

# Fuzzy C-Means Clustering for Customer Segmentation

I Made Dhanan Pradipta<sup>1\*</sup>, Agus Eka Anwar Wahyudi<sup>2</sup>, and Sri Aryani<sup>3</sup>

<sup>1,2</sup>Departement of Electrical and Computer Engineering, Post Graduate Program, Udayana University

<sup>3</sup>Departement of Electrical and Computer Engineering, Faculty of Engineering, Udayana University

\*Email: [agus.anwar1989@gmail.com](mailto:agus.anwar1989@gmail.com)

**Abstract—** The right marketing strategy for a company can increase revenue for the company. A product will not be bought even known if the product sold is not clear its usefulness. Therefore, market segmentation is used to predict market demand based on market groupings such as Measurability (buyer characteristics and characteristics), Accesability (segmentation selected by company) and Substantiability (large segment and profitable to serve). Customer segmentation is used to determine the status of potential customers to choose the right market strategy for the company. Clustering method used is fuzzy c-means, this method is determined based on the existence of each data point by the degree of membership. Segmentation using this clustering method will generate customer labels such as superstar, golden, typical, occasional, everyday, and dormant.

**Index Terms—**Fuzzy C-means, Clustering, Segmentation, customer

## I. INTRODUCTION

Tight competition business makes many companies begin to identify the market that only produce products to be marketed in the future. This market share is used to concentrate the distribution of products of the region or places of high interest to the products. Market segmentation is one of the marketing strategies used to enable participants to form a strong foundation with respect to achieving the goals of the entrepreneur. The current market growth is considered to be a lot of changes due to changes in consumer consumption level which increased in 2011 by 4.7% higher than the average of 4.4%, Accompanied by increased community income. (Insurance Group 2013). In this case, the management company should be able to know the market that will be a prospect to sell the product [1].

Customer segmentation is a computer-based information system used to determine customer loyalty by analyzing data transactions owned. The result of this segmentation is to know customers who have high loyalty and low loyalty. Calculations in segmenting customers using the Fuzzy C-Means Clustering Method.

### A. Segmentation market

Market segmentation is a heterogeneous market dividing activity from a single product into homogeneous market units. Market segmentation is basically a strategy based on the

philosophy of marketing management whose orientation in consumers. There are four criteria that must be fulfilled market segmentation so that market segmentation process can be executed effectively and beneficial for company [9], that is

1. Measurable, meaning that the market segment can be measured, either the magnitude, or the extent and the purchasing power of that market segment.
2. Accessible, meaning that the market segment can be achieved so that it can be served effectively.
3. Large enough (Substantial), so it can be profitable if served.
4. Can be implemented (Actionable), so that all programs that have been arranged to attract and serve the market segment that can be effective.

### B. Clustering

Clustering is the process of grouping so that all members of each partition have an equation based on a matrix given. 6 Data objects located within the cluster must have a resemblance whereas those not in one cluster have no similarity. Cluster analysis or group analysis is a data analysis technique that aims to classify individuals / objects into groups that have different properties between groups, so that individuals / objects that lie in one group will have homogeneous properties. The methods in the cluster-based segmentation include iteration, K-means, fuzzy C-means, neural network, kohonen, and various other cluster techniques [2].

### C. Fuzzy C-Means

The Fuzzy C-Means method was first discovered by Jim Bezdek in 1981, the Fuzzy C-Means method is a clustering technique which is the existence of each data point is determined by the degree of membership. Fuzzy C-Means concept is the first to determine the cluster center, which is marks the average location of each cluster. In the initial conditions, the center of the cluster is not accurate, each data point has a membership degree for each cluster. By fixing the cluster center and the degree of membership of each data point repeatedly, it will be seen that the center of the cluster will move towards the right location. This loop is based on the minimization of objective function that describes the distance from the data point given to the cluster center which is weighted by the degree of membership of the data point [8]. The algorithm of Fuzzy C-Means is as follows:

TABLE II  
CLASS SEGEMENTATION

Customer Class	Description
Superstar	a. Customer with high loyalty b. High value c. High frequency d. High transaction
Golden customer	a. Second high value b. High frequency c. Average transaction
Typical customer	a. Average Value and frequency b. Average transaction
Occational customer	a. Second lowest frequency after dormant customer b. Low recency c. High expenditure
Everyday shopper	a. Enhancement in transaction b. Low expenditure c. Average to low value
Dormant customer	a. Lowest frequency b. Lowest recency

1. Determination of the cluster center that marks the average location for each cluster, with initial conditions inaccurate.
2. Each data has a membership degree for each cluster.
3. With loops based on the minimization of objective functions, the cluster center and the membership value are improved. So the location of the cluster can be in the correct position

*D. RFM Method*

RFM method function is to determine the variables to be used in measuring the level of purchasing a product by the customer. RFM model is a model to determine consumer segmentation based on recency, frequency, and monetary.

1. Recency is the time range (in units of days, months, years) of the last transaction made by the consumer until now.
2. Frequency is the total number of transactions or the average number of transactions in a period.
3. Monetary is the average amount of costs consumers spend in a unit of time.

The distribution of customers to the retail company is divided into six based on the value of RFM owned by the customer [12].

II. METHOD

The method used in determining attribute is RFM (Recency, Frequency, Monetary) method and for clustering using Fuzzy c-means method.

*A. RFM Method*

Distribution of attributes based on RFM can be found based on the following table:

TABLE I  
RFM METHOD

Attribute	Variable	Domain Value
Recency	Long	$0 \leq r < \text{Max\_r1}$ hari
	For a while	$\text{Max\_r1} < r < \text{max\_r2}$ hari
	Recently	$\text{Max\_r2} < 1$
Frequency	Rarely	$0 \leq f < \text{max\_f1}$ transaksi
	Rather often	$\text{Max\_f1} < f < \text{max\_f2}$ transaksi
	Often	$\text{Max\_f2} < f < \text{max\_f3}$ transaksi
	Very often	$\text{Max\_f3} < f$
Monetary	Very low	$0 \leq m < \text{max\_m1}$ rupiah
	Low	$\text{Max\_m1} < m < \text{max\_m2}$ rupiah
	Average	$\text{Max\_m2} < m < \text{max\_m3}$ rupiah
	Rather high	$\text{Max\_m3} < m < \text{max\_m4}$ rupiah
	High	$\text{Max\_m4} < m < \text{max\_m5}$ rupiah
	Very high	$\text{Max\_m5} < m$

From table I , it can be seen that classes that will be used in segmenting customers are superstar, golden, typical, occasional, everyday, and dormant [11].

*B. Fuzzy C-Means Clustering*

Clustering method which is used is FCM. FCM have relations with RFM method to determined attribute that is used for clustering.

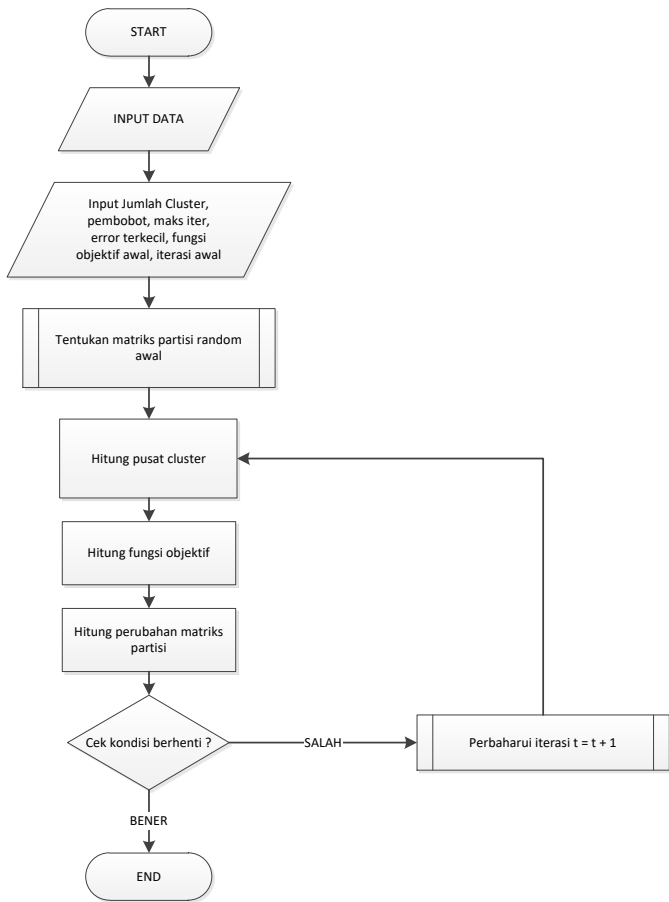


Fig. 1 Fuzzy C-means flowchart

Algorithm FCM is as follows.

1. Input data to be grouped,  $X$  is a matrix with size  $n \times m$  ( $n$  = number of data samples,  $m$  = attribute of each data).  $X_{ij}$   $i$ -th sample data ( $i = 1,2,3, \dots, n$ ),  $j$ -attribute ( $j = 1,2,3, \dots, m$ ).
2. Determine the number of clusters ( $c$ ). The rank for the partition matrix ( $w$ ), maximum iteration ( $MaxIter$ ), the smallest expected error, the initial objective function ( $P_0 = 0$ ), and the initial iteration ( $t = 1$ ).
3. Generate random numbers  $\mu_{ik}$ ,  $i = 1,2, \dots, n$ ;  $k = 1,2, \dots, c$  as elements of the initial partition matrix  $U$ .
4. Calculate the center of the  $k$ -cluster:  $V_{kj}$ , with  $k = 1,2, \dots, c$ . And  $j = 1,2, \dots, m$ , using the equation (1).

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w x_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \tag{1}$$

Information :

- $V_{kj}$  = cluster  $k$  center for attribute  $j$
- $\mu_{ik}$  = degree of membership for the sample  $i$  in *cluster*  $k$
- $X_{ij}$  = data  $i$ , attribute  $j$

1. Calculate the objective function in iteration  $t$  using the equation (2).

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \tag{2}$$

Keterangan :

- $V_{kj}$  = center of cluster  $k$  for attribute  $j$
- $\mu_{ik}$  = degree of membership for the sample  $i$  in *cluster*  $k$
- $X_{ij}$  = data  $i$ , attribute  $j$

$P_t$  = objective function in iteration  $t$

2. Calculate matrix partition changed using equation (3).

$$\mu_{ik} = \frac{\left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{1}{w-1}}}{\sum_{k=1}^c \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{1}{w-1}}} \tag{3}$$

with  $i = 1,2,3, \dots, n$ , and  $k = 1,2, \dots, c$

- $V_{kj}$  = center of cluster  $k$  for attribute  $j$
- $\mu_{ik}$  = degree of membership for the sample  $i$  in *cluster*  $k$
- $X_{ij}$  = data  $i$ , attribute  $j$

3. Check condition :

if :  $(|P_t - P_{t-1}| < \xi)$  or  $(t > MaxIter)$  then its stopped, if not:

$t = t + 1$ , loop step 4.

### III. RESULT AND DISCUSSION

System description is a figure of the whole process performed on the system and modules that will be applied in the manufacture of the application. In Figure 2 shows the process that occurs from the input process until the output process.

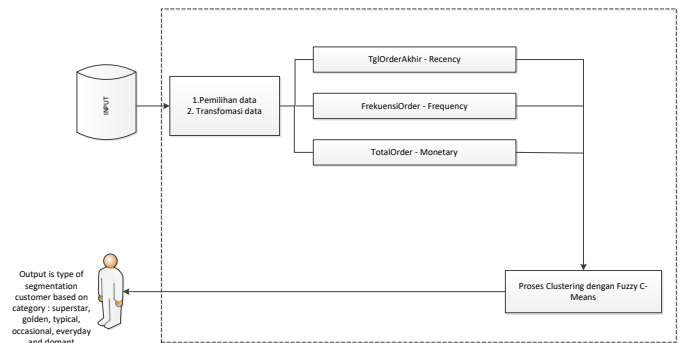


Fig. 2 System description

explained the process begins with data input and subsequent preparation of data by selecting data to be used and performed data transformation by RFM (Recency, Frequency, Monetary) method. The result of the RFM process will be clustering process using Fuzzy C-Means method. The clustering results are classified by customer segmentation based on the level of purchase of output products in the form

of customer segmentation based on the level of product purchases, namely superstar, golden, everyday, occational and dormant.

A. Case Study

An example of a calculation using Fuzzy C-Means with the input of the RFM value as an attribute.

1. The first step, perform input data in the form of tables that have been made based on Fuzzy RFM

TABLE III  
EXAMPLE DATA

Data	R(1)	F(2)	M(3)
1	5	7	20000
2	7	3	78291
3	9	1	2909
4	3	6	123211
5	1	4	23214

1. Secondly, input the number of cluster, weight, max iter, smallest error, objective function, and initial iteration.
- 2.

TABLE IV  
ATTRIBUTE DATA

Number of cluster	c	3
Weight	2	2
Max Iter	Maxiter	3
Error	E	0.01
Objective function	P0	0
First Iteration	iter	1

3. Third, rise the number of  $\mu_{ik}$  i = the number of data, k = the number of cluster. rise data use data random.

TABLE V  
DATA RISEN

I	K1	K2	K3
1	5	14	2
2	9	17	15
3	14	15	22
4	3	4	2
5	12	6	12

3. Fourth, calculate the center of cluster using equation (1).

TABLE VI  
CENTER OF CLUSTER

i	$\mu_{ik}$			Xij			Ui1	Ui2	Ui3
	K1	K2	K3	R(1)	F(2)	M(3)	$\wedge_w$	$\wedge_w$	$\wedge_w$
1	5	14	2	5	7	20000	25	196	4
2	9	17	15	7	3	78291	81	289	225
3	14	15	22	9	1	2909	196	225	484
4	3	4	2	3	6	123211	9	16	4
5	12	6	12	1	4	23214	144	36	144
							455	792	861

Result of the center cluster calculated

TABLE VI  
CENTER OF CLUSTER USING RFM

Vkj	R1	F2	M3
1	5.8923	6.7354	26073.5164
2	2.7340	3.4141	37888.5151
3	13.4219	2.0754	26558.7584

4. Fifth, calculate the objective function using equation (2)

TABLE VII  
CLUSTER 1

i	cluster 1			Tot1
	$(X_{ij} - V_{ij})^2$			
1	0.79	3.22	36887601.46	36887605.47
2	34.71	14.40	2726665593.51	2726665642.62
3	23.93	33.58	536594820.04	536594877.55
4	8.36	0.63	9435690720.14	9435690729
5	154.30	7.81	8176834.04	8176996.15

5. Seventh, check the stopped condition. In this step, the condition is measured by equation  $(|Pt - Pt-1| < \xi)$  atau  $(t > MaxIter)$ .

TABLE VIII  
RESULT OF CLUSTERING

i	K1	K2	K3
1	0.32	0.28	0.38
2	0.38	0.23	0.38
3	0.23	0.52	0.24
4	0.36	0.27	0.35
5	0.03	0.91	0.04

the result already qualified so there is no need for repetition to looping step. The iteration is performed only once because the result of the first iteration has exceeded the maximum previously known iteration. So, the result of this clustering can be seen in table 9.

TABLE IX  
RESULT OF CLUSTERING

I	K1	K2	K3
1			X
2	X		
3		X	
4	X		
5		X	

The result of clustering can be known by doing the comparison and taken from the most great value. The data with the largest value of the comparison result then the data included into one cluster such as data 1 included into cluster number 3, data-2 in cluster 1, data-3 in cluster 2, data-4 in cluster 1, and data-5 on cluster 2.

IV. CONCLUSION

Market segmentation is customer segmentation based on customer loyalty. Basically, consumers in choosing or buying something in accordance with their respective behavior. Purchase decisions made by consumers is based on the needs and properties owned. Therefore, customer segmentation is needed to treat customers appropriately by differentiating

customers with high loyalty and low loyalty. The method used in customer segmentation is Fuzzy C-Means and for the method of determining the attributes used is using Fuzzy RFM (Recency, Frequency, Monetary) method.

REFERENCES

[1] Betalya. 2009. *Konsep Data Mining dan Knowledge Discovery in Database*. Universitas Gunadarma. Diunduh : 10 Oktober 2014.

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

[3] 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*. Tersedia di [http://www.community.udayton.edu/provost/it/training/documents/SPSS\\_CRI](http://www.community.udayton.edu/provost/it/training/documents/SPSS_CRI) SPWPlr.pdf. Diunduh : 10 Oktober 2014.

[4] Erick, Lisangan. 2010. *Perancangan Data Warehouse Pengolahan Persediaan Buku PT. Gramedia Asri Media Makasar*. Bandung : Politeknik Telkom.

[5] Kusuma Wijaya, Alvian. 2014. *Implementasi Data Mining dengan Algoritma Fuzzy C-Means Studi Kasus Penjualan di UD Subur Baru*. Semarang : Fasilkom Udinus

[6] Kristanto, Titus. 2013. *Analisa Data Mining Metode Fuzzy Untuk Customer Relationship Management Pada Perusahaan Tour & Travel*. Surabaya : Institut Teknologi Surabaya.

[7] Mishra. 2009. *Customer Relationship Management: Implementation Process Perspective*. Turkey : Department of Computer Engineering Atilim University.

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

[9] Nurbaity Lubis, Arlina. 2004. *Strategi Pemasaran Dalam persaingan Bisnis*. Sumatra Utara : Universitas Sumatra Utara

[10] Project, Cipsed. 2011. *Teknik Untuk Melakukan Analisa Pasar*. Gorontalo: State University of Gorontalo [UNG] Entrepreneurship ToT Program

[11] Santosa, Budi. 2007. *Segmentasi Pasar dengan Ant K-Means Clustering*. Surabaya : Institut Teknologi Sepuluh November.

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