

Implementation of Data Mining Technique for Customer Relationship Management (CRM) on Online Shop Tokodiapers.com With Fuzzy C-Means Clustering

Lisna Zahrotun

Department of Informatics, Faculty of Industrial Technology
Ahmad Dahlan University
Yogyakarta, Indonesia
Lisna.zahrotun@tif.uad.ac.id

Abstract - Online business competition is very tight. One strategy that can be done is to retain customers. By applying the concept of CRM (Customer Relationship Management), the online shop can identify customers, so customers can get special facilities in the appropriate marketing strategies according to their needs. This research discusses about how data mining process from customer data at shop online tokodiapers.com. The process of data mining begins with clustering process using Fuzzy C-Means (FCM) method which is then tested by clustering using purity test. The result of this research is a web-based data mining application that is able to segment customer data on tokodiapers.com using fuzzy C-Means clustering method. The test is done using purity test showed that the best grouping is found on the grouping with 4 clusters with the value of purity test 0.79.

Keywords: Data mining, Fuzzy C-Means, Clustering

I. INTRODUCTION

The problem that often arises in the company is the level of customer loyalty was undetected. This is due to the rapid growth of data transactions and the limited ability to segment with manual calculations. Customer segmentation is defined as "The process of dividing the customer into different, meaningful, and homogeneous subgroups based on various attributes and characteristics"[1]. Customer segmentation was once used to assist the bank in dividing the market into client groups based on different needs, characteristics or behaviors which may require a separate product or marketing mix[2]. Customer segmentation is also used for customer management and customer supply chain[3]. Efficiency of customer relationship management using fuzzy query method has been done to improve customer information [4]. Grouping using the Fuzzy C-Means method is done to divide the customers into two groups using the LRFM model variable (length, frequency, frequency, monetary value) [5]. Implementation of customer relationship management as a business solution [6].

Online shop tokodiapers.com is one of the shops that deal in baby supplies product. This online sale has been running for five years. In addition to online sales

tokodiapers.com also has an offline store. Every day this online shop serves customers more than 100 customers, either just asks or want to buy at the online store. To serve the customers, the admin system of onlineshop tokodiapers.com using online web, BBM, Whatsapp and SMS. This will certainly take a long time. In addition, marketing still think that all customers needs are the same, this makes customers who do not need the products feel annoyed with the bids that always flooded.

Large amounts of data require a search in a set of data that called data mining [7]. Stacked customer data can be utilized to prepare for customer's needs that are important factors to gain advantage in providing products and services them. Managers should prioritize their customers and focus on the main customers and think that when they can not provide good service then they will lose the customers. This is important because if customers switch to another company it will impact on the company's reputation, loss and credibility of the company [5]. One method of data mining is Fuzzy C-Means (FCM). FCM has a high degree of accuracy and fast computing time [8].

By using data mining from customer data at tokodiapers.com online shop hence the customer identification can be done using Customer Relationship Management (CRM). By applying the CRM concept online shop can identify potential customers by segmenting customers. The purpose of the customer segmentation process is to know the customers behavior and implement the right marketing strategy so as to bring profits for the company. The role of customer segmentation is very important. So to answer the problem of customer service and marketing to customers that fit the target shop online tokodiapers.com then "Implementation of Data Mining Techniques for Customer Relationship Management (CRM) On Tokodiapers.com Online Shop With Fuzzy C-Means Clustering Method" is made.

II. METHODS

2.1 Data Research

The research data used in this study is customer data from diapers.com stores as many as 132 customer data.

2.2 Data Mining

Data mining is the process of searching for patterns of interesting and hidden (hidden pattern) of a large data set stored in a database, data warehouse, or other data storage [9]. Data mining, often referred to as knowledge discovery in database (KDD). Phase data mining stage is shown in Figure 1.

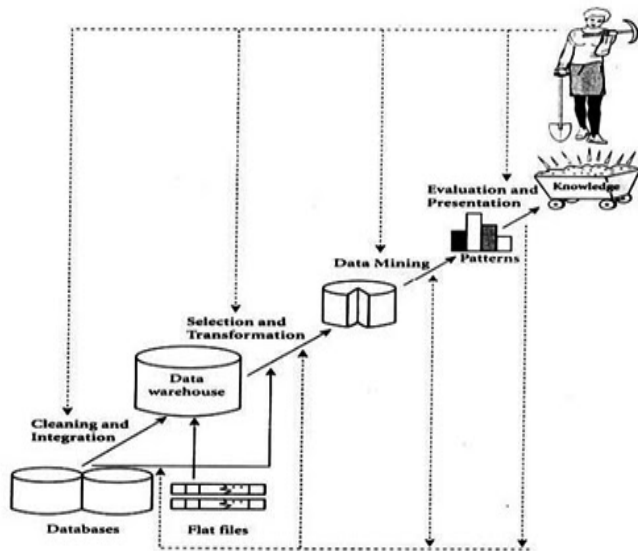


Figure 1. Steps of Knowledge Discovery in Database[10]

2.3 Data Mining in CRM

CRM has four dimensions, namely [11]:

1. Customer identification (customer identification).
2. Building attraction to customers (customer attraction).
3. Maintaining existing customers (customer retention).
4. Develop customer (customer development).

The four dimensions of CRM can be seen as closed cycle in Customer Management System. Each dimension has the same common goal that is to build a deeper understanding of the customer and increase customer value. In relation to the achievement of these objectives, data mining techniques can be used to discover the characteristics and behavior of hidden customers in the database. The data mining engineering framework in the four dimensions of CRM is described in Figure 2.

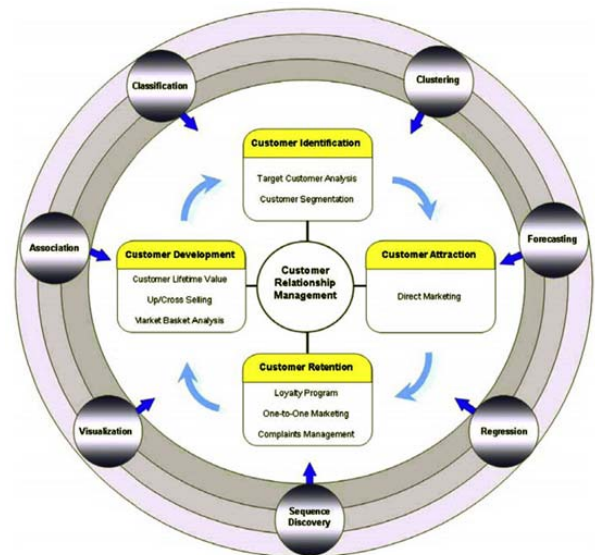


Figure 2. The Data Mining Engineering Framework in CRM[12]

2.4 Fuzzy C-Means Clustering Method

Fuzzy C-Means is one of the data grouping techniques in which the existence of each data point in a cluster is determined by the degree of membership. The Fuzzy C-Means method includes a supervised clustering method in which the number of cluster centers is determined in the clustering process. The algorithm of fuzzy c-means is as follows [13]:

1. Input data to be grouped, i.e X, a matrix of size $n \times m$ (n = number of data samples, m = attribute of each data). X_{ij} the i -th sample data ($i = 1, 2, \dots, n$), the j -attribute ($j = 1, 2, \dots, m$).
2. Determine the number of clusters (c), the rank for the partition matrix (w), maximum iteration (\maxIter), the smallest expected error (ξ), the initial objective function ($P_0 = 0$), and the initial iteration ($t = 1$).
3. Generate random numbers η_{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, with $k = 1, 2, \dots, c$; and $j = 1, 2, \dots, m$, using the following equation

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

with:

V_{kj} = the center of the k -cluster for the j -attribute

μ_{ik} = degree of membership for the i -th sample data in the k -cluster

x_{ij} = i -data, j -attribute

5. Compute the objective function on the t iteration using the following equation:

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

with:

- V_{kj} = the center of the k-cluster for the j-attribute
- μ_{ik} = degree of membership for the i-th sample data in the k-cluster
- X_{ij} = i-data, j-attribute
- P_t = objective function on the t iteration

6. Calculate the partition matrix change using the following equation:

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

With $i = 1, 2, \dots, n$; and $k = 1, 2, \dots, c$.

Where :

- V_{kj} = the center of the k-cluster for the j-attribute
- μ_{ik} = degree of membership for the i-th sample data in the k-cluster
- x_{ij} = i-data, j-attribute

7. Check Stop condition:

If $(|P_t - P_{t-1}| < \epsilon)$ or $(t > Maxiter)$ then stop

If not: $t = t + 1$, repeat step 4.]

2.3 Fuzzy FRM Method

The RFM analysis consists of three dimensions, [14]

1. Recency, i.e time range (in units of days, months, years) of the last transaction made by the customer to date.
2. Frequency, i.e the total number of transactions or the average number of transactions in a period.
3. Monetary, which is the average amount of customer's purchase value in a time unit

2.4 Purity Test

Purity test is done to know good of bad cluster that resulted in clustering process. Here's the formula used in purity test:

$$Purity(\Omega, C) = \frac{1}{N} \sum_k \max_j |\omega_k \cap c_j| \quad (4)$$

Where :

$\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$ is the cluster set $\omega_k \rightarrow$ the data set in ω_k

$C = \{c_1, c_2, \dots, c_j\}$ is the set of classes $c_j \rightarrow$ the data set in c_j

Clustering is bad if the purity value is close to 0, and good if the purity value is close to 1.

III. RESULT

In this study, data obtained from tokodiapers.com processed through the data mining stage:

3.1 Data Cleanup

It is the stage to clean up data that contains noise, data that has a blank value and inconsistent data to be removed. The original 132 initial data after the process of cleaning the data into 108 data.

3.2 Data Selection

In the selection process this data is done the selection of attributes of the data that existed before. The result of the selection process is to choose from 7 attributes that is no, date, name, address, no phone, item name and price become 4 attributes namely date, name, address, and price.

3.3 Data Transformation

Transformation aims to present data that is ready to do the mining process. In this study the data will be divided into three namely

- a. Regency is used to find out how long consumers have become customers, which in this study used in units of years.
- b. Frequency is used to find out how often consumers make transactions in one month,
- c. Monetary is used to know the average amount of purchase price made by the consumer every month.

3.4 Data Mining Process

Data mining process is made using php my admin. The results of the C-Means fuzzy data mining process for the results of calculating equations 1 of the cluster center search are shown in Figure 3, the result of the equation 2 of the objectivity function shown in Figure 4, the calculation result of Equation 3 of the matrix change shown in Figure 5 and the cluster results shown in Figure 6

Figure 3. Cluster Center calculation results

Figure 4. Result of Objectivity Function Calculation

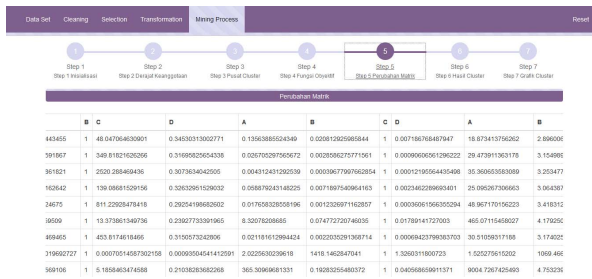


Figure 5. Matrix Calculation Results

The figure shows a software interface titled 'Cluster Yang Terbentuk'. It displays four clusters of data points. Each cluster is represented by a table with columns: No, Tanggal, Nama, Alamat, No. Telp, Nama Barang, and Harga. Cluster 1 has 7 data points, Cluster 2 has 7 data points, Cluster 3 has 3 data points, and Cluster 4 has 3 data points.

Figure 6. Clustering results using fuzzy C-Means

In the application that has been made, three experiments with different parameters has been done. The results of the experiment are as follows:

1. Experiment 1

In experiment 1 produced 2 cluster centers with regency, frequency and monetary values shown in Table 1.

Table 1. Experiment Results with 2 clusters

Cluster	Regency	Frecuency	Monetary	Data Total
1	1.53	1.80	456.658	12
2	2.33	1.08	109.163	96

With the amount of data on cluster 1 and cluster 2 shown in Figure 7

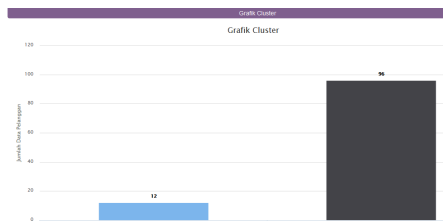


Figure 7. The Display of Graph 2 Cluster

2. Experiment 2

In experiment 2 produced 3 cluster centers with regency, frequency and monetary values shown in Table 2

Table 2. Experiment Results with 3 clusters

Cluster	Regency	Frecuency	Monetary	Data Total
1	1.41	2.13	513.397	28
2	2.57	1.00	86.701	9
3	1.76	1.37	186.904	71

With the data total on cluster 1 and cluster 2 in Figure 8

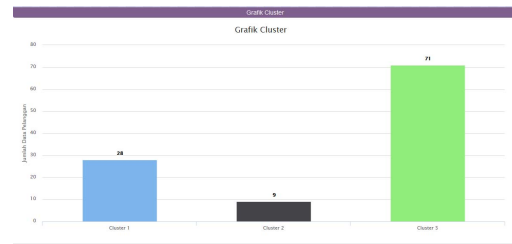


Figure 8. Graph 3 Cluster View

3. Experiment 3

In experiment 3 produced 4 cluster centers with regency, frequency and monetary values shown in Table 3

Table 3. Experiment Results with 4 clusters

Cluster	Regency	Frecuency	Monetary	Data Total
1	1.74	1.43	357.149	7
2	2.60	1.00	80.618	64
3	1.43	2.47	583.929	5
4	1.77	1.24	160.581	32

With the amount of data on cluster 1 and cluster 2 shown in Figure 9



Figure 9. Graph 3 Cluster

3.5 Pattern Evaluation

The pattern evaluation process is done by using purity test. The results of the test are shown in Table 4.

Table 4 Purity Testing Results

N o.	Total Cluster	Weig hting rank	Max Iteration	Smal lest Error	Object ivity Functi on	Initi al iterat ion	Pur ity
1.	2	2	50	0.001	0	1	0.71
2.	3	2	100	0.005	0	1	0.45
3.	4	2	70	0.005	0	1	0.79
Average							0.65

Purity test that has been done using the data set of 108 customer data by combining the parameters of the number of clusters, weighting rank, max iteration, smallest error, objectivity function and initial iteration obtained the result that the best purity value of 0.79 from the third experiment is to produce 4 clusters .

3.6 Knowledge Representation

From purity test result obtained 4 best cluster with character of customer:

1. The first cluster obtained 7 customer data with the characteristics of the customer has an average purchase of

- 357,149 per month is long enough to be a customer that is 1.74 years and has a frequency of buying at least 1 time per month purchases.
2. The second cluster obtained 64 customer data with the characteristics of customers have an average purchase of 80,618 per month is long enough to be a customer that is 2.6 years and has a frequency of buying at least 1 time per month.
 3. The third cluster obtained 5 customer data with the characteristics of the customer has an average purchase of 583,929 per month is long enough to be a customer that is 1.43 years and has a frequency of buying at least 2 times the purchase each month.
 4. The fourth cluster obtained 32 customer data with the characteristics of the customer has an average purchase of 160,581 per month is long enough to be a customer that is 1.77 years and has a frequency of buying at least 1 time per month purchases.

IV CONCLUSION

From the research that has been done can be concluded:

1. A web-based data mining implementation has been produced that is capable of segmenting consumers with the C-Means fuzzy clustering method.
2. Purity test that has been done using a data set of 108 customer data by combining the parameters of the number of clusters, weighted rank, max iteration, smallest error, objectivity function and initial iteration obtained the result that the best purity value of 0.79 from the third experiment is to produce 4 clusters.

ACKNOWLEDGEMENT

This research has been supported by the research grant of RISTEK DIKTI with the Research Scheme of Beginner Lecturer (PDP)

REFERENCES

- [1] K. W. Choung, J. T. Kwok, M. H. Law, and K. C. Tsui, *Mining Customer Product Ratings For Personalized Marketing Decision Support System*. 2003, pp. 231–243.
- [2] Maghfirah and Te. B. Adji, “Appropriate Data mining Technique and Algorithm for Using in Analysis of Customer Relationship Management (CRM) in Bank Industry,” in *Seminar Nasional Aplikasi Teknologi Informatika (SNATi)*, 2015, pp. 7–10.
- [3] L. Bryant, “Customer Knowledge Management In Marketing and Supply Chain Management by Looking at Data Mining,” vol. 21, no. 1, pp. 1–11, 2009.
- [4] C. Chiu, H. Ku, I. Kuo, and P. Shih, “Customer information system using fuzzy query and cluster analysis,” *J. Ind. Prod. Eng.*, vol. 31, no. 3, pp. 134–145, 2014.
- [5] A. Ansari and A. Riasi, “Customer Clustering Using a Combination of Fuzzy C-Means and Genetic Algorithms,” *Int. J. Bus. Manag.*, vol. 11, no. 7, pp. 59–66, 2016.
- [6] K. Aiaots, “Implementing Customer Relationship Management in Hansa busines Solutions,” 2001.
- [7] N. P. P. Yuliani, I. K. G. D. Putra, and N. K. D. Rusjayanti, “Customer Segmentation Through Fuzzy C-Means and Fuzzy RFM Method,” *J. Theor. Appl. Inf. Technol.*, vol. 78, no. 3, pp. 380–385, 2015.
- [8] K. Hammouda and F. Prof. Karaay, “Comparative Study of Data Clustering Techniques,” 2000.
- [9] P. A. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. 2006.
- [10] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kauffman, 2006.
- [11] R. . Swift, *Accelerating Customer Relationship - using CRM and Relationship Tecnologies*. Prentice Hall inc, 2000.
- [12] E. W. T. Ngai, D. C. . Li Chiu, and Chau, *Expert System with Applications*. 2009, pp. 2592–2602.
- [13] J. Yan and J. Power, *Using Fuzzy Logic (ToWard Intelligent System)*. New York: Prentice Hall inc, 1994.
- [14] Chen, “Understanding Customer Relationship Management (CRM) People, Process And Technology,” vol. 9, p. 5, 2003.