International Journal of Engineering & Technology, 8 (1.9) (2019) 322-326

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Customer Segmentation Using Fuzzy C-Means Method and Fuzzy Rfm

Paramita Mayadewi[1], Dahliar Ananda[2], Tria Nur Paramadina[3]

[1,2,3]Department of Informatics Management, School of Applied Science, Telkom University, Bandung 40257, Indonesia
**Corresponding author E-mail: paramita@tass.telkomuniversity.ac.id*

Abstract

The study was conducted to explore the application of data mining in customer segmentation for laundry business. Intense competition in similar business encourage the company to manage its customer optimally. With a large number customers, the problem that has to be faced is how to determine potential customers. The process conducted is to divide customers into several segments with the aims to build customer profiles based on patterns of transactions that have been carried out. Customer profile that is created is a profile that shows the potential level of the customer. There are five categories of potential customers form highest to lowest. Implementations is done using two methods of data mining, namely clustering, and segmentation. Clustering method using Fuzzy C-Means algorithm while segmentation using Fuzzy RFM (Recency, Frequency, and Monetary) models. Studies conducted succeeded in grouping customers based on transcations conducted (Recency, Frequency, and Monetary). Therefore the mining results can be used to assist companies in the process of identifying the customer and also as an alternative marketing strategy.

Keywords: *data mining, customer segmentation, Fuzzy C-Means, Fuzzy RFM Model*

1. Introduction

The organization needs to develop innovation activities in order to capture the need of customers, so that can increase customer satisfaction and retention in face of complexity and competitiveness of business today. In connection with that matter, the company should be able to identify the best customers by increasing the company's understanding of customer needs as an individual so as to maintain loyalty to the company. The purpose of the customer segmentation process is to determine customer behavior and apply the right marketing strategy to be profitable for the company.

Cluster analysis is a method often used in customer segmentation [1]. The basic notion of cluster analysis is classifying customers into groups that are similar to one another. In data mining method, data processing to identify customer can be done using Fuzzy C-Means Clustering algorithm. In the process of clustering, some data will form groups (clusters) that contain similar data. The clustering results will be segmented into classes of customers whi have been determined using a classification model-Fuzzy RFM. The result of this segmentation will be used as a material to identify each customer.

The use of Fuzzy C-Means dan Fuzzy RFM methods are widely applied in Customer Relationship Management (CRM). Mahboubeh K, et.al [2], using RFM analysis to assess the behavior of customers in shopping. The study was conducted at health and beauty company, which then proposes marketing and sales strategy for the company. A study conducted by Zhao Dan [3], is applied RFM model and clustering K-Means algorithm to identify and assist evaluation of indigent student loan subsidies. Young Sung, et al, apply RFM analysis for a predictive recommedation in u-commerce [4]. Toly Chen, using Fuzzy method for mining sales data and build flexible customer clusters [5]. This paper discusses the use of data mining techniques using Fuzzy C-Means and Fuzzy RFM methods for customer segmentation in laundry company. Implementation of this method is done by studying customer profiles to find the appropriate class for each customer transaction patterns.

This paper is organized as follows. Section 2 decribes our study methodology. Section 3 provides a description of result . Section 4 draws a conclusion.

2. Fuzzy C-Means and Fuzzy Rfm

Customer segmentation is the division of potential customers in a given market into discrete groups. That division is based on the customer having similar enough in needs and buying characteristics.

The clustering algrorithm can be used to analyzed a characteristic of data, cluster identification and a result of monitoring data model. The model of operator data mining are built for searching the good cluster and characteristic distinctly.

Copyright © 2019 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the process of customer segmentation, as the input is a database that contains transactions customer. Data will be grouped using a Fuzzy C-Means algorithm to perform clustering process on the data prior to the process customer segmentation using Fuzzy RFM.

Fuzzy C-Means is one where data classification technique where each data point in a group (cluster) is determined by degree of membership. The Fuzzy C-Means method is a supervised clustering method where the number of cluster centers is determined in the process ot clustering. By combining the fuzzy logic C-Means with the RFM model, the results of customer segmentation can be used to give an assessment of the customer scoring and determines customer profile more precisely than the model RFM standalone [6].

The RFM method used for determine variable of measuring purchase products by customers. Variable can determine as recency, frequency and monetary [7]. RFM analysis consists of three dimensions [8]:

1. Recency, that span of time (in days, month, year) of the last transaction made by consumers today.

2. Frequency, is the total amount of the transaction or the average number of transactions during the period

3. Monetary, which is the average amount of consumer purchases value in a unit time.

Segmentation cluster in the laundry company divided by eight characterictic with RFM values of customers.

Table 4: Customer Characteristic in RFM Values

Class	Linguistic Variable Description			Label Member
	Recency	Frequency	Monetary	
C1	New	Rarely	High	Gloden A
C2	New	Rarely	Low	Occational
C3	New	Often	High	Superstar
C4	New	Often	Low	Everyday A
C5	Old	Rarely	High	Golden C
C6	Old	Rarely	Low	Dormant
C7	Old	Often	High	Golden B
C8	Old	Often	Low	Everyday B

Fuzzy RFM uses the concept of fuzzy sets and membership functions [9]. In fuzzy RFM method, an important indicator in determining the classification is the membership degree of different classes and the total value of fuzzy RFM method. Consumer segmentation process will be done by calculating the center cluster membership degree from each cluster to all classes of fuzzy RFM model using equation [9]:

$$\mu_A(x)=\left(\prod_{i=1}^{m}\mu_i(x)\right)^{(1-\gamma)}\left(1-\prod_{i=1}^{m}(1-\mu_i(x))\right)^{(\gamma)} \quad (1)$$

µA is membership degree for each class; µi is membership degree of each variable linguistic in fuzzy RFM; A is classes in fuzzy RFM model; I is variable linguistic in fuzzy RFM; X is central cluster; ɣ is gamma, usually have value 0.5.
For each membership degree that is obtained, can be determined the class of cluster, i.e., the class which has the highest degree of membership. Each class of fuzzy RFM model has consumer label that states the characteristics of each class of consumer.

2.1. Implementation

Studies conducted following the steps in the methodology CRISP-DM (Cross Industry Standard Process for Data Mining). Stages in the CRISP-DM can be seen in figure 1:

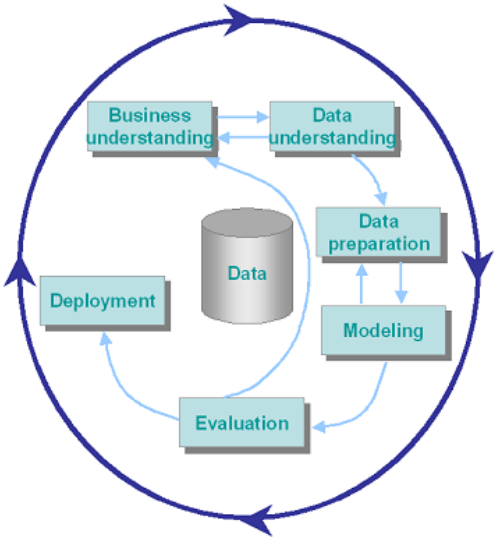


Fig 1: CRISP-DM Stages [10]

2.2. Business Understanding

Business understanding phase focus on understanding business needs based on business appraisal. This understanding aims to determine a pattern that would look for in the data mining process.
Implementation of case studies applied on laundry company. The company was built in 2009 and quite successful, have more than seven outlets in various cities in Java. With the development of increasingly large data, making the company difficult to identify customers with manual calculations. Companies need to identify their customers in connection with the loyalty to their company. Company's business objectives are as follow:

1. Increase and maintain the number of customers, especially customers who could potentially become regular customers
2. Increase corporate profits

Based on understanding of business objectives, data mining purpose is to explore the knowledge of customer's transaction pattern based on retention, frequency and monetary to determine the potential customers.

2.3. Data Understanding

This phase is understanding the needs of data related to the achievement of business goals and data mining objectives. This phase begins with the initial data collection that is done through literature, interviews and observation of processes in the laundry company. Based on business understanding, the necessary data in data mining process is the customer data and customer transaction data.

2.4. Data Preparation

The process in this phase encompass three main points:

1. Data Selection. Select the data that will be used in the process of data mining. In this phase, select the appropriate attributes. Figure 2, describe the attributes that will be used in the process of data mining. It depicted in diagram relationships.

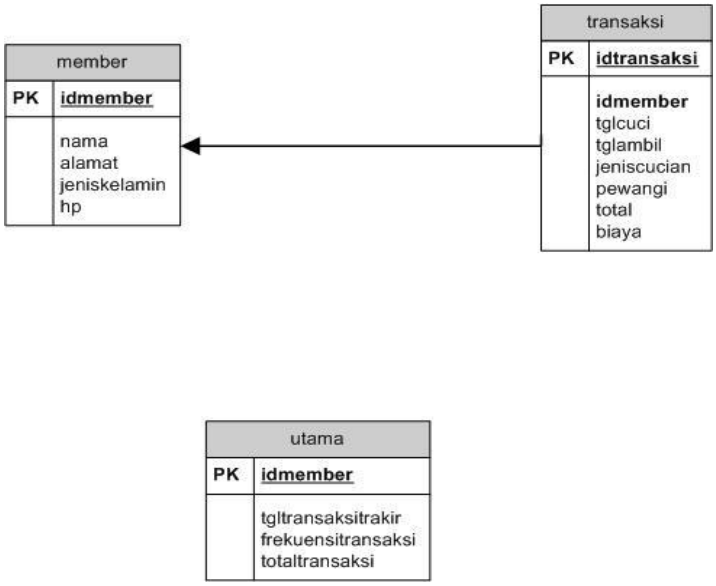


Fig 2: Laundry Relationship Scheme

Data used in the mining process is the transaction data in 2014 which has 2832 transaction records from 145 customers.

2. Data Preprocessing. This phase ensure that the selected data are free from noise data and missing values. The process of handling the missing values is assign a default value for string data and calculation the amount of the highest value for numeric data.

3. Data Transformation. In this phase, data is transformed to be used by data mining tools. The process is done by grouping the selected attributes into a single table with de-normalization. The table below shows the transformation of the data.

Table 1: Data Transformation

Initial Data	Transformed Data
IdMember	ID Member
TglTransaksiTerakhir	Indicate recency, the date of the last transaction made by member
FrekuensiTerakhir	Indicate frequency, total value of transaction carried out by member.
TotalTransaksi	Indicate monetary, amount of money according to a given transaction.

2.5. Modelling

As described previously, this study uses two approaches to conduct customer segmentation, i.e. Fuzzy C-Means (FCM) clustering algorithm and Fuzzy RFM model. Fuzzy C-Means (FCM) algorithm is used to process clustering data and RFM method for determining member classes based on recency, frequency and monetary. After getting the result of clustering (as shown in table 2), the data will be processed by RFM method.

Table 2: Calculation Center Cluster

ni1*xi1	ni1*xi2	ni1*xi3	ni2*xi1	ni2*xi2	ni2*xi3
0,0058	0,09	0,0868	0,0316	0,49	0,4724
0,0103	0,0278	0,0281	0,1652	0,4452	0,4497
0,16	0,153	0,16	0,36	0,3443	0,36
0,0619	0,64	0,5884	0,0039	0,04	0,0368
0,0671	0,0835	0,0799	0,151	0,1878	0,1798

In the RFM fuzzy model, variables recency, frequency, and monetary, each divided into two fuzzy sets:

a. Recency is divided into two fuzzy sets, namely NEW (0-25 days) and OLD (21-r days).
b. Frequency is divided into two fuzzy sets, namely RARELY (0-40 transactions) and OFTEN (30-f transactions).
c. Monetary is divided into two fuzzy sets, LOW (0-500.000 rupiah) and HIGH (350.000 rupiah – m).

Table 3: Fuzzy Rfm Domain Values

ATTRIBUTE	LINGUISTIC VARIABLE	VALUE
RECENCY	NEW	$0 \leq r < 25$ days
	OLD	$r > 21$ days
FREQUENCY	RARELY	$0 \leq f < 40$ transactions
	OFTEN	$f > 30$ transactions
MONETARY	LOW	$0 \leq m < 500.000$ rupiah
	HIGH	$m > 350.000$ rupiah

Fuzzy RFM used trapezoid graph for dispart the domain value. The graph of domain value from fuzzy RFM will show in figure 3.

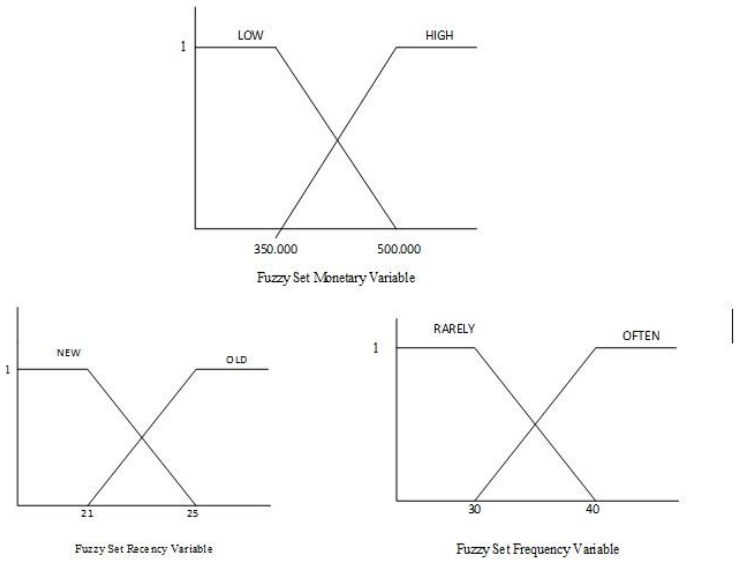


Fig 3: Chart Linguistic Variables Fuzzy RFM

The range of values given in variable retention, frequency and monetary in table 3 are used to calculate membership degree of the cluster center using equation (1). Based on membership degree that has been obtained, the class of the cluster can be determined as shown in table 4.

Table 4: Description Linguistic Variable And Label Member

Class	Description Linguistic Variable			Label Member	Score
	Recency	Frequency	Monetary		
C1	New	Rarely	High	Golden A	60 p
C2	New	Rarely	Low	Occational	20 p
C3	New	Often	High	Superstar	100 p
C4	New	Often	Low	Everyday A	40 p
C5	Old	Rarely	High	Golden C	50 p
C6	Old	Rarely	Low	Dormant	0 p
C7	Old	Often	High	Golden B	70 p
C8	Old	Often	Low	Everyday B	30 p

The next process is determining and sorting of potential customer up to less potential, namely Superstar, Golden, Every Day, Occasional and Dormant (shown in table 5) which is divided into several types of consumers based on customer activity level.

Table 5: Potential Customer Based Group

Class	Description Linguistic Variable			Label Member
	Recency	Frequency	Monetary	
C3	New	Often	High	Superstar
C1	New	Rarely	High	Golden A
C7	Old	Often	High	Golden B
C5	Old	Often	High	Golden C
C4	New	Often	Low	Everyday A
C8	Old	Often	Low	Everyday B
C2	New	Rarely	Low	Occational
C6	Old	Rarely	Low	Dormant

2.6. Evaluation and Deployment

Evaluation phase was performed to assess the achievement of business objectives from the proposed model and try to determine the business reason towards efficiency the proposed model. The next step that can do is develop a business strategy. According to the results of grouping and ranking the potential customers, different business strategy can be developing and this should be done by experts, that a well informed about the business strategy.
The deployment step begins with an evaluation and ending with a strategy to implement the results of data mining into the business. The implementation plan and disseminate the results of this study will be the next study phase.

3. Conclusion

Results of studies that have been carried out, the combination of Fuzzy C-Means and Fuzzy RFM is able to produce potential customers into eight characteristics based on factors of recency, frequency and monetary.
Grouping customers into different groups can assist decision makers in identifying market segments more clearly so as to develop more effective business strategies.

References

[1] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," ACM Comput. Surv., vol. 31, no. 3, pp. 264–323, 1999.
[2] M. Khajvand, K. Zolfaghar, S. Ashoori, and S. Alizadeh, "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study," Procedia Comput. Sci., vol. 3, pp. 57–63, 2011.
[3] D. Zhao, "Integrating RFM model and cluster for students loan subsidy valuation," 2008 Int. Semin. Bus. Inf. Manag. ISBIM 2008, vol. 2, pp. 461–464, 2009.
[4] Y. et al. Sung Cho, "Incremental Weighted Mining based on RFM Analysis for Recommending Prediction in u-commerce," Int. J. Smart Home Vol. 7 No. 6 (2013)., 2013.
[5] T. Chen and Y. Wang, "Fuzzified FCM for Mining Sales Data and Establishing Flexible Customer Clusters," Int. J. Hybrid Inf. Technol., vol. 5, no. 4, pp. 79–84, 2012.
[6] N. I. Putu, P. Yuliari, I. K. Gede, D. Putra, N. I. Kadek, and D. W. I. Rusjayanti, "Customer Segmentation Through Fuzzy C-Means and Fuzzy Rfm Method," vol. 78, no. 3, pp. 380–385, 2015.
[7] I. Chen and K. Popovich, "Understanding customer relationship management (CRM)," Bus. Process Manag. J., vol. 9, no. 5, pp. 672–688, 2003.
[8] A. Tsiptis and A. Chorianopoulus, Data Mining Techniques in CRM. United Kingdom: John Wiley and Sons, 2009.
[9] D. Zumstein, "Customer Performance Measurement: Analysis of the Benefit of a Fuzzy Classification Approach in Customer Relationship Management," Challenges, no. March, 2007.
[10] "CRISP-DM 1.0. Step by Step Data Mining Guide." 2000.
[11] T. C. H. Wu, "A Fuzzy Set Approach for Analyzing Customer RFM Data," pp. 2–5.