Journal of Theoretical and Applied Information Technology
31st August 2015. Vol.78. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CUSTOMER SEGMENTATION THROUGH FUZZY C-MEANS AND FUZZY RFM METHOD

NI PUTU PUTRI YULIARI[1],I KETUT GEDE DARMA PUTRA[2], NI KADEK DWI RUSJAYANTI[3]
Department Of Information Technology, Engineering Faculty in Udayana University
Bukit Jimbaran, Bali, Indonesia, Telp. +62361703315
E-mail: putriyuliari@gmail.com[1],Ikgdarmaputra@gmail.com[2],dwi.rusjayanti@gmail.com[3]

ABSTRACT

This research aims for finding the potential customer use data transaction. This causes, the company is difficult to arrange customers who have high and low loyalty and this research have a application for customer segmentation to help analyzing transaction data with Fuzzy C-Means for clustering and Fuzzy RFM for identify the customers. Softwares used to conduct this experiment are Microsoft SQL Server to saving the database and Matlab as the tools. The results of this segmentation for four experiments are two classes. Its has superstar I and Occasional H for each number cluster and then for the best number of cluster for this experiments are two clusters according MPC method.

Keywords: *Customer segmentation, Clustering, Fuzzy c-means, Fuzzy RFM, MPC (Modified Partition Coefficient)*

1. INTRODUCTION

The company is difficult to arrange customers who have high and low loyalty. This is caused of data transaction growth fast and limited ability to segmenting with manual computation. Amount of data from furniture company has important information to segmenting, the process of finding in a set of data called data mining[1]. Data mining is used to give services to customers based on views or insights of customers with CRM strategy. Every relation with customers can make a benefit to the company. Profitable relationship is done by analyzing data transaction of customers. Because of that marketing is important to dividing customers[2]. Volume data is continues to grow and can not be analysis with manual[1]. The application of data mining can help in analyzing customers to determine level of loyalty customer.

Effective segmentation leads to competitive advantage, recognition and exploitation of new market opportunities, selection of the appropriate target market, enhanced differentation and positioning, and increased profitability. Despite the appealing strategic and tactical benefit of market segmentation, cluster analysis remain the most favoured method.[6] The basic of idea of cluster analisys is to divide a heterogeneous customers market into homogeneous sub-grups[9]. But, some information is inevitably lost when object are grouped. Information loss is not problematic but it can result in the wrong conclusions[8]. Hence, there is no succesful segmentation without an appropriate clustering algorithm[7].

Therefore, this research have a application for customer segmentation to help analyzing transaction data in a furniture company, the application is developing method of Fuzzy C-Means and Fuzzy RFM. Software used to conduct this experiment is Microsoft SQL Server for saving database and Matlab as the tools. This application used fuzzy clustering algorithm with Fuzzy C-Means method, the algorithm have been selected because this method can make data grouped by the cluster. Fuzzy RFM (Recency, frequency, monetary) method used to choose customer with high or low loyalty from the result data of Fuzzy C-Means method. Fuzzy RFM can determine customer to the class with level loyalty their have.

2. CUSTOMER SEGMENTATION

Segmentation is process for divided customers to the some cluster with category of the loyalty customer for build the market strategy. The Characteristic of segmentation are made by bussiness rule. The clustering algorithm can be analize characteristic of data, cluster identification and result of monitoring data model. The model of operator data mining are build for searching the

Journal of Theoretical and Applied Information Technology
31st August 2015. Vol.78. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

good cluster and characteristic distinctly.

Figure 1 explaining about system process for customer segmentation, the input of this system is database from the company. Database from the company was choosed use data preparation process, data will divided to third group with Fuzzy RFM parameter. Data with paramater will be clustering with Fuzzy C-Means method and then apply MPC method for validity cluster. Output from this application is class category of the customer.

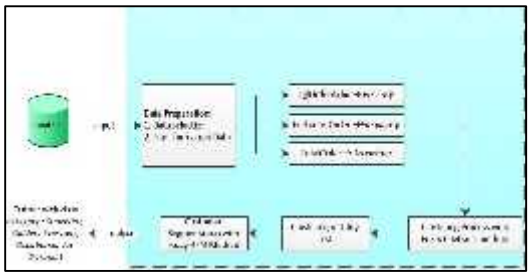

Figure 1: System Process

3. CLUSTERING

The RFM method used for determine variable of measuring purchase products by customers. Variable can determine as recency, frequency and monetary[3].

a. Recency is range time (day, month, year) from end transaction until this time by customers purchase.
b. Frequency is transaction total or transaction average in once period.
c. Monetary is average cost total of customers in once period.

Segmentation cluster in the retail company divided by six characteristic with RFM values of customers[2].

Table 1:Customer Characteristic with RFM Values.

Customer Class	Description
Superstar	a. Customers with high loyalty b. High values c. High frequency values d. Highest transaction
Golden Customer	a. Second high values b. High frequency values c. Standard transaction values
Typical Customer	a. Have standard value and transaction values
Occational Customer	a. Second of the last frequency values after dormant customer b. Lowest recency c. Highest transaction
Everyday Shopper	a. Have raising transaction b. Low transaction c. Have value with middle until low scale
Dormant Customer	a. Lowest frequency and value b. Lowest recency

Attribute distribution base on RFM will show in Table 2.

Table 2: Domain Value RFM

Atribute	Linguistic variable	Domain Value
Recency	Long Time Ago Rather Longer Recently	0 ≤ r < Max_r1 day Max_r1 day< r < Max_r2 day Max_r2 day < r
Frequency	Seldom Rather Frequent Often Very Often	0 ≤ f < Max_f1 transaction Max_f1 transaction < f < Max_f2 transaction Max_f2 transaction <f< Max_f3 transaction Max_f3 transaction < f
Monetary	Very Low Low Rather Low Rather High High Very High	0 ≤ m < Max_m1 Rupiah IDR Max_m1 < m < IDR Max_m2 IDR Max_m2 < m < IDR Max_m3 IDR Max_m3 < m < IDR Max_m4 IDR Max_m4 < m < IDR Max_m5 IDR Max_m5 < m

Fuzzy RFM used trapezoid graph for dispart the domain value. The graph of domain value from fuzzy RFM will show in figure 2.

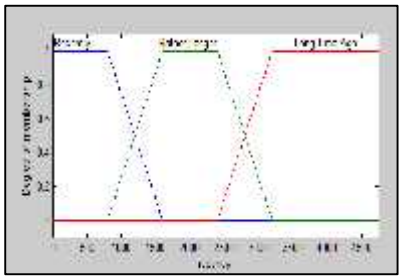

Figure 2a : Fuzzy RFM Recency

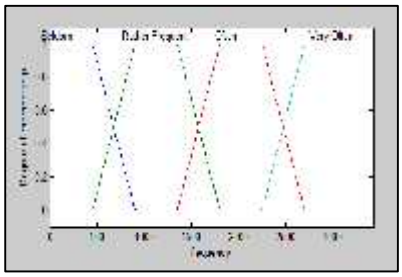

Figure 2b : Fuzzy RFM Frequency

Journal of Theoretical and Applied Information Technology
31st August 2015. Vol.78. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

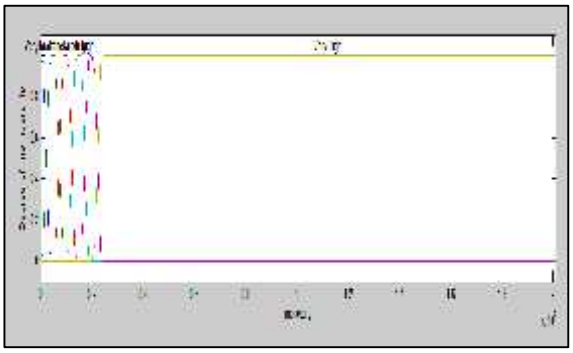

Figure 2c : Fuzzy RFM Recency

Customer segmentation process will do with computing membership degree of the centroid from every cluster with all class from fuzzy model. Its computing used equation from zumstein[11]

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

Explanation:

μ_A = Degree of membership for every class
μ_i = Degree of membership for every linguistic variable in Fuzzy RFM
A = Class in RFM Model
i = Linguistic Variable
x = Centroid
γ = *Gamma*, using value 0,5

Table 2 explain about limit of class for customer segmentation as superstar, golden, typical, occasional, everyday and dormant. The FCM approach has been applied to data clustering. Then for validation test of data using MPC method, its has for make sure the best number of clusters. The algorithm of MPC method is[10]:

$$MPC(c) = 1 - \frac{c}{c-1}(1-(c)) \quad (2)$$

The C value is the centroid and then $MPC_{(c)}$ is index value of MPC when cluster have c value.

4. RESULT AND DISSCUSION

4.1 Arsithecture Data Analysis

Arsithecture data for customer segmentation divided to 3 part which are data selection, preprocessing and transformation. Data selection is using transaction data of furniture company. Preprocessing step base on RFM method used customerID, order date and unit price from database. And the final step is transformation data to RFM.

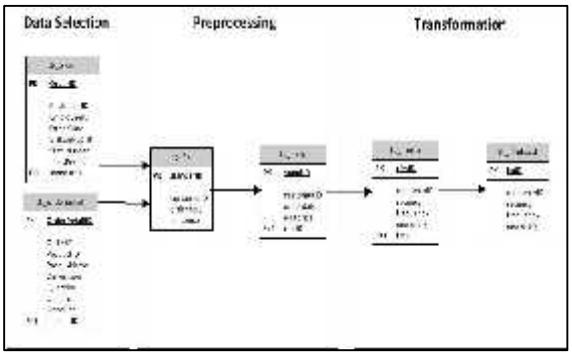

Figure 3 : Step of Architechture Data Selection

Matlab R2014b application has applied for implement FCM method. The ODBC is used for make relation between Matlab and SQL server. If Matlab and SQL server has a relation then the selection data can do based RFM method.

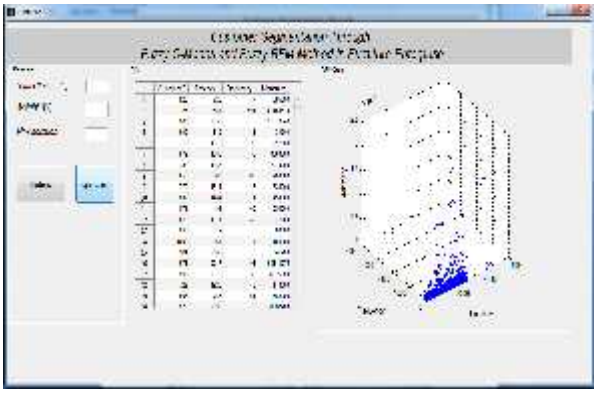

Figure 4 : RFM Data in 3D Graphic

Figure 4 explaining about data after execute to RFM and the graphic have information about dissemination data of the company. User can input total cluster, weight and maximum iteration.

4.2 Experiment

In this section, the experiment is taken to demonstrate with two until five clusters using same weight and maximal iteration, and then we implement FCM to classify the customers. The figures 5 are about dissemination of data base on total cluster.

Journal of Theoretical and Applied Information Technology
31st August 2015. Vol.78. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

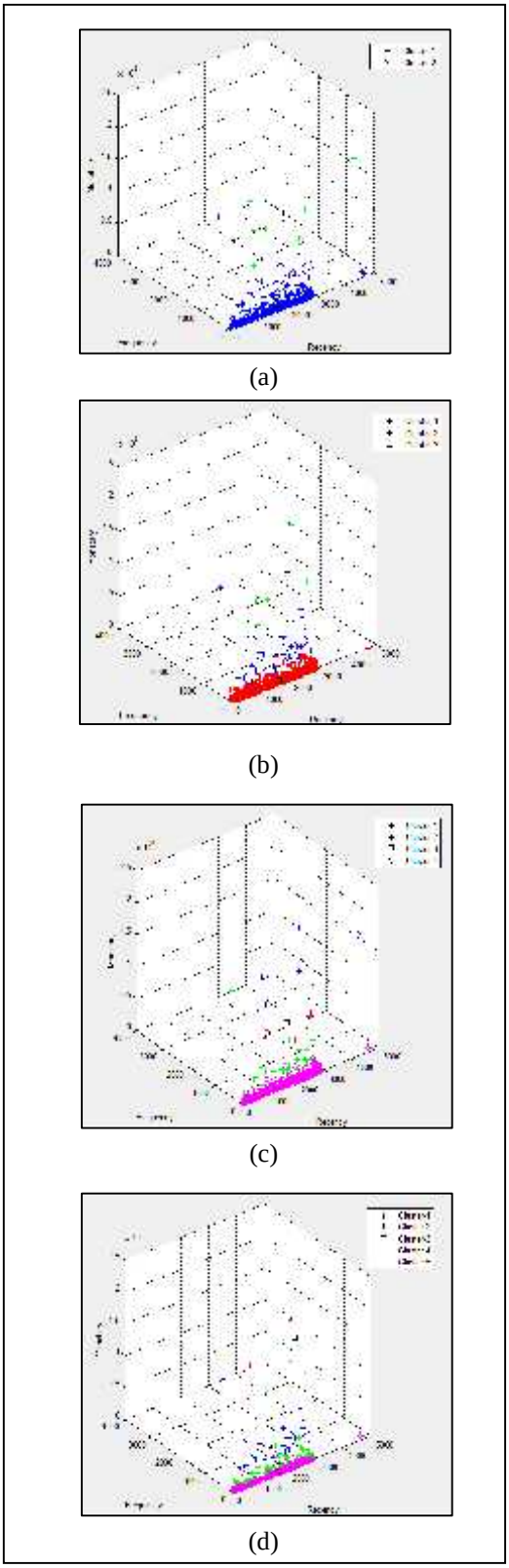

Figure 5: Dissemination Data Graph (a) 2 Clusters, (b) 3 Clusters, (d) 4 Clusters, (e) 5 Clusters

The result of the clustering process will used to find class of customer with Fuzzy RFM method, before finding the class use Fuzzy RFM. We should to implement MPC method for validition cluster to make sure the result of the clustering is right. We have the best number of cluster according MPC method is two clusters.

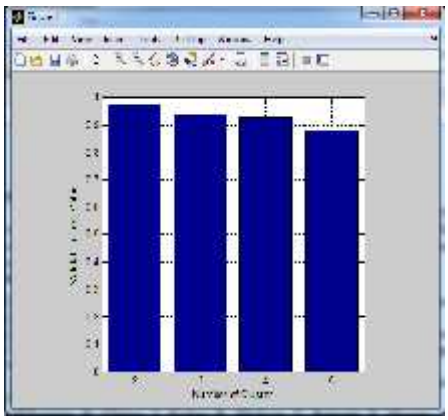

Figure 7: Cluster Comparison

The best value for number of cluster (close to one in range 0-1) with Fuzzy RFM method, it means quality of the cluster is more good for clustering data. The comparison of number of cluster with MPC Method showed in figure 7 which are 2, 3, 4 and 5 clusters and the best result is two clusters with value of validation test is closed to one. This comparison aims is for knowing number of cluster with good result for dissemination data.

An error in the usage of cluster number will cause an error in class identification and will affect the marketing strategy of a product in the company which will cause loss in company profit.

The best cluster number from MPC method is two clusters. The implement of FCM method for two clusters with 38 iterations improvement and the minimal objective value is 0.3855 will shown in the Table 3. Take customer 925 as an example, the membership of belonging to cluster 1 is 0,9999, while cluster 2 (0,0001) is lower than cluster 1. As a result, if we classify customers with according to highestvalues of membership, the customers 925, 687, 902, 710, 879, 733, 856, 779, and 802 in the same class. While customer 23 in cluster 2.

Journal of Theoretical and Applied Information Technology
31st August 2015. Vol.78. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3: The Classification Result Using FCM with 2 Number Cluster

CustomerID	Cluster 1	Cluster 2	Classification Result
925	0.9999	0.0001	1.0000
23	0.1029	0.8971	2.0000
687	0.9303	0.0697	1.0000
902	0.9994	0.0006	1.0000
710	0.9994	0.0006	1.0000
879	0.9971	0.0029	1.0000
733	0.9999	0.0001	1.0000
856	0.9996	0.0004	1.0000
779	0.9995	0.0005	1.0000
802	0.9998	0.0002	1.0000

In this case, use the value of cluster center for compared with Fuzzy RFM Method. The cluster center of two clusters shown in Table 4.

Table 4: Value of the Cluster Center

Class Code	R	F	M	Score
K26	1	1	0.9641	0.9819
K30	1	1	1	1

In the table 4, class code K30 is code for occasional H class and K26 for superstar I. Superstar customer is the customer with high loyalty, and the customer with Occasional class has low frequency. So, in this research we have two customer characters that is high loyalty (Superstar) and low frequency but highest transaction (Occasional).

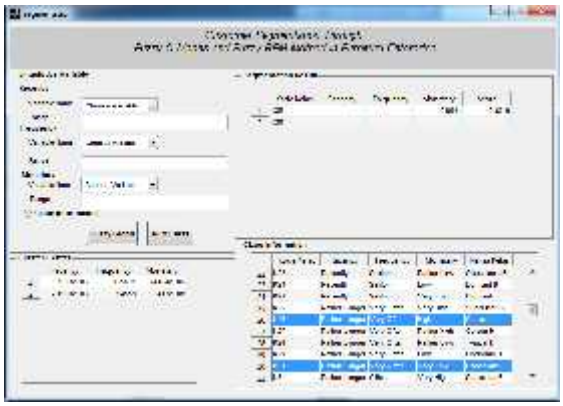

Figure 6: The Result of Customer Identified by Fuzzy RFM Method

This research has two result, that is class of the customers and the best number of clusters. the application result can knowing potential customer from class their have. Customers with included to cluster 1 will have Superstar class, and cluster 2 is Occasional class. From the result, the company can make a decision for subjected their customers. For the next experiment according Fuzzy C-Means and Fuzzy RFM method should use two clusters.

5. CONCLUSIONS

The customer segmentation application is developing method of Fuzzy C-Means and Fuzzy RFM with data transaction. Its has can build cluster with superstar customers class from comparing membership degree of the centroid with the class of Fuzzy RFM Method. The result of four experiment with same weight and maximum iteration have two class dominant which are superstar I and Occasional H. The cluster validity test with MPC (Modified Partition Coefficient) have the best number of cluster for FCM is two clusters.

6. FURTHER RESEARCH DIRECTION

The experimental results supported the usefulness of the proposed methodology. In the future, other clustering methods can be fuzzified in similar ways for the same purpose and not only focus on customer segmentation but also determine of the marketing target.

REFRENCES:

[1] Betalya. *Konsep Data Mining dan Knowledge Discovery in Database*. Universitas Gunadarma. 2009

[2] Tsiptsis Antonios Chorianopoulos. *Data Mining Techniques in CRM: Inside Customer Segmentation.* NewCaledonia: Antony Rowe Ltd, Chippenham, Wiltshire. 2009

[3] Chen. *Understanding Customer Relationship Management (Crm) People, Process And Technology*. Vol. 9 No. 5, 2003

[4] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 2000. *CRISP-DM 1.0 : Step-by-Step Data Mining Guide*. 2000

[5] Ching-hsue, Cheng. *Classifying the segmentation of customer value via RFM model and RS theory*. Expert Systems with Applications 36 (2009) 4176–4184.

[6] Dolnicar, S., Crouch, G. I., Devinney, T., Huybers, T., Louviere, J., and Oppewal, H. *Tourism and discretionary income allocation heterogeneity among households. Tourism Management*, 2009, 29(1):44–52.

[7] Dolnicar, S. *Using cluster analysis for market segmentation–typical misconceptions, established methodological weaknesses and some recommendations for improvement*.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Australasian Journal of Market Research, 2003, 11(2):5–12.

[8] Franke, N., Reisinger, H., and Hoppe, D. R*emaining within–cluster heterogeneity: a meta–analysis of the "dark–side" of clustering methods*. Journal of Marketing Management, 2009, 25(3/4):273–293.

[9] Punj, G. and Stewart, D. W. *Cluster analysis in marketing research: review and suggestions for application*. Journal of Marketing Research, 1983,20(2):138–148.

[10] Nugraheni, Yohana. *Data mining dengan metode fuzzy untuk customer relationship manangement pada perusahaan retail*. Denpasar : Universitas Udayana. 2011

[11] Zumstein, D. *Customer Performance Measurement : Analysis of the Benefit of a Fuzzy Classification Approach in Customer Relationship Management.* Switzerland : University of Fribourg. 2007