Effective Marketing Strategy Determination Based on Customers Clustering Using Machine Learning Technique

Muhammad Ridwan Andi Purnomo¹, Abdullah Azzam¹, Annisa Uswatun Khasanah¹

¹Department of Industrial Engineering, Faculty of Industrial Technology, Universitas Islam Indonesia, INDONESIA

Email: ridwan_ie@uii.ac.id

Abstract. Marketing is one of the high cost activities in product sales. Therefore, effective marketing is a must in a company and it should be able to encourage customers to purchase more products. One of the efforts to determine effective marketing strategies is clustering the customers and formulating correct actions for every customer cluster. Today, most of companies have digital data including customer transaction data. Techniques to analyse digital data to discover knowledge behind the data is also developed from time to time. One of the techniques in digital data analysis that receives major attention from researchers is machine learning; a technique to enable computer to do learning in analysing the data. This study presents the process of customer clustering to determine effective marketing strategy using a machine learning technique. Customers would be analysed based on 3 parameters, which are last date of coming (recency/ R), purchase frequency (frequency/ F) and total money spent for product purchased (monetary/ M). Such method is known as RFM method. Result of this study shows that the proposed machine learning could be used to cluster the customers and the customer clusters could be used as the basis for marketing manager to determine suitable marketing strategy for every customer clusters.

1. Introduction

Today's competition is harder than before due to very rapid product trend changes and customer behaviour that is more difficult to predict. Product sales are very dependent on marketing activities that carried out by company. When customer behaviour becomes more complicated, then, common marketing strategy that doesn’t consider customer behaviour becomes less effective. Marketing is one of the activities in products selling that requires high costs. Therefore, less effective marketing will result in higher products selling prices without any added value. This will result in declining product competitiveness in the market.

Currently, Information Technology (IT) has been widely used to record business data of a company. Customer detail, products and sales transactions data would be easier to manage when digitally recorded. Customer transaction data actually reflects the customer behaviour. Therefore, if the transaction data could be analysed, then the company would be able to find out the customer behaviour. From the customer behaviour that has been identified, the company would be able to
formulate marketing strategies in accordance with customer behaviour. That will increase the success of product marketing so that it will encourage customers to increase their product purchases.

Customer clustering based on transaction data could be carried out to group customers into several clusters. Customers in a cluster would have similar behaviour and would have different behaviour with customers in other clusters. From company point of view, when customers have similar behaviour in a cluster, then the company would be able to formulate suitable marketing strategy for those customers. Therefore, the formulated marketing strategy would be more effective compared to general marketing strategy. In customer clustering, several parameters are required. Some parameters that proposed by previous researchers are Recency, Frequency and Monetary (RFM). In RFM, customers would be clustered based on their last visit when purchased the products (R), number of their visit in a time interval (F) and the amount they spent to purchase the products in a time interval (M) [1]. RFM is a branch of data mining technique which extracted customer’s behaviour with 3 parameters to reduce complexity [2]. Due to its simplicity, RFM technique is widely used by previous researchers to study about customer’s behaviour [3, 4].

2. Literatures Survey
Several studies about customer segmentation have been investigated by previous researchers. A study about customer segmentation that aims to cluster the customer [5]. In a cluster, the customers have similar characteristics and effective marketing strategy could be applied in order to increase customer’s loyalty and profitability. The study shows the customer’s behaviour and commercial activity on cash back website was determined by how the customer’s role within cash back website’s social network.

Another study about customer segmentation through investigation of how customer uses multi-channel and media in modern retail has been investigated by previous researchers [6]. In such study, the researchers used latent-class clustering technique to cluster the customers based on purchase channels and media touch points. Result of the study showed a comprehensive customer’s behaviour based on cross reference between purchase channels and media touch points when they conducted purchases.

Study about customer segmentation in customised product has been investigated by previous researchers [7]. The study showed that it was harder to identify pattern of customer preferences because of diversity of the products sold. In that study, two data mining techniques have been compared; which were clustering and subgroups discovery; in order to identify market segments and rules to identify the customer preferences. Result of the study showed that the customer segments and the association rules could be used as the references to identify customer’s order pattern.

Market basket analysis (MBA) that is one of the data mining techniques has been implemented to conduct customer visit segmentation [8]. In that study, customer visit was identified based on the purchased products. It aimed to identify the shopping intention or mission behind the visit. A semi-supervised feature selection approach that used the product taxonomy as input and customised categories as output was proposed. Result of the study could be used as the basis for effective customer-based product campaign and layout improvement of the investigated store.

Customer segmentation usually works based on set of data. Therefore, data mining techniques are widely used in that field. A previous study investigated hybrid soft computing approach for customer segmentation [9]. There were two modules applied to process a set of data. In the first module, a clustering technique was implemented to cluster existing customer, while in the second module, a hybrid feature selection was implemented based on filtering and multi-criteria decision making (MCDM) technique. Final output of such study was extracted IF-THEN rules in order to predict value of the customers.

Similar study about rules extraction for customer segmentation has been conducted by previous researchers [10]. In that study, apriori algorithm based on clustering techniques has been implemented. There are three clustering techniques that were utilised namely k-means, k-medoids and Self
Organising Map. The clustering techniques have been implemented based on presence, frequency and monetary value. Result of the study was customer profiles based on the clustering result.

Customer segmentation has impact to the demand supplies from the previous supply chain. Different customer segmentation could be considered as the reference to determine product supply strategy. Previous researchers have conducted a study about product substitution of demand supply in panic buying behaviour [11]. There are two condition considered in such study which are panic and supply disruption condition. In the panic condition, wholesaler retains the inventory to satisfy retailers, while in supply disruption condition, product substitution strategy that accepted by the retailers are implemented. The study aims to maximise profit when both conditions are occurred.

Several previous studies show that segmenting the customers is usually conducted using clustering technique. The customers could be clustered based on several criteria, such as product preference [12], emotion [13] and region and product demand [14]. Besides customer, products also could be clustered based on customer purchasing pattern. A study about product recommendation to particular customers based on customer purchasing data in the past has been conducted by previous researchers [15]. Customers and product purchased is connected through a 0-1 matrix indices customers purchased the products (1) or not (0). In such study, a method to avoid overlap cluster was also proposed so that at the end a perfect products-customers matrix could be obtained. Based on the perfect cluster matrix, then products could be recommended to customers who have similar behaviour with the customers in the same product cluster.

A study about vehicle product cluster using fuzzy clustering and genetic algorithm has been investigated by previous researchers [16]. The reason of using fuzzy clustering was because traditional clustering technique could not satisfy real time clustering. The genetic algorithm was used to determine optimal number of clusters. A data set with 7 criteria for clustering has been used to test the proposed method and the result showed that the proposed method is effective to do clustering in customer demand and product orientation.

3. The Method
This study took place in a furniture company that has online and offline shops with integrated data. Data were collected using JavaScript Object Notation (JSON) system. There are 2 master data, which are customer and product master data; and 1 transaction data that is sales data that records transaction of customers that have purchased the products. There are 192 registered customers and 37 kinds of furniture products. The sales data are extracted from year 2016 until year 2018, and there are 30240 records. The extracted data would then be saved in a temporary RFM table that consists of 8 fields which are transaction_id, invoice_id, product_code, customer_id, product_name, quantity, invoice_date and unit_price.

Recency parameter value was obtained from subtraction of current date and last date of customer when purchased the products. Frequency parameter value was obtained from number of transactions of every customer, while monetary parameter value was obtained by summing product price purchased by customer. Data warehousing and parameters value calculation were conducted by writing and executing several Structured Query Language (SQL).

4. Result and Discussions
Figure 1 shows the description of the recency calculation by using “.describe()” function in python programming language.
From the description depicted in Figure 1, it could be known that standard deviation of the recency is still relatively high; therefore, a clustering technique is required in order to group the customers into several clusters with minimum inertia. This study uses K-mean clustering as it tries to separate samples in several clusters of equal variances and minimising the inertia; that could be measured using Sum of Squared Error (SSE) value. Figure 2 shows SSE value versus number of clusters.

From Figure 2, it could be known that start from 4 cluster numbers, the SSE value is not significantly changed; therefore; 4 cluster numbers would be selected to cluster the customers, and Figure 3 shows the description of the clustering result.

Figure 3 shows that mean of every cluster is significantly different and it means that customers in a cluster have significant different recency value from customers in another cluster.
Purchase frequency aspect is calculated based on number of transactions for every customer, instead of number of invoices. Such way would be fairer because in an invoice, there are could be more than one transaction. Figure 4 shows description of purchase frequency aspect calculation.

Figure 4. Description of purchase frequency calculation

Figure 4 shows that purchase frequency gap and its standard deviation is relatively high. It means that customer purchase behaviour is different. In term of marketing, specific treatment to encourage customers to increase their purchase frequency is required. Customers clustering for purchase frequency would be carried out using same method as in recency, which is K-mean clustering method. In order to determine effective cluster number, same parameter as in recency, which is SSE value would also be used for such purpose. Figure 5 shows the SSE value versus number of clusters. Based on Figure 5, four cluster numbers would be selected as number of customer clusters. Figure 6 shows the description of purchase frequency clustering result.

Figure 5. SSE value versus number of cluster graph

| Frequency_Cluster | count | mean | std | min | 25%  | 50%  | 75%  | max  |
|-------------------|-------|------|-----|-----|------|------|------|------|
| 0                 | 36.0  | 102.972222 | 15.020911 | 64.0 | 96.75 | 107.5 | 113.00 | 123.0 |
| 1                 | 64.0  | 144.466750  | 9.831870  | 126.0 | 136.00 | 145.0 | 154.00 | 158.0 |
| 2                 | 58.0  | 173.310345  | 9.717423  | 158.0 | 164.25 | 173.0 | 181.50 | 192.0 |
| 3                 | 34.0  | 212.794118  | 16.255124 | 194.0 | 199.50 | 206.5 | 222.75 | 262.0 |

Figure 6. Description of clustering result based on purchase frequency aspect
From Figure 6, it could be seen that mean value of every cluster is significantly different and it means that customers in a cluster have significant different purchase frequency with customers in another cluster.

Monetary aspect reflects total money amount spent by the customers; therefore, monetary aspect is calculated by multiplying number of products purchased by customers with its price. It could be understood that when customers spent a lot of money, it means that the customers were interested with the products and when they purchased the product from the shop, it means that the customers are loyal customers. Figure 7 shows description of monetary aspect calculation.

From Figure 7, it could be analysed that that monetary value gap and its standard deviation is still relatively high. It could be said that customer loyalty is different. In term of marketing, specific treatment to increase their loyalty is required. Customers clustering for monetary aspect would be carried out using same method as in previous two aspects, which is K-mean clustering method. Number of clusters would be analysed based on SSE value as well. Figure 8 shows the SSE value versus number of clusters. Based on Figure 8, four cluster numbers would be selected as number of customer clusters. Figure 9 shows the description of clustering result based on monetary aspect.

![Figure 7. Description of monetary aspect calculation](image)

![Figure 8. SSE versus number of cluster graph](image)
From Figure 9, it could be seen that mean value of monetary aspect of every cluster is significantly different and it means that customers in a cluster have significant different monetary value with customers in another cluster.

After conducted customers clustering based on RFM aspect, then, the final step is scoring the customers by summing all of the aspects score. Figure 10 shows the result of the final scoring.

Based on Figure 10, it could be seen that the customers could be clustered into 10 clusters. However, in term of marketing strategy, 10 different marketing strategies would be cost consuming and sometime would be not effective. Therefore, the 10 customer clusters would be compressed into 3 clusters, that could be labelled as ‘less-loyal’ customers, who have final score from 0 to 2, ‘slightly loyal’ customers, who have final score from 3 to 6 and ‘loyal’ customers, who have final score from 7 to 9. After being compressed, then customer cluster profile could be identified and several marketing strategy alternatives could be formulated. Table 1 shows the customer cluster profile and the marketing strategy alternatives.

| Customer Cluster | Profile | Marketing Strategy Alternatives |
|------------------|---------|---------------------------------|
| 1                | Last visit is more than 3 months ago, visit frequency is middle and money spent is middle. | Introducing new low to middle price products to encourage customers to come again. |
| 2                | Last visit is more than 1 month ago, visit frequency is middle and money spent is middle. | Introducing new middle price products and product bundling with |
spent is much.

| 3 | Last visit is just last week, visit frequency is high and money spent is very much. |
|---|--------------------------------------------------------------------------------|

Introducing new high price products and give touching and warm service to the customers.

5. Conclusion and Recommendations

From the case study, it could be concluded that customer purchase behaviour is not the same and marketing strategy must be formulated based on the customer purchase behaviour. RFM could be used as reference aspects to cluster customers in order to identify their purchase behaviour. In term of customer clustering, K-mean technique; as supervised clustering method; could be used to cluster the customer behaviour by determining in advance the cluster number. In real application, cluster number could be the same with number of marketing strategy formulated by the company. Combination of RFM and K-mean clustering could be used as machine learning technique to train a computer to conduct a data mining based on human expert reasoning. For further research, it is recommended to use unsupervised clustering technique by defining cluster radius to control homogeneity of the customer behaviour. By using the proposed technique, marketing strategy could be formulated more effective because customer’s behaviour in a cluster would be more homogeny.

6. Acknowledgment

The authors would like to appreciate Minister of Research, Technology and Higher Education, Republic of Indonesia and Directorate of Research and Community Service, Islamic University of Indonesia for supporting this research through a leading university basic research scheme, year 2019.

7. References

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

[2] Hu, Y.H., and Yeh, T.W. (2014) Discovering valuable frequent patterns based on RFM analysis without customer identification information, *Knowledge-Based Systems*, 61, 76-88.

[3] Song, M., Zhao, X., E, H., and Ou, Z. (2017) Statistics-based CRM approach via time series segmenting RFM on large scale data, *Knowledge-Based Systems*, 132, 21-29.

[4] Hu, Y.H., Huang, T.C.K., and Kao, Y.H. (2013) Knowledge discovery of weighted RFM sequential patterns from customer sequence databases, *The Journal of Systems and Software*, 86, 779-788.

[5] Ballestar, M.T., Grau-Carles, P., and Sainz, J. (2018) Customer segmentation in e-commerce: applications to the cashback business model. *Journal of Business Research*, 88, 407-414.

[6] Nakano, S., and Kondo, F.N. (2018) Customer segmentation with purchase channels and media touchpoints using single source panel data. *Journal of Retailing and Consumer Services*, 41, 142-152.

[7] Brito, P.Q., Soares, C., Almeida, S., Monte, A., and Byvoet, M. (2015) Customer segmentation in a large database of an online customized fashion business. *Robotics and Computer-Integrated Manufacturing*, 36, 93-100.

[8] Griva, A., Bardaki, C., Pramatari, K., and Papakiriakopoulos, D. (2018) Retail business analytics: Customer visit segmentation using market basket data. *Expert Systems With Applications*, 100, 1-16.

[9] Khalili-Damghani, K., Abdi, F., and Abolmakarem, S. (2018) Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries. *Applied Soft Computing Journal*, 73, 816-828.

[10] Silva, J., Varela, N., Lopez, L.A.B., and Millan, R.H.R. (2019) Association rules extraction for customer segmentation in the SMEs sector using the apriori algorithm, *Procedia Computer Science*, 151, 1207-1212.
[11] Tsao, Y.C., Raj, P.V.R.P., and Yu, V. (2019) Product substitution in different weights and brands considering customer segmentation and panic buying behaviour. *Industrial Marketing Management*, 77, 209-220.

[12] Zhang, J., Simeone, A., Gu, P., and Hong, B. (2018) Product features characterization and customers' preferences prediction based on purchasing data, *CIRP Annals - Manufacturing Technology*, 67, 149-152.

[13] Poirson, E., and Da Cunha, C. (2019) A recommender approach based on customer emotions, *Expert Systems With Applications*, 122, 281-288.

[14] Wang, Y., Assogba, K., Liu, Y., Ma, X., Xu, M., and Wang, Y. (2018) Two-echelon location-routing optimization with time windows based on customer clustering, *Expert Systems With Applications*, 104, 244-260.

[15] Balakrishnan, J., Cheng, C.H., Wong, K.F., and Woo, K.H. (2018) Product recommendation algorithms in the age of omnichannel retailing - An intuitive clustering approach, *Computers & Industrial Engineering*, 115, 459-470.

[16] Jie, L., Ailing, G., and Lianyong, Y. (2011) Application of product cluster method based on fuzzy cluster and genetic algorithm, *Procedia Engineering*, 15, 4930-4935.