CUSTOMERS CLUSTERING BASED ON RFM SCORE USING GENETIC ALGORITHM

Muhammad Ridwan Andi Purnomo[1], Nur Riana Fajarwati[2]

Department of Industrial Engineering, Faculty of Industrial Technology,
Islamic University of Indonesia, Yogyakarta, Indonesia.
[1]ridwan_ie@uii.ac.id, [2]riana.inung@gmail.com

ABSTRACT

Today, companies should select the right marketing strategy for their customers for being competitive in turbulent environment. It is required for keeping existing potential customers as the main profits resource. There are several methods usually used to analyse customers loyalty. Most of conventional methods usually just consider frequency of product consumption to analyse customers characteristics. However, such methods have lack in analysis of money spent by customers and time between arrival of customers. Actually, customers who came frequently in a period of time and customers who came rarely but spent so much money could not be neglected in the customers characteristic analysis. RFM considers these 3 factors, consumption interval (R), frequency (F) and money amount (M) in analysing customers loyalty. In this research, raw data about customers transaction would be analysed using RFM technique. Further, clustering technique would be applied in order to group customers who have similar characteristics. The clustering was conducted using Genetic Algorithm (GA), and further it called as GA-Clustering. The clustering analysis is supervised clustering and the number of cluster was determined based on number of CRM strategy that have been determined in advance through discussion with firm's owner. The result of this research is customers cluster based on RFM value and suitable CRM strategy that must be applied to every customers cluster.

Keywords*: customers loyalty, customers relationship management, RFM, clustering, genetic algorithm*

1. INTRODUCTION

In dynamic and competitive business, a company must have innovative marketing strategy includes understanding of customer needs, improve customer satisfaction and customer retention. The fulfilment of particular customer requirements is one of key factors for the success business. In other words, selection of right marketing strategy for particular customers must be the main concern to the company in dynamic competition. It is important for companies to identify potential customers that have high value in dynamic and global competition (Chang et al., 2007). The potential customers must be maintained by understanding their characteristics and defining suitable marketing strategy so that they can be loyal customers. Customer Relationship Management (CRM) is a paradigm to strengthen relationship between company and customers. CRM is everything that is related to satisfaction of customer's needs. CRM helps the company for better distinguish, more efficient allocation of resources to group of customers that have the highest profitability (Soeini and Fathalizade, 2012). Moreover, in order to achieve better customer retention and profitability, the company needs to customise strategies and fulfil different customer needs by allocating resources effectively and efficiently (Huang et al., 2009).

There are several techniques to analyse customers characteristic. One of the recent technique that start being used widely is Recency, Frequency and Monetary (RFM). Such model is a characteristic-based model used to analyse the behaviour of customers and then make prediction (Hughes, 1996; Yeh, 2008). RFM model has been widely applied in marketing database and it is a common tool to develop marketing strategies (Wei, et al., 2010). RFM is easy to

use and generally could be implemented quickly. Three variables in RFM are simple, these variables could be easily calculated based on customer transactions data that is stored in database. The output of RFM model could be processed further, such as using clustering technique, in order to determine suitable marketing strategies. Several clustering techniques are available in the literature, and one of the well known technique is k-means algorithm. However, the k-means algorithm needs initial solution that is generated randomly, hence, the quality of the result would be depended to the initial solution as well. To overcome such problem, several meta optimisation techniques have been proposed by previous researchers (Liu et al., 2011).

Actually, meta optimisation technique have potentiality to solve complicated case such as clustering. When it used in clustering case, the meta optimisation technique could find cluster centres and at the same time could find members of each clusters. Hence, the meta optimisation technique doesn't need to be embedded in a clustering technique. It could be used to conduct clustering directly. In this research, Genetic Algorithm (GA) as one of well known meta optimisation technique, would be used as clustering technique after RFM analysis. GA uses random process and stochastic transition in searching the solution. Beside, GA uses an agent population instead of single agent and it would increase the possibility of the GA to find global optimum solution.

2. RELATED WORKS

Several researches about customers clustering to identify the characteristic of customers have been investigated by previous researchers. Weng and Liu (2004) applied modified clustering technique to build an online personalised recommendation system based on customer profile as well as product profile. Validation of the proposed system is conducted through experiment. When a user use the system and the user doesn't click the recommended product by the system, then the system will adjust the grouping process automatically. The proposed system could make correction by itself so that the system is stable.

McCarty and Hastak (2007) investigated RFM, Chi-Square Automatic Interaction Detection (CHAID), and logistic regression as analytical methods for direct marketing segmentation. It is found that CHAID tends to be superior than RFM when the response rate to a mail is low and the mail would be a relatively small portion of the database. However, RFM is an acceptable procedure in other circumstances. RFM may focus too much attention on transaction information and ignore individual information like values, lifestyles, and so on that may help a company to market the product to their customers better.

Lumsden et al. (2008) applied RFM model to distinguish customer value according to pre-purchase motivations of membership initiation in all-inclusive travel vocation club in America. In such study, Recency is defined as the number of periods in year since the last visit. Frequency is defined as the number of visits to clubs on holiday, while Monetary defined as the average expenditure of customer in a time period.

Cheng and Chen (2009) used RFM model and Rough Set theory to classify the segmentation of customer value in C-company dataset in Taiwan's electronic industry. This study firstly utilises RFM model to yield quantitative values as input attributes. Then k-means algorithm is used to cluster customer value. Finally, it employed rough set to mine classification rules that help company in determining an excellent CRM.

Study about the use of RFM combined with customer grouping has been conducted by Wei et al. (2011). In such study RFM score of customers would be clustered using self-organising map neural network to segment dental patients of a children's dental clinics in Taiwan. That study is just to identify customer's profile of dental clinics and there is no CRM strategies proposed to every cluster formed.

From review above, it could be seen that RFM has received major attention from researchers to analyse customers characteristics. However, it would be useful if the analysis could be continued to

ISSN : 1978-774X

customer clustering in order to determine appropriate strategy for CRM. GA, as optimisation tool, has potentiality to be used as clustering method. Such idea would be investigated in this research.

3. THE METHOD

Case study of this research took place in a Day and spa Salon in Jogjakarta, Indonesia. Customer characteristic would be analysed based on RFM score and the analysis would be continued to clustering using GA so that suitable marketing strategy for each customers cluster could be implemented well. In this research, GA-clustering is stand alone technique. It means that the GA would not be embedded in a clustering algorithm. Following sub sections explain in detail about RFM model and GA-clustering development.

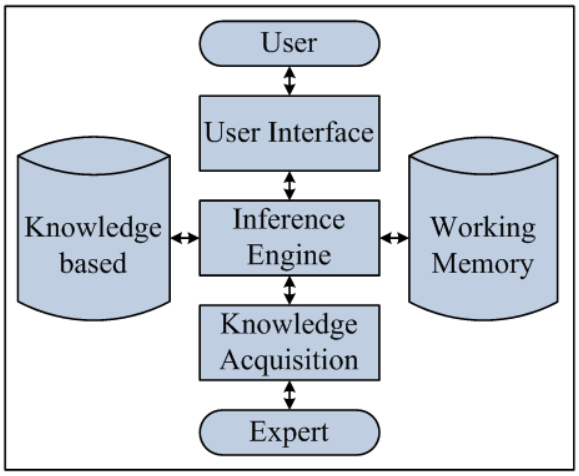

Figure 1. General structure of ESCAPP

3.1. RFM Development

For RFM model development, data about customer transactions was used, which start from year 2009 until 2013. The data was taken from the firm's database. It means, the data was come from registered customers only, which was around 215 customers. Unregistered customers is neglected since the data is not well documented and impossible to be analysed. The data is about time arrival and amount of money spent at that time. In order to get valid analysis, because the data is about amount of money, hence the data must be converted to future value. The conversion is conducted by considering bank rate, in this case is Bank Indonesia rate, in February, 2013 which is 5.75%.

3.1.1. Recency

Recency refers to interval between the time that the latest consuming behavior happens and present. The shorter the interval, the bigger the Recency value. For Recency value, customers will be scored by particular symbols based on quintiles concept. It is conducted by sorting the customers according to Recency of customers. The top 20% of customers would be scored with 5, while the next 20% customers would be scored with 4, and so forth. In this research, the measurement of the recent transaction period is monthly period.

3.1.2. Frequency

Frequency refers to the number of transactions in a particular period. For example, two times of one year, two times of one month. The many the frequency, the bigger Frequency value. Higher frequency score indicates greater customer loyalty. A customer having a high score of frequency implies that customer has great demand for the product and is more likely to purchase repeatedly. Scoring for Frequency was conducted by similar way with scoring for Recency.

3.1.3. Monetary

Monetary refers to consumption money amount in a particular period or by the average money amount per purchase. Marcus (1998) suggested that it is better to use the average purchase amount rather than total accumulated purchase amount. The much money amount, the bigger Monetary value. Same with Recency and Frequency scoring, customers were scored by the average of total purchase amount. Customers who have lowest monetary is scored with 1 while customers who have highest monetary is scored with 5. Result of RFM analysis is shown in Table 1.

Table 1. RFM model result

No	Customer No	R	F	M
1	1	4	5	2
2	3	5	5	3
3	4	4	5	4
4	5	1	3	4
5	6	1	5	1
.	.	.	.	.
.	.	.	.	.
215	218	5	2	3

3.2. GA Development

GA is an optimisation algorithm. GA would be used to cluster the data above, hence, clustering problem must be viewed as optimisation problem. The objective of the optimisation problem is minimising the distance between each data to every cluster centre that would be selected before. Therefore, the decision variable to be found by the GA is the cluster centres that could minimise the total distance each data to the nearest cluster centre. The number of cluster to be formed is same with the number of CRM strategy that would be applied to the customers. There are 5 CRM strategies that have been determined in advance through discussion with the owner of the firm namely initiative information, positive impression, contact initiative, behaviour forming, and mind share to customer. Since the number of cluster has been determined in advance, hence this clustering problem is supervised clustering problem.

3.2.1. Chromosome Representation

In this GA-clustering, chromosome is individual in a population, representing the solution of the problem. In this research, the solution to be found is cluster centres, hence, the chromosome was constructed using real number which represents the data that would be cluster centre. Since there are 5 clusters to be formed, hence a chromosome would have 5 locus. Figure 1 shows example of a chromosome.

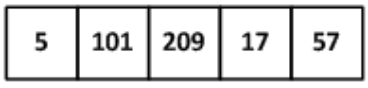

Figure 1. Chromosome Representation

Based on Figure 1 above, then data number 5, 101, 209, 17 and 57 would be cluster centre for cluster number 1, 2, 3, 4 and 5 respectively.

3.2.2. Fitness Function

The objective of this GA-clustering is to minimise the total absolute distance between each data to nearest cluster centre. Equation 1 shows formula of the total distance while Equation 2 shows the fitness function of the chromosome.

$$TD = \sum_{i=1}^{215} Min\left(\left|d_i - c_j\right|, \forall j, j = 1,..,5\right) \quad (1)$$

$$Fitness_k = \frac{1}{TD} \quad (2)$$

where: *TD* = total distance
d = RFM score of data
i = index of data
c = cluster centre
j = index of cluster centre
k = index of chromosome

3.2.3. Crossover Operation

Since the chromosome is relatively short, then one-cut point crossover is used in this GA-clustering. The crossover operation would be executed based on a rate value, called as crossover probability. In this research, the crossover probability is set to 30%. Figure 2 shows the mechanism of the one-cut point crossover.

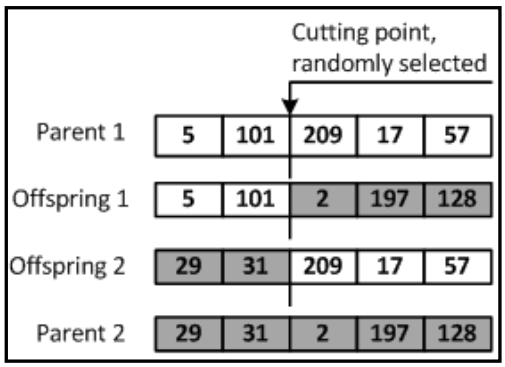

Figure 2. Mechanism of One-Cut Point Crossover

3.2.4. Mutation Operation

In this GA-clustering, mutation process was conducted by random selections of a locus in the parent chromosome and insert it in random location. It called as insertion mutation operation. Same with crossover

operation, mutation was also executed based on some rate value, called as mutation probability. In this research, the mutation probability is set to 80%. It is quite high because mutation operation could be used by GA to escape from monotone searching. Since the chromosome is relatively short then high rate of mutation would not randomising the searching process. Figure 3 shows mechanism of insertion mutation operation.

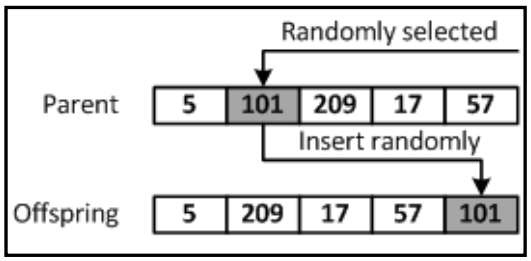

Figure 3. Mechanism of Insertion Mutation

3.2.5. Termination Condition

Basically, there are several termination condition in GA. The most commonly used by researchers are number of generation and convergence level. This research considers both stopping criteria. Initially, GA would be ran for some number of generations. After finished, the result would be analysed from the searching graph. If GA already convergence for more than half number of generation, then the result would be considered as optimum result. On the other hand, if GA is convergence for less than half number of generation, then the GA would be continued for some number of generations.

4. RESULT AND DISCUSSIONS

4.1. GA-Clustering

After GA has been ran for 1000 generations, GA proposes that data number 118, 90, 207, 92 and 69 must be cluster centres. From optimisation point of view, in GA there is no guaranty that the solution provided is optimum solution. However, there is an approximation method to justify that the result of GA is optimum or near optimum solution. Such approximation is carried out by analysing the searching graph of GA. Basically, in GA, one of the major issue is avoiding local optimum. The local optimum could occur when GA couldn't explore solution domain well and the major factor causes this problem is the diversity of the chromosomes. If the diversity of the chromosome is low, then GA would have poor reference to find the solution. Such phenomenon could be analysed from the searching graph of GA. If average value of chromosome's fitness fits the best value of chromosome's fitness then the GA could be considered being trapped in local optimum. Figure 4 shows the searching process of the GA-clustering.

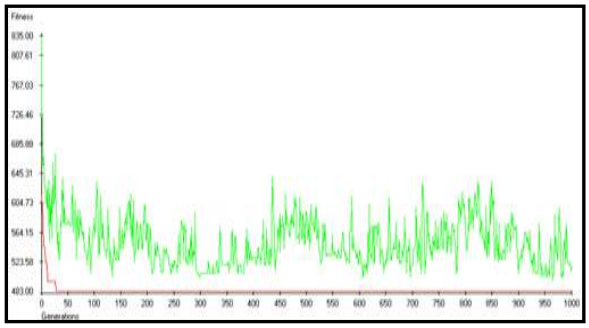

Figure 4. Searching Process of GA

It could be seen from Figure 4 that the average value of chromosome's fitness (the green line) is never fit the best chromosome's fitness (the red line). It means that the diversity of chromosomes could be maintained along searching process. Hence, it could be said fairly that the proposed GA-clustering is not trapped in local optimum condition and the solution provided could be said as optimum or near optimum solution. From Figure 4, it could be seen as well that the GA could find a solution and the solution could be maintained, without any improvement, for more than half of number of generation. Hence, it could be claimed fairly that the proposed GA is able to provide optimum solution. After the cluster centres were found, then each data could be grouped to the closest cluster centre. Table 2 shows profile of each cluster.

4.2. CRM Strategies Determination

CRM strategies for each cluster is defined through brainstorming with the owner of the firm. The CRM strategy for each cluster and example of actions in every strategy is shown in Table 3.

Table 2. Cluster Profile

Cluster number	RFM Score R	F	M	Profile
1	2	2	4	Customers in this cluster do not make purchase for long time. It could be seen from average interval time purchase, which is 13.5 months. These customers have the lowest recency than others. The number of transaction is rare, in average only twice per year. However, customers in this cluster spent high money amount per purchase, in average IDR 213.580,00 per purchase.
2	2	3	1	Customers in this cluster could be categorised as passive customers. The purchase frequency is not very often, in average only four times per year. Beside, the average money spent per purchase are the lowest, it is only in average IDR 82.071 per purchase. Customers in this cluster could be said as almost lost customers.
3	4	1	4	The last purchase of customers in this cluster is in average of 3 months ago. The transaction frequency are the lowest, in average only once per one year. However, the money spent is highest, in average IDR 249.978,00. The monetary value of customers in this cluster is the biggest.
4	4	4	2	The latest purchase of customers in this cluster is in average of 2.5 months ago. The purchase frequency of customer in this cluster is the highest, in average of eleven times per one year. Customers in this cluster could be categorised as frequent customer or repeat customers. However, customers in this cluster spent little money, in average IDR 103.907,00 per purchase.
5	3	4	4	Customer in this cluster could be categorised as potential customer. The latest purchase of customers in this cluster is in average of 6 months ago. Although the last purchase was in average of 6 months ago, the money spent is relatively high, in average IDR 175.328,00 per purchase. Beside, the purchase frequency is often, in average of 8 times per year.

Table 3. CRM Strategy for Each Cluster

CRM Strategy	Example of actions	Cluster
Initiative information	Differentiate products and services, increase quality of existing product or service.	3
Give positive impression	Encourage the customers to come again (ex: by member get member program), keep environment clean and beauty, humbly employee, easy and simple service, pre-booking through website, and so on.	2
Contact initiation	Contact the customers and treat them as very important customers, give luxurious services, introducing high class products or services and give a gifts.	1
Behaviour forming	Contact or inform customers regularly to update the customers with new products or services. Ask feedback from the customers to show that the firm prioritise customer satisfaction.	5
Mind share	Show to customers that firm's products or services are better than the competitor's products or services, conduct promotion in commercial medias, explain product knowledge to customers.	4

5. CONCLUSION

Based on the explanation above, it could be concluded that data about customer transactions could be analysed further using RFM model. In RFM, customer's loyalty would be analysed from 3 factors, interval of purchase, purchase frequency and money amount. Such method could be used to analyse potentiality of customers even though they just come to the firm rarely, but spent so much money. RFM score of customers could be clustered so that customers with similar characteristic could be grouped together, and GA is very potential to do that. Based on the experiment, GA able to cluster the RFM score of customers without any initial solution. The proposed GA also shows good performance when being used to conduct clustering. Based on profile of every cluster, then appropriate CRM strategy could be applied to every customers cluster and it is hoped the strategy would be the effective strategy to maintain existing customers of the firm.

6. REFERENCES

(a) Chang, E.C, Huang, S.C, Wu, H.H & Lo, C.F. 2007. A case study of applying spectral clustering technique in in the value analysis of an outfitter's customer database. *Proceeding of International Conference on Industrial Engineering and Engineering Management*.

(b) Cheng, C.H., & Chen, Y.S. 2009. Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications,* 36: 4176-4184.

(c) Huang, S.C, Chang, E.C, & Wu, H.H. 2009. A case study of applying data mining technique in an outfitter's customer value analysis. *Expert Systems with Applications*, 36: 5909-5915.

(d) Hughes, A. M. (1996). Boosting response with RFM. *Marketing Tools,* 3: 4-10.

(e) Liu, Y., Wu, X., & Shen, Y. 2011. Automatic clustering using genetic algorithm. *Applied Mathematics and Computation,* 218: 1267-1279.

(f) Lumsden, S.A, Beldona, S., & Morrison, A. M. 2008. Customer value in an all-inclusive travel vacation club : An Application of the RFM framework. *Journal of Hospitality & Marketing Management, 16*: 270-285.

(g) Marcus, C. 1998. *A practical yet meaningful approach to customer segmentation. Journal of Consumer Marketing*, 15: 494-504.

(h) McCarty, J., & Hastak, M. 2007. Segmentation approaches in data-mining: A comparison of RFM,CHAID,and Logistic regression. *Journal of Business Research,* 60: 656-662.

(i) Soeini, R. A., & Fathalizade, E. 2012. Customer segmentation based on modified RFM Model in the insurance industry. *International Proceeding of Computer Science and Information Technology*.

(j) Wei, J.T., Lin, S.Y., & Wu, H.H. 2010. A review of the application of RFM model. *African Journal of Business Management,* 4: 4199-4206.

(k) Wei, J.T., Lin, S.Y., Weng, C.C., & Wu, H.H. 2011. Using RFM model in a children's dental clinic for market segmentation. *Proceeding of the 11th International DSI and the 16th APDSI Joint Meeting*.

(l) Weng, S.S, & Liu, M.J. 2004. Personalized product recommendation in e-commerce. *Proceeding of International Conference on e-Technology, e-Commerce and e-Service*.

(m) Yeh, I.C, Yang, K.J, & Ting, T.M. 2008. Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications,* 36: 5866-5871.

AUTHORS BIOGRAPHIES

Muhammad Ridwan Andi Purnomo is a lecturer in Department of Industrial Engineering, Faculty of Industrial Technology, Universitas Islam Indonesia, Yogyakarta. He received her Master of Science and Philosophy of Doctor from National University of Malaysia in 2006 and 2011 respectively. His research interests are in the area of Production Planning & Control, intelligent system, robotics, automation and image processing. His email address is <ridwan_ie@uii.ac.id>.

Nur Riana Fajarwati was undergraduate student in Department of Industrial Engineering, Faculty if Industrial Technology, Universitas Islam Indonesia. She obtained her bachelor with a research on data mining and RFM topic as its focus. Her email address is <riana.inung@gmail.com>.