaving with customers according to their history has been growth dramatically. Customer Segmentation has been used as the basis for understanding and grouping customers. Everymonth, there are thousands of customers are geenrated.Every day there is a transaction process performed by Customer. The process generates a lot of data. One of the most widely applied methods for segmenting customers is RFM model. This study aims to perform customer segmentation by utilizing data mining process based on RFM model and by using techniques Clustering. The algorithm used for cluster formation is K-Means algorithm.

The goal of this project is to analyze and segment the customers of an e-commerce company by using the RFM approach of given data. This will enable the e-commerce company to optimize their retention and acquisition strategies. Many businesses get most of their revenue from their 'best' or regular customers. Since the resources that a company has, are limited, it is crucial to find these customers and target them. It is equally important to find the customers who are dormant/are at high risk of churning to address their concerns. For this purpose, here I am trying to Build an unsupervised learning model which can enable company to analyze their customers via RFM (Recency, Frequency and Monetary value) approach.

The result of this research can be used by company to know customer category, and then the company will know how to maintain the customer owned.

Index Terms **- Data Mining; RFM Model; Cluster Analysis; Customer Segmentation, K-Means Algorithm.**

I. INTRODUCTION

As it is well known by marketers, customers have various kinds of needs and wants. Companies have used several segmentation criteria and techniques to better identify and understand customer groups and provide preferable products and services to them in order to satisfy these different needs and wants. Also, segmentation is important that the company can create profitable segments and react to the selected segment based on its competitive advantages. However, many marketers have difficulty in identifying the right customer segments to organize marketing campaigns.

Customer is always the first priority of every business; it has also proven many times that customer-oriented organizations are successful and ever growing in cooperate world. This was figured out by many organizations and they are trying to implement customer centric approach as their work criteria. To get into the shoes of the customers and try to merge according to the new trend followed by customers is constantly in generating huge profit.

Customer satisfaction is critical in customer relationship management to confirm an increasing graph of customer loyalty and retention.

When a person goes into a retail store and purchases a few items, the following basic data points should be generated:

1. Customer ID
2. Address/Contact number of the customer (Demographic information)
3. Invoice number
4. Product name and product code
5. Quantity
6. Unit Price
7. Date and time of the transaction

II. LITERATURE SURVEY

In recent research we have found that a case study of using data mining techniques to segment customers for an IT solution is presented. The objectives of this research are to construct a customer segmentation model based on 3 customer demographics and purchase behaviors and to help business better understand its customers and support their customer-centric marketing strategy. Regarding to the customers demographic data and RFM values generated from purchase behaviors, customers have been segmented using the K-means clustering technique into numerous groups based on their similarity, and the profile for each group is identified based on their characteristics. Accordingly, recommendations are provided to the business on marketing strategy and further analysis.

In today's business competition, customers are the main focus of the company to maintain its excellence. Companies must plan and use clear strategies in serving customers [1][2]. The company's primary focus is not on how to get new potential customers but how to sell more products to the existing customers because the cost that companies must in curt acquire new customers is much more expensive than to retain existing customers [3].

RFM model is extremely useful in ceitbuilds customer segmentation model effectively. However, simplicity threatens the power of RFM and the models need to be made and be improved by a manual process. Moreover, RFM models cannot confront with changes in the business and managers should handle them by adhoc decisions. In this paper, we found the best definitions for R, F and M to have a dynamic RFM model and also using K-Means in order to propose R+FM model which builds customer segmentation model dynamically.[6]

There are 4customer categories that demand company to give different service to customer. RFM technique is based on three simple customer attributes, namely Recency of purchase, Frequency of purchase, and Monetary value of purchase. The values of recency, frequency and monetary are combined to form RFM scores. For example, in a five category ranking system, there are about 125 possible RFM scores and the highest RFM score is 555. RFM scores clearly shows the categories of different consumers. The best customer search with the highest RFM scores. In this paper, the ranking1-4 isused to evaluate the customer retention.[7]

The model used by the researcher is RFM (Recency,Frequency, Monetary) commonly used to perform the last visit time grouping, visit frequency, and revenue obtained by the company [4]. There as on why continuing to use the RFM model is that it is easy to use and quickly implemented in companies, and in addition RFM is easily understood by managers and marketing decision makers[5].

P. Anitha et al. applied the k-means algorithm to the RFM model to evaluate the buying behaviour of users in different regions [12]. Siti Monalisa et al. applied the RFM model to property insurance, customer investment, telecommunications services , healthcare , and FMCGindustries [13].

III. A CASE STUDY

The dataset used in this case study is credit sales data on Nine Reload Credit Server. At the company there is a lot of data stacking, thousands of transactions every month. You can imagine how difficult it would be if you had to analyze the data manually one by one. The researchers tried to analyze the data as much as 82,648 customer transactions. The model proposed in determining the profitable customer is described in Figure 1 which shows the steps to determine the profitable customer.

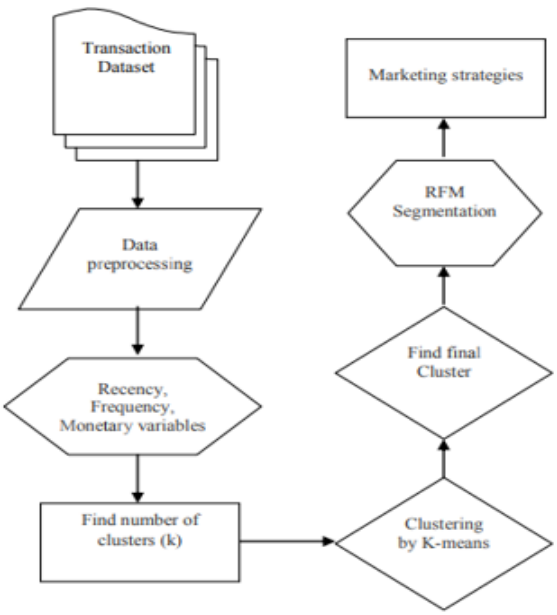

Fig. 1 Architecture of Customer SEgmentaion Implementation

In this study the database used is the data collected from the transaction as much as 5,37,979 customers id and with 12 different attributes are there.Table 1is anexample of a database.

Table 1 Sample Database

CustomerID	Item Code	InvoieNo	Date of purchase	Quantity	Time	price per Unit	Price	Shipping Location	Cancelled_status	Reason of return	Sold as set
4355	15734	398177	29-10-2017	6	3:36:00 PM	321	1926	Location 1			
4352	14616	394422	05-10-2017	2	2:53:00 PM	870	1740	Location 1			
4352	14614	394422	12-10-2017	2	2:53:00 PM	933	1866	Location 1			
4352	85014B	388633	22-08-2017	3	2:47:00 PM	623	1869	Location 1			
4352	15364	394422	10-10-2017	2	2:53:00 PM	944	1888	Location 1			
4349	14618	397122	27-10-2017	1	12:43:00 PM	256	256	Location 1			
4343	15364	368432	13-02-2017	-4	2:46:00 PM	922	-3688	Location 1	TRUE		
4341	85014B	377109	14-05-2017	3	9:22:00 AM	677	2031	Location 1			
4341	85014A	377109	12-05-2017	3	9:22:00 AM	692	2076	Location 1			
4341	85014B	390217	07-09-2017	6	2:47:00 PM	670	4020	Location 1			
4341	85014B	389462	04-09-2017	6	10:55:00 AM	674	4044	Location 1			
4339	85014B	361897	11-12-2016	3	5:48:00 PM	624	1872	Location 1			
4333	14614	375503	24-04-2017	2	12:17:00 PM	875	1750	Location 1			
4331	14616	394146	01-10-2017	2	11:55:00 AM	851	1702	Location 1			
4331	14614	372098	23-03-2017	2	10:30:00 AM	861	1722	Location 1			

Table 2 Attributes of Database

Column Name	Description
CustomerID	Unique idetifier for each Customer
Item Code	Unique id for each product
InvoieNo	Unique id for each purchase
Date of purchase	Date on which the purchase was made
Quantity	Number of items bought for each product
Time	Time at which the purchase was made
price per Unit	Price of single unit of item purchased
Price	total purchase price
Shipping Location	Dilivery Loaction
Cancelled_status	Status of Cancellation
Reason of return	Reason for return of product
Sold as set	Was the product sold with another product/ Offer

3.1 Data Mining

Data mining is a process that uses statistics, mathematics, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases. Data mining is a part of knowledge discovery data which is a useful, unknown, and hidden information extraction process from data.

3.2 DataPreparation

At this stage the database structure will be prepared so as to simplify the mining process. The data set also consist of some duplicate entries & Null values. Provided data set is between the period 02-12-2016 to 19-12-2017 (Approx 1 year). The preparation process includes three main things :selection, pre-processing, and transformation data. This process also carries out the selection of attributes that are adjusted to the data mining process. Here we first remove null values, duplicate values and negative values form database. After removing all above things now database size is 368829.

On this dataset now we perform operations to find best , good , loyal and lost customer and result will be published soon.

3.3 Customer Segementaion

Customer segmentation is a process of dividing the Customers into several segments or categories in a way such that they are very similar to one another relevant to marketing, such as gender, age , interests ands pending habits[8]. Inorder to segment customers into different groups or segments RFM variables are widely used.

Customer segmentation based on RFM (Recency,Frequency& Monetary) variables is very popular now a days to partition the customers pursuant to their characteristics and behaviour. Market segmentation means the combination of buyer's or customers into several segments or groups using some criteria such that they share or represent something that is common. Market segmentation allows different companies or business farms target and select different categories of customers in term so their characteristics and behaviour for buying products or services or their interests on certain products [9].

Factors for segmentation for a business to business marketing company:

1. Industry
2. Number of employees
3. Location
4. Market cap/Company size
5. Age of the company

Factors for segmentation for a business to consumer marketing company:

1. Demographic: Age, Gender, Education, Ethnicity, Income, Employment, hobbies, etc.
2. Recency, Frequency, and Monetary: Time period of the last transaction, the frequency with which the customer transacts, and the total monetary value of trade.
3. Behavioral: Previous purchasing behavior, brand preferences, life events, etc.
4. Psychographic: Beliefs, personality, lifestyle, personal interest, motivation, priorities, etc.
5. Geographical: Country, zip code, climatic conditions, urban/rural areal differentiation, accessibility to markets, etc.

IV. CLUSTERING ALGORITHMS FOR SEGMENTATION

Clustering can be defined as the process of organizing objects in a database into clusters/groups such that objects with in the same cluster have a high degree of similarity, while objects belonging to different clusters have a high degree of dissimilarity.

Clustering is an unsupervised learning approach in which the raw data are classified in a way so as to discover the inherent patterns that exist in data[10]. Clustering partitions data points into smaller number of clusters such that objects are very similar within a cluster and very discordant to objects in other clusters.Clustering techniques [6,7] fall into a group of un directed data mining tools. The goal of undirected data mining is to discover structure in the data as a whole.

4.1 K-meansClusteringAlgorithm

K-means is a well-known unsupervised, iterative, partitioning learning algorithm in the field of data mining. It is used in solving various clustering problem especially for large datasets. The algorithm has two separate parts. In first part, Knumber of centers are selected randomly. K is initially fixed. In second part, every data object is taken to the closest center . In order to choose the initial value of K, self-organising map can be used.

V. Conclusion

The results of this study can be used as a decision support system in the credit business to map customers and to know potential customers. We are possible to improve the accuracy of recommendation using RFM method for item segmentation and clustering by item category so as to be able to reflect the attributes of items. As a result of that, we can propose the personalized recommendation system using k-means clustering of itemcategory based on RFM.[6]

The main purpose of this research was to segment the customers id from the transaction data of 82,648 based on RFMmodel, and furthermore clustering analysis was performed by usingK-Means

REFERENCES

[1] InaMaryani[1], DwizaRiana[2], RachmawatiDarmaAstuti[3], AhmadIshaq[4], Sutrisno[5], Eva Argarini Pratama[61,2,3] STMIKN usa Mandiri Jakarta,[4,5,6] UniversitasBina, Customer Segmentation based on RFM model and Clustering Techniques With K-Means Algorithm

[2] Maryani, Ina, and Dwiza Riana. 2017. "Clustering and Profiling of Customers Using RFM for Customer Relationship Management Recommendations." 2017 5thInternationalConferenceonCyberandITServiceManagement, CITSM 2017,2–7.https://doi.org/10.1109/CITSM.2017.8089258.

[3] Tama, Bayu Adhi. 2010. "Penetapan Strategi PenjualanMenggunakan Association Rules DalamKonteks CRM."JurnalGenericVol.5(No.1):35–38.

[4] Wongchinsri,Pornwatthana,andWerasakKuratach.2016. "A Survey -Data Mining Frameworks in CreditCard Processing." 2016 13th International ConferenceonElectricalEngineering/Electronics,ComputerTelecommunications and InformationTechnology, ECTI-CON2016.

[5] PeimanAlipourSarvari,AlpUstundag,and HidayetTakci.2014."PerformanceEvaluationofDifferentCustomer Segmentation ApproachesBased on RFM andDemographicsAnalysis."Kybernetes43(8):1209–23.https://doi.org/10.1108/K-01-2015-0009

[6] MohammadrezaTavakoli, MohammadrezaMolavi, Vahid, Masoumi , MajidMobini, SadeghEtemad and Rouhollah Rahmani , Customer Segmentation and Strategy Development based on User Behavior Analysis, RFM model and DataMiningTechniques:ACaseStudy, 2018 IEEE 15th International Conference on e-Business, Engineering (ICEBE), Page No.119-126, IEEE

[7] Sabbir Hossain Shihab, Shyla Shyla Afroge, Sadia Zaman Mishu, Department of Computer Science & Engineerin Rajshahi University of Engineering & Technology Rajshahi-6204, Bangladesh , RFM Based Market Segmentation Approach UsingAdvanced K-means and Agglomerative Clustering:AComparative Study,2019InternationalConferenceonElectrical,ComputerandCommunicationEngineering(ECCE),7-9February,2019, IEEE

[8] W. Fox, L. Kaufman and P. Rousseeuw, "Finding Groups in Data: AnIntroduction to Cluster Analysis.", *Applied Statistics*, vol. 40, no. 3, p.486,1991

[9] W.Pride,O.Ferrell,B. Lukas,S.Schembri,O.NiininenandR.Casidy,*Marketing principles*, 3rd ed. South Melbourne, Victoria : Cengage,2018,p.200.

[10] P. Fader, B. Hardie and K. Lee, "RFM and CLV: Using Iso-ValueCurves for Customer Base Analysis", *Journal of Marketing Research*,vol. 42,no.4,pp.415-430,2005.

[11] Yong HuangBusiness SchoolSichuanUniversity MingzhenZhang*Business School Sichuan University, Yue He Business School Sichuan University Chengdu,China , Research on improved RFM customer segmentation model based on K-Means algorithm , 2020 5th International , Conference on Computational Intelligence and Applications (ICCIA), Page no. 24-27, 2020, IEEE

[12] P. Anitha,Malini M. Patil. RFM model for customer purchase behaviorusing K-Means algorithm[J] .JournalofKingSaudUniversity-ComputerandInformationSciences,2019.

[13] Siti Monalisa,PutriNadya,Rice Novita. Analysis for Customer Lifetime Value Categorization with RFM Model[J].ProcediaComputerScience,2019,161.