RFM-BASED CUSTOMERS CLUSTERING FOR PRECISE INDUSTRIAL MARKETING STRATEGY FORMULATION

M.R.A. Purnomo[1], A.R. Anugerah[2], A. Azzam[1] A.U. Khasanah[1] and M.N. Alfareza[3]

[1]Faculty of Industrial Technology,
Universitas Islam Indonesia, Yogyakarta,
55581, Indonesia.

[2]Institute of Tropical Forestry and Forest Products,
Universiti Putra Malaysia, Serdang,
43400, Selangor, Malaysia.

[3]College of Engineering,
National Taiwan University of Science and Technology,
Da'an District, 106335, Taipei City, Taiwan.

Corresponding Author's Email: [1]ridwan_ie@uii.ac.id

Article History: Received 22 May 2020; Revised 6 December 2020; Accepted 16 April 2021

ABSTRACT: In the information age, digital marketing which covers data processing and spreading of digital information is an effective marketing strategy alternative. In industrial sector, marketing is one of the costly activities, therefore, it needs to be carried out accurately. This study discusses an accurate industrial marketing strategy formulation based on customers purchasing behaviour. Customer sales data were analysed by a well-known method, namely Recency, Frequency, and Monetary (RFM) combining with data analytics technique. Results indicated that data analytics could improve the preciseness of the conventional RFM by considering non-profit customers and provide real-time data. This research gives another insight on improving conventional RFM analysis effectiveness as a simple and widely used tool for marketers.

KEYWORDS: *RFM; Customers Clustering; Data Analytics; Precise Marketing*

1.0 INTRODUCTION

Information plays an important role in industrial sector, especially in marketing. There are a lot of ineffective marketing strategies that lead to insignificant increment in product sales. Besides, it sometimes degrades product or company reputation. Marketing must be carried out at a maximum customer orientation through the advances in technology and information technology [1].

It is very important to consider customers behaviour in products marketing. There are two approaches in customers'-oriented marketing. First is product tailoring to the related customers [2]. In this approach, marketing strategy is initiated with customer study to understand its profile. The next step is to develop products and communicate it to the customers through all marketing channels. Lastly, improve the marketing staff's skill to effectively market the products [3].

The second approach is customers profiling that also use customer data to formulate marketing strategies. This approach is suitable when there are a lot of products to market. In this approach, the customers are offered with correct products according to their preference. Therefore, they will receive correct marketing approach and are encouraged to purchase. Numerous researchers have used this approach to formulate marketing strategy, as for instance [4-7].

There were several previous studies in clustering that have been carried out with different objectives. As for instance, Motlagh et al. [8] did a clustering analysis on electricity load time series data to predict electricity load from time to time and to prepare the electricity supply in the future. Clustering analysis was also used to predict customer churn rate in the telecommunication business [9-10]. In logistic, customer clustering technique was used as the basis for product distribution optimisation based on multi-criteria variables [11-12].

Clustering technique can be used to increase customer's willing to pay (WTP) for expensive product. He et al. [13] conducted a study to attract customer WTP for a green housing (GH) by clustering the customers based on their characteristics. From the customer clustering analysis, researcher could identify factors that influence customer' WTP for a GH to develop marketing strategy.

This research aims to develop precise marketing strategies according to customers recency, frequency, and monetary (RFM) data in the furniture industry. Three variables in RFM reflect the customers purchasing behaviour, thus marketing strategy can be formulated. The RFM is used because it is able to explore the quantitative characteristics of customers and their value could be reflected by the most recent consumption, frequency in normal consumption and monetary cost of customers [14]. It is very important to consider responsiveness of the customers to promotions (recency), engaged and satisfaction level of customers with the products (frequency) and economic ability of customers in purchasing products (monetary), in formulating the marketing strategies.

RFM as a marketing technique was considered easy to follow and relatively cheap. Several researchers have developed a neuromarketing technology to detect brain response by analysing customer' brain activities [15]. Such technology was developed to minimise bias or unclear customer' answers on questionnaire. However, collecting samples and developing such technologies to have precise results requires more time and money.

RFM technique has been combined with other approach to increase their preciseness, as for instance: considering both customers and product perspectives [16], applying business intelligent to predict customer purchase [17] and integrating it with Analytical Hierarchy Process (AHP) [18]. The novelty in this study is the use of data analytics in the RFM technique. According to the literature, this mixed technique was not received intensive attention from the previous researchers. In this study, data analytics is used to discover knowledge behind the large size of sales data. It would be the reference for the marketing experts in formulating precise marketing strategy besides the RFM result.

2.0 MATERIALS AND METHODS

This research employed more than 100,000 data from one of furniture companies in Indonesia as a case study. The data covers customer' purchasing information from January 2013 to the end of November 2020. There are two main steps in this study: first is data warehousing to extract the purchasing data information and secondly is grouping the customers according to three main variables: recency, frequency, and monetary.

2.1 Data Acquisition

In the data acquisition process, the first step is data extraction from the database using JSON system. Data acquisition was conducted through data warehousing of transactional data and integrating it to last transaction (recency), purchase frequency and total money spent (monetary). There are 2 steps in the data warehousing:

Step 1: Create a view object that consist of three main data: last transaction, purchase frequency and total money spent. In this research, the view object is named as rfm_value. The SQL script is shown as follow:

*SELECT customer_id, MAX(transaction_date) AS last_coming_date, COUNT(*) AS frequency, SUM(quantity * unit_price) AS monetary FROM transactions group by customer_id*

Step 2: Create a view object that showed the summary of RFM value for each customer. In this research, the view object is named as summary_rfm_value. The SQL script is shown as follow:

SELECT customer_id, DATEDIFF((CONVERT((SELECT MAX(last_coming_date) FROM rfm_value), CHARACTER)), (CONVERT((last_coming_date), CHARACTER))) AS recency_value, ((SELECT MAX(frequency) FROM rfm_value)-frequency) AS frequency_value, ((SELECT MAX(monetary) FROM rfm_value)-monetary) AS monetary_value FROM rfm_value

2.2 Customers RFM Data

Customer clustering with RFM technique is conducted through 4 main steps: recency, frequency and monetary score calculation for each customer, and lastly customer clustering based on combined RFM scores.

2.2.1 Recency Analysis

Recency score is calculated by subtracting latest purchase date of customer with purchase date of every customer. Equation (1) shows the formula to calculate the recency score such as

$$R_c = \mathrm{Max}(\mathrm{purchase\_date}) - \mathrm{purchase\_date}_c ; c = 1,2,3,\ldots C \qquad (1)$$

where Rc is the recency value of customer c, purchase_date is the customer' purchase date in the data warehouse, c is the customer index and C is the number of distinctive customers in the data warehouse.

The recency score is then be grouped into 20-day interval. The interval was suggested by the marketing experts in the observed furniture company. Figure 1 shows the customers recency distribution data. Results show that there were more than 50 customers and 30 customers who did not make a transaction for more than 2 months and 1 month, respectively.

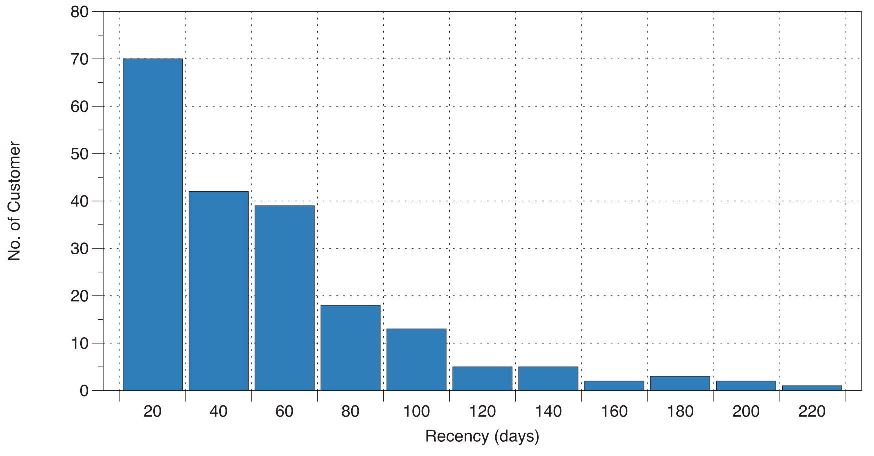


Figure 1: Customers recency value distribution

2.2.2 Frequency Analysis

Frequency analysis is based on the purchase frequency data on each customer. The frequency score is calculated by counting the customer visit during analysis period. Figure 2 shows the distribution of purchasing frequency in the observed furniture company. In order to retrieve customer data based on the frequency value, the SQL was executed such as

Select Count (customer_id) customer_id Group by customer_id

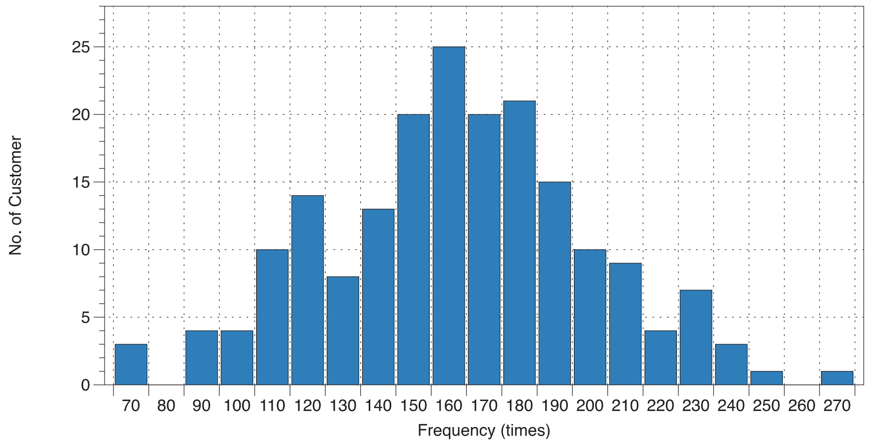


Figure 2: Customers frequency value distribution

The data shows similar pattern with normal distribution curve. It means the number of customers with average frequency score is more and number of customers with left-extreme or right-extreme frequency score is less.

2.2.3 Monetary Analysis

Monetary analysis was conducted based on the money spent by every customer to purchase the products. Similar with frequency analysis, monetary data were retrieved by applying an SQL:

*Select sum (product_price * purchase_quantity) Group by (customer_id)*

Monetary data is one of the factors in determining the product sales. For better understanding and representation, data were classified in IDR 100,000 base value as shown in Figure 3.

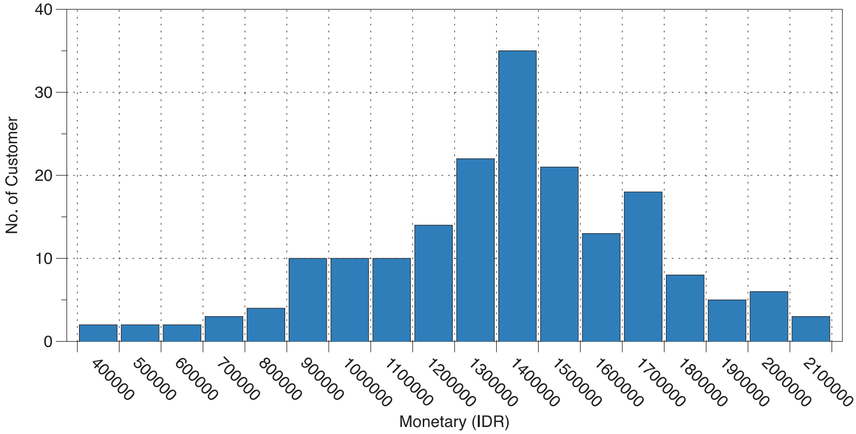


Figure 3: The average money spent distribution

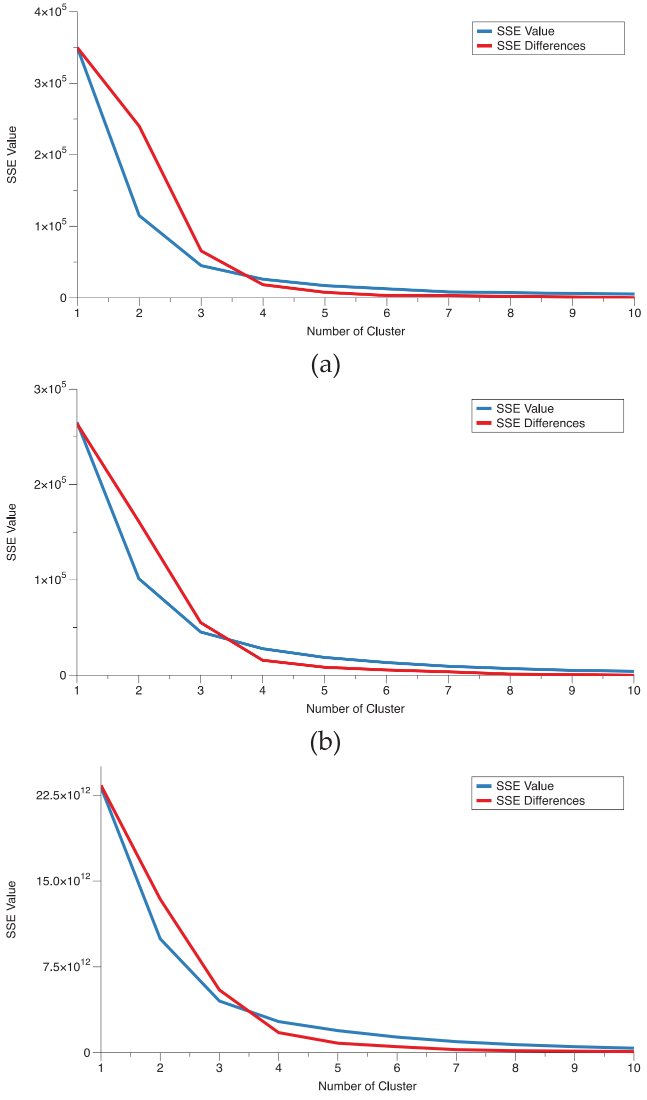
Figure 3 shows that the monetary pattern is normal. Most of customers spent money in average and only few customers spent extremely small and big amount of money. However, the graph also shows slightly skew to the right pattern. It means that the products are interesting for the customers. For this reason, the company has chances to increase their sales through effective marketing strategy implementation.

2.3 K-Means Clustering

In this study, K-mean clustering is utilised due to large and high dimension of data in the warehouse [19]. It is realized that there is a trade-off between clustering quality and cost to develop marketing strategy. The higher number of clusters means the greater number of marketing strategies are required thus the marketing cost becomes higher. In order to solve this trade off, the number of clusters would be determined based on Sum of Squared Error (SSE) as formulated in Equation (2). SSE represents a total distance of all cluster members from the cluster centre and it was used as the basis of cluster number determination. Figures 4(a) to 4(c) show the SSE value distribution for 10 cluster scenarios.

$$SSE = \sum_{i=1}^{n} (r_i - \bar{r})^2 \tag{2}$$

where i is the data index, n is the number of data and r is the recency value.



(a)

(b)

(c)

Figure 4: SSE value distribution to determine number of marketing strategies according to (a) recency, (b) frequency and (c) monetary score

The SSE value for clusters 4 to 10 is not significantly different in all variables: recency, frequency, and monetary. For this reason, the customers were grouped into 4 clusters to minimise marketing cost. The descriptive statistic details on each cluster are shown in Table 1.

Table 1: Description of clustering result based on recency score

Recency (days)								
Cluster	Count	Mean	STD	Min	25%	50%	75%	Max
0	11	167.182	23.219	132.00	152.00	163.00	185.00	201.00
1	33	88.485	16.131	68.00	75.00	83.00	99.00	126.00
2	53	45.830	10.483	30.00	37.00	44.00	54.00	65.00
3	95	13.053	8.530	0.00	6.00	12.00	20.00	29.00
Frequency (times)								
Cluster	Count	Mean	STD	Min	25%	50%	75%	Max
0	36	102.972	15.021	64.00	96.75	107.50	113.00	123.00
1	65	144.692	9.920	126.00	136.00	145.00	154.00	159.00
2	57	173.561	9.612	160.00	165.00	173.00	182.00	192.00
3	34	212.794	16.255	194.00	199.50	206.50	222.75	262.00
Monetary (in Rupiah)								
Cluster	Count	Mean	STD	Min	25%	50%	75%	Max
0	37	8.4e+05	1.5e+05	470817	8e+05	8e+05	9e+05	11e+05
1	90	1.3e+06	1.0e+05	1076971	12e+05	13e+05	14e+05	14e+05
2	51	1.6e+06	1.1e+05	1464860	15e+05	16e+05	17e+05	18e+05
3	14	2.1e+06	1.8e+05	1865409	19e+05	20e+05	21e+05	26e+05

According to recency score, most of the customers were located in cluster 3 with the lowest recency score. It means, most of them are loyalist with average days to the last purchase within 13 days or less than 2 weeks. The overall results show that the number of recencies was skewed to the right, which means there was a tendency of customer to be a loyalist. However, 5.7% of current customers were expected to not make a purchase again.

Frequency and monetary shared similar pattern, the smallest STD can be found in clusters 1 and 2; however, slightly higher deviation can be found in clusters 0 and 3. This means the behaviour in clusters 1 and 2 were similar while varied in other clusters. Although customers were already grouped in clusters 1 and 3, but they actually have different behaviour, thus must be analysed together with recency, frequency and monetary.

3.0 RESULTS AND DISCUSSION

In order to get comprehensive customers' profile, the final step in RFM analysis is to combine scores from three main variables: recency, frequency and monetary. In this study, the combination was carried out by the mean value. Table 2 shows the combination of each score.

Table 2: Recency, frequency and monetary combination

Final Score	Recency (days)	Frequency (times)	Monetary (in Rupiah)
0	174.666667	79.666667	6.485973e+05
1	91.714286	105.428571	8.556969e+05
2	95.928571	125.071429	1.034655e+06
3	54.685714	123.114286	1.038851e+06
4	49.240000	148.840000	1.301631e+06
5	32.257143	156.514286	1.327513e+06
6	35.280000	176.400000	1.467248e+06
7	14.300000	180.350000	1.605146e+06
8	19.444444	205.888889	1.744422e+06
9	12.700000	228.100000	2.079259e+06

In RFM, basically customers are scored using discrete value from 111 to 555. Customers with score equals to 555 are categorized as potential customers while customers with score equals to 111 are categorized as lost customers [20]. In this research, customers were firstly classified into 10 clusters (from 0-9) to reduce the possibility of customers to be included in a group with different buying behaviour [21]. In this case, customers under final score 0 to 2 is classified as potential lost customers and those under score 7 to 9 are categorized as potential customers. These 10 clusters were then classified into 3 new categories (cluster) according to the number of proposed marketing strategies as shown in Table 3.

Table 3: Proposed marketing strategies on each cluster

Cluster	Marketing Strategies
Cluster 1 (score 0-2)	Encourage the customers to repurchase in store by giving discounted price for expensive products.
Cluster 2 (score 3-6)	Proactively inform new products arrival, promotional programs, and customer relationship programs.
Cluster 3 (score 7-9)	Provide excellent services when they are coming to store

In this study, non-potential customers are still used (customers in clusters 1 and 2) as a target market. It is according to the fact that products in this retail are interesting and the graph in the monetary value shows the product sales pattern is skew to the right. In another words, it is highly potential for current customers to spend more and make repeated order. This action was taken to answer the findings from previous research that found RFM ignores the non-profit customers [22].

The marketing strategy in cluster 1 was formulated based on the condition that customers did not make purchase for a long time. However, their frequency score was not significantly different with cluster 2. The money was relatively low spent. Moreover, the marketing strategy in cluster 2 was formulated because the customers did not make purchase in slightly long time. The frequency score was good with 123 to 156 times buying experience but relatively spent small money (IDR 1,000,000 to IDR 1,300,000). The proposed company is to introduce new products or programs to the customers in this cluster.

The marketing strategy in cluster 3 was to maintain and increase retention of the loyal customers. The last time customers under this group going for shopping is 2 weeks (14 days) and spent the most compared to another cluster. In order to support the marketing strategies, a mobile application is strongly recommended to develop. It can be used as communication media on particular information because customers are using their mobile phone frequently in daily activities.

This research was able to tackle several opinions on the disadvantages of RFM as a marketing tool. First, the possibility of customers to be included into a group with different buying behaviour and clusters is minimised. Second, the past opinions that believed RFM analysis ignores the non-profit customers is considered and tackled. Third, the SQL allows new transaction data to be added periodically. This is important because customers' purchasing behaviour and data were actively changing [23]. Once new data is added into the system, SQL would re-calculate and re-suggest customers that is suitable with certain marketing strategy.

Although, the number of score groups is enlarged into 10 (Table 2), the purchasing behaviour of customers with score 0-2 and cluster 3-6 are relatively similar. It is possible if several data fall in cluster border area [23]. In this study, some customers were fall in the border area and their profiles were not significantly different with other clusters. This means the customers could be considered as member of the other nearby cluster with some degree of membership value. Therefore, more than one marketing strategy could be applied. For such analysis, a fuzzy database system is proposed by implementing fuzzy-SQL method for further research. This method allows the marketing experts to analyse with some part of the RFM variables.

4.0 CONCLUSION

Ineffective marketing strategy is frequently happened to any business sectors and resulting to insignificant increment in product sales. RFM as an easy-to-follow, relatively inexpensive, and widely-know method was proposed to develop precise marketing strategy. However, conventional RFM has several weaknesses, as for instance: the possibility of customers to be included in different purchasing behaviour, the ignorance of non-profit customers, and not dynamic when the new data were added. In this research, the conventional RFM was integrated with data analytics to improve preciseness and to tackle several opinions on RFM' weakness. In this research, RFM was classified into 10-based score and was used as a basis to identify clusters that suitable with proposed marketing strategy. the use of data analytics in RFM could discover the real customers' behaviour. Customers with lowest RFM score is not always considered as lost customers. The result RFM analysis supported by data analytics could be used as a reference by marketing experts in formulating precise marketing strategy to encourage all of the customers to increase the product sales. This research gives another insight on improving conventional RFM analysis effectiveness as a simple and widely used tool for marketers around the world.

ACKNOWLEDGMENTS

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

REFERENCES

[1] B.A. Abishovna, "The principle of effective marketing management", *Procedia - Social and Behavioral Sciences*, vol. 109, pp. 1322 – 1325, 2014.

[2] S. Lian, Y. Xu and C. Zhang, "Family profile mining in retailing", *Decision Support System*, vol. 118, pp. 102-114, 2019.

[3] A.D. Gregorio, I. Maggioni, C. Mauri and A. Mazzucchelli, "Employability skills for future marketing professionals", *European Management Journal*, vol. 37, no. 3, pp. 251-258, 2019.

[4] C. Calvo-Poral and J.P. Levy-Mangin, "Profiling shopping mall customers during hard times", *Journal of Retailing and Customer Services*, vol. 48, pp. 238-246, 2019.

[5] D. Vicari and M. Alfo, "Model based clustering of customer choice data", *Computational Statistics and Data Analysis*, vol. 71, pp. 3-13, 2014.

[6] A. Mahbubi, T. Uchiyama and K. Hatanaka, "Capturing consumer value and clustering customer preferences in the Indonesia halal beef market", *Meat Science*, vol. 156, pp. 23-32, 2019.

[7] V. Holy, O. Sokol and M. Cerny. "Clustering retail products based on customer behavior", *Applied Soft Computing*, vol. 60, pp. 752–762, 2017.

[8] O. Motlagh, A. Berry and L. O'Neil, "Clustering of residential electricity customers using load time series", *Applied Energy*, vol. 237, pp. 11-24, 2019.

[9] S. Mitrovic, B. Baesens, W. Lemahieu and J. De Weerdt, "tcc2vec: RFM-informed representation learning on call graphs for churn prediction", *Information Sciences*, vol. 557, pp. 270-285, 2021.

[10] I. Bose and X. Chen, "Detecting the migration of mobile service customers using fuzzy clustering", *Information & Management*, vol. 52, no. 2, pp. 227-238, 2015.

[11] Y. Wang, X. Ma, Y. Lao and Y. Wang, "A fuzzy-based customer clustering approach with hierarchical structure for logistics network optimization", *Expert Systems with Applications*, vol. 41, no. 2, pp. 521-534, 2014.

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

[13] C. He, S. Yu, Q. Han and B. de Vries, "How to attract customers to buy green housing? Their heterogeneous willingness to pay for different attributes", *Journal of Cleaner Production*, vol. 230, pp. 709-719, 2019.

[14] M. Song, X. Zhao, E. Haihong, and Z. Ou, "Statistics-based CRM approach via time series segmenting RFM on large scale data", in International Conference on Utility and Cloud Computing, Shanghai, China, 2016, pp. 282–291

[15] N.A. Mahamad, M.K.M. Amin, and O. Mikami. "Evaluation neuromarketing technique on consumer satisfaction using EGG Imaging", *Journal of Advanced Manufacturing Technology*, vol. 13, no. 2(2), pp. 11-22, 2019.

[16] R. Heldt, C.S. Silveira and F.B. Luce. "Predicting customer value per product: From RFM to RFM/P", *Journal of Business Research*. vol. 127, pp. 444-453, 2021.

[17] S. Haghighatnia, N. Abdolvand and S. Rajaee Harandi. "Evaluating discounts as a dimension of customer behavior analysis", *Journal of Marketing Communications*, vol. 24, no. 4, pp. 321-336, 2018.

[18] S. Monalisa, P. Nadya and R. Novita. "Analysis for Customer Lifetime Value Categorization with RFM Model", *Procedia Computer Science*, vol. 161, pp. 834–840, 2019.

[19] F.G. Tari and Z. Hashemi, "Prioritized K-mean clustering hybrid GA for discounted fixed charge transportation problems", *Computers & Industrial Engineering*, vol. 126, pp. 63-74, 2018.

[20] A. Alizadeh Zoeram and A.R. Karimi Mazidi, "New approach for customer clustering by integrating the LRFM model and fuzzy inference system", *Iranian Journal of Management Studies*, vol. 11, no. 2, pp. 351-378, 2018.

[21] A. Dursun and M. Caber, "Using data mining technique for profiling profitable hotel customers: An application of RFM analysis", *Tourism Management Perspectives,* vol. 18, pp. 153-160, 2016.

[22] M-F. Băcilă, A. Rădulescu, and I.L. Marar, "RFM based segmentation: An analysis of a telecom company's customers", International Conference on Marketing from Information to Decision, Cluj-Napoca, Transylvania, 2012, pp. 52-62.

[23] J.R. Miglautsch, "Thoughts on RFM scoring", *Journal of Database Marketing & Customer Strategy Management,* vol. 8, no. 1, pp. 67-72, 2000.