International Journal of Electrical and Computer Engineering (IJECE)
Vol. 12, No. 6, December 2022, pp. 6132~6139
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i6.pp6132-6139

Indonesian pharmacy retailer segmentation using recency frequency monetary-location model and ant K-means algorithm

Ghea Sekar Palupi[1], Muhammad Noor Fakhruzzaman[2]
[1]Department of Informatics, Faculty of Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia
[2]Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga, Surabaya, Indonesia

Article Info

Article history:

Received Aug 25, 2021
Revised Jun 17, 2022
Accepted Jul 10, 2022

Keywords:

Ant K-means
Logistics
Machine learning
Retailer segmentation
Sustainable industry

ABSTRACT

We proposed an approach of retailer segmentation using a hybrid swarm intelligence algorithm and recency frequency monetary (RFM)-location model to develop a tailored marketing strategy for a pharmacy industry distribution company. We used sales data and plug it into MATLAB to implement ant clustering algorithm and K-means, then the results were analyzed using RFM-location model to calculate each clusters' customer lifetime value (CLV). The algorithm generated 13 clusters of retailers based on provided data with a total of 1,138 retailers. Then, using RFM-location, some clusters were combined due to identical characteristics, the final clusters amounted to 8 clusters with unique characteristics. The findings can inform the decision-making process of the company, especially in prioritizing retailer segments and developing a tailored marketing strategy. We used a hybrid algorithm by leveraging the advantage of swarm intelligence and the power of K-means to cluster the retailers, then we further added value to the generated clusters by analyzing it using RFM-location model and CLV. However, location as a variable may not be relevant in smaller countries or developed countries, because the shipping cost may not be a problem.

Corresponding Author:

Muhammad Noor Fakhruzzaman
Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga
Jl. Dr. Ir. H. Soekarno, Surabaya, Indonesia
Email: ruzza@stmm.unair.ac.id; ruzcmc@gmail.com

1. INTRODUCTION

As an established country, Indonesia has one of the most lucrative pharmacy industries in South-East Asia. A state-owned company dealing with pharmaceutical product distribution is the current market leader in pharmacy, with sporadic retailers popping up across the country. In addition, medical devices, medical personnel, and medicine have all played an important role during the pandemic in ways people probably had not quite noticed before. Increasing demand has prompted the government to add the medical device and pharmaceutical sectors to the Making Indonesia 4.0 priority sectors list. As the ASEAN Economic Community gradually integrates pharmaceutical regulations, other pharmacy distribution companies in ASEAN countries were coming to Indonesia for the open market share. The integrated pharmacy company data stated that Vietnam has grown by 20% in ASEAN, compared to Indonesia which only grown by 12 to 15%. This data shows that Indonesia pharmacy distribution companies needed to improve if they want to stay in the race [1]. However, Indonesia is also lacking in terms of pharmaceutical product exports compared to Vietnam and Malaysia, other ASEAN countries [2]. If those two countries fill up the pharmacy market in Indonesia, the current leading distribution company may fall behind.

Therefore, to improve the performance of the Indonesian pharmacy industry, there is a need to enhance the customer service of the pharmacy distribution company to its retailers. Although acquiring a new retailer is urgent, maintaining the loyalty of the current retailer is far more important, considering the lower cost compared to acquisition. Retaining a current retailer is critical because losing a retailer means losing all sources of revenue [3]–[5].

Nonetheless, retaining current retailer comes with challenges. One of the challenges we want to tackle is identifying the unique characteristics of each retailer. Tailoring retention strategy based on retailer's characteristics ensures the optimal effort given. The retailer also feels personally cared by the distribution company with the tailored approach [6]–[9]. With that, segmenting retailer based on their characteristic groups act as an initial stage to develop unique retention strategy in each retailer clusters.

Previous studies used unsupervised machine learning method such as K-means clustering, with sales data acquired from each retailer can be analyzed and clustered based on the parameters such as directorate item, principal, sales channel, customer name, ship to, month, and value [10]–[12]. Those data can be used to cluster retailers based on similar buying behavior. Combined with swarm intelligence algorithm such as ant Clustering, the retailer segmentation cluster can be generated [13]–[16]. Market segmentation is an approach to cluster market player with like-characteristics given the market's heterogenistic nature [17]. Its main purpose is to build customer model as potential promotion target and simplify the process of tailoring promotion strategy for each customer clusters. As targeted market segment often assures profit and relevance, a careful approach regarding segmentation process is needed to be done with optimal cost and result. One of the approaches is to identify RFM-location value (recency, frequency, monetary, and location) to better segment the market [10], [18], [19]. To analyze each segment, previous study proposed customer lifetime value and customer loyalty matrix using customer buying frequency and monetary value as the axis, as shown on Figure 1 [20]. While Chang and Tsay [21] proposed a more developed customer loyalty matrix which used recency, frequency, monetary, and length (location) to determine the customer segment.

Recency frequency monetary (RFM)-location is derived from basic RFM model which incorporate location as a new parameter. The location parameter is derived from the actual distance between main distribution company point and each retailer. This parameter is deemed important because the closer the distance, the cost of distribution will also become cheaper, turning to profits. Location acts as a factor of value based on the distribution cost and risk factor [18], [22]–[24]. Thus, RFM-location method is useful to give information of each generated clusters. Finally, the information from RFM-location can be used to identify which retailer cluster has the most customer lifetime value, giving the sense of priority in retaining the retailers in the clusters. Also, the strategy can be further tailored according to each characteristic in each cluster.

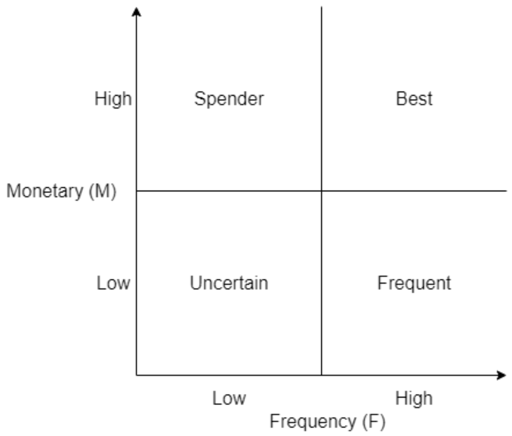

Figure 1. Customer loyalty matrix [20]

2. RESEARCH METHOD

This study consisted of several steps to answer the research question, which is, how to optimally segment pharmacy retailer and give relevant information to each segment in order to devise a tailored strategy. Data collection is the most important step in this study, by which we collected the data from a renowned Indonesian pharmaceutical company. However, the company mentioned is undisclosed by request. The steps taken are depicted on Figure 2 in details.

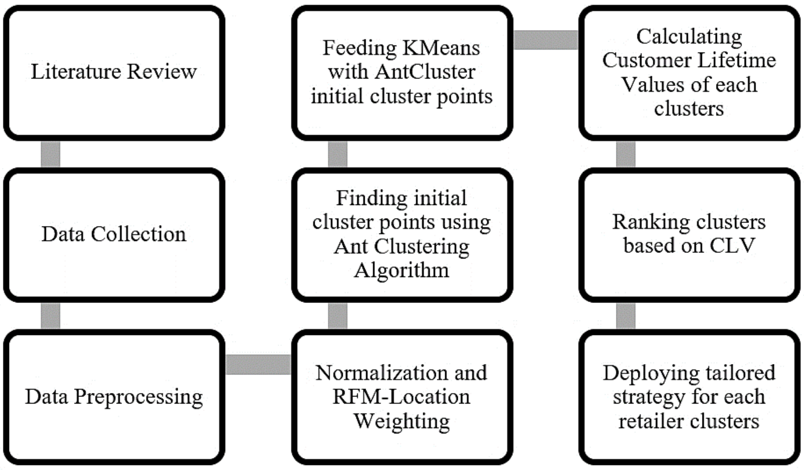

Figure 2. Research phases

2.1. Data collection and preprocessing

The data was collected through ethical method by asking one of the state-owned pharmacy distributor companies in Indonesia. Monthly sales act as the dataset for this study, the data consisted of seven columns which are directorate item, principal, sales channel, customer name (retailer), ship to, month, and value. The data was then cleaned for missing values and standardized for its entry, and then the attributes were selected based on their relevancy with the research purpose. The final dataset consisted of remaining five attributes, which are principal, customer name, ship to, month, and value.

2.2. Normalizing and RFM-location weighting

Generated clusters from unsupervised learning method often gives minimal information regarding the cluster, so, retailer segmentation needed to be done. To give specific information of each cluster, customer lifetime value was calculated using 4 parameters, which are recency, frequency, monetary, and location (RFM-location). The equation is as:

$$C^j = w_L C_L^j + w_R C_R^j + w_F C_F^j + w_M C_M^j$$

C^j is RFM-location valuation for cluster j, $C_R^j C_F^j C_M^j C_{Location}^j$ are normalized RFM-location for cluster J, and w_R w_F w_M $w_{Location}$ are RFM-location weights. Each retailers known RFM-location value from the dataset, combined with distance to the distributor acquired using Google maps, are then normalized. Min-Max method was used for the normalization process, it is then multiplied with each company-set weight for each retailer. The RFM-location weight set by the company were R=0.1, F=0.3, M=0.4, and L=0.2. The results of the multiplication are then used for the clustering process.

2.3. Clustering process

To put together retailers in like-characteristic groups, a hybrid clustering algorithm is employed. Ant K-means is a hybrid of ant clustering algorithm and K-means. By using ant clustering algorithm to identify initial cluster centers, the centroid for K-means algorithm can be further informed, resulting in better clusters. Ant colony optimization (ACO) is one of the methods in swarm intelligence which mimics ant behavior in seeking food source and homing to their nest by shortest path possible. Combined with K-means clustering algorithm, the algorithm works as depicted on Figure 3.

The hybrid algorithm is then coded using MATLAB and given required parameters, such as *Pdirection*, *Pdrop*, *Tcreate*, *Pload*, *Pdestroy*, *Tremove*, and *MaxIteration*. The program is then run 40 times to look for the optimal result. To evaluate the clustering task, the sum-of-squares within clusters (SSW) value is calculated, the smallest SSW value in resulting clustering runs is then chosen as the optimal generated clusters.

K-means by itself required a substantial amount of time to converge. Meanwhile, ant clustering algorithm has quick convergence without any prior information about the data [25]. By combining ant clustering algorithm and K-means, the clustering problem can converge faster and more efficient. Also, previous study found that ant K-means is a stable performing algorithm for solving a various industrial clustering problems with more promising results [26], [27].

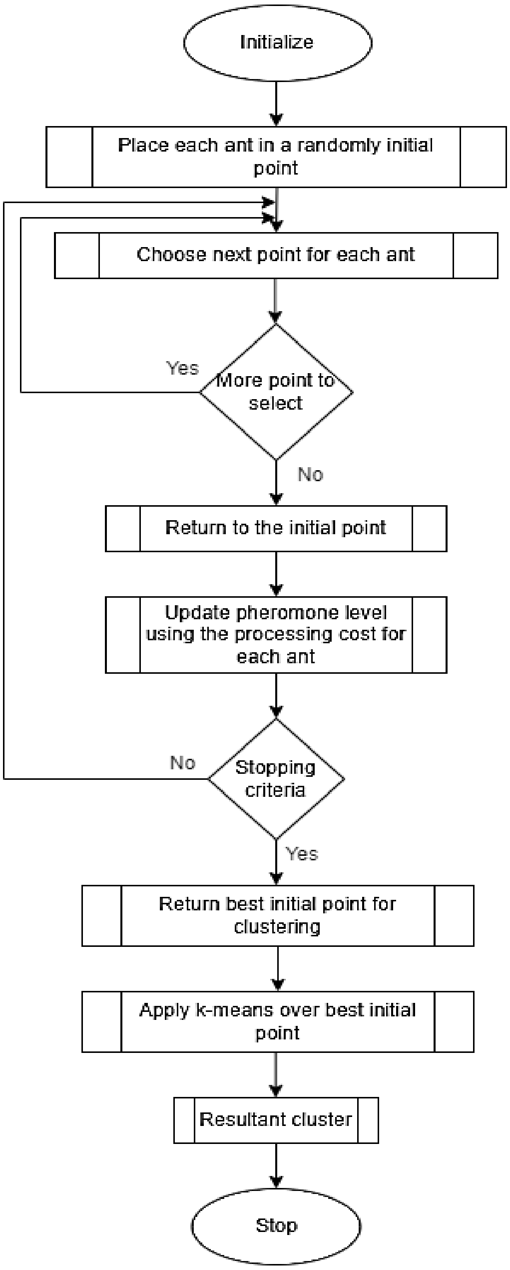

Figure 3. Ant K-means algorithm flowchart

2.4. Ranking retailer segment and deploying strategy

After clustering process is done and clusters are generated, they were then re-evaluated. The clusters were evaluated for its Dunn index and mean RFM-location value, and if it needs to be done, some clusters have to be re-combined based on its RFM-location value. This step is to achieve unique clustering, thus maximizing the effect of the tailored strategy.

3. RESULTS AND DISCUSSION

After the clustering is done, each process was calculated for its SSW value. Then, the clustering process that yielded 13 clusters were chosen with SSW=0.3848. To verify whether the clustering process is considered ideal, Dunn index is used to validate. Depicted on Figure 4, 13 clusters are optimal result with Dunn index value of 0.04321. Then the mean of RFM-location index value is calculated from the data of generated clusters. It is done to normalize each RFM-location variable from each generated cluster. The mean of RFM-location index value of each cluster is depicted on Table 1.

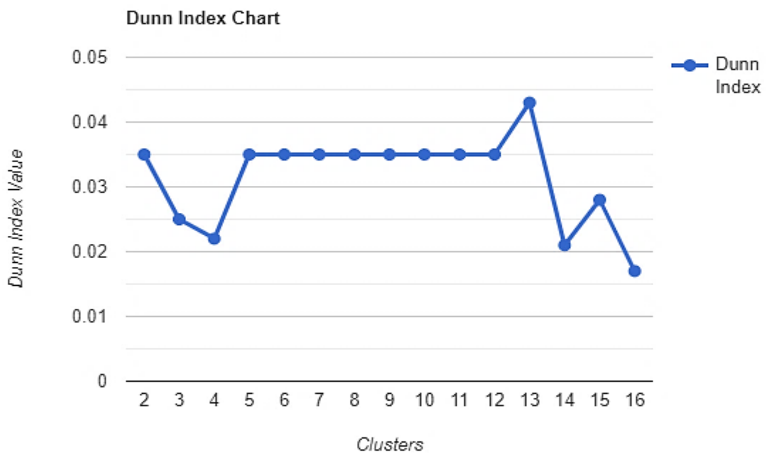

Figure 4. Dunn index chart

Table 1. Mean RFM-location index value

Cluster	C_R	C_F	C_M	C_L
Cluster 1	0.042245	0.098980	0.001223	0.007506
Cluster 2	0.039107	0.161786	0.005506	0.008020
Cluster 3	0.000087	0.294338	0.008711	0.002329
Cluster 4	0.031319	0.149231	0.003779	0.006639
Cluster 5	0.008759	0.121684	0.005805	0.003342
Cluster 6	0.052158	0.074330	0.000699	0.008130
Cluster 7	0.067842	0.008534	0.000302	0.005449
Cluster 8	0.012130	0.158195	0.001762	0.006798
Cluster 9	0.009926	0.241046	0.010742	0.004521
Cluster 10	0.009292	0.205749	0.003029	0.004457
Cluster 11	0.030496	0.191977	0.006365	0.008204
Cluster 12	0.006319	0.066758	0.001380	0.003003
Cluster 13	0.009057	0.016246	0.000482	0.003068

To ensure that the CLV calculation will yield more accurate rankings, a re-evaluation is needed to be done to check whether generated cluster already has unique pattern or not. Each cluster's RFM-location mean value is then compared with total cluster's RFM-location mean value. The total mean of RFM-location values is R=0.023, F=0.159, M=0.004, and L=0.005. The comparation shown that some clusters have identical characteristic with each other. Listed below are some clusters with indistinguishable characteristic: i) cluster 1, 4, and 6 with characteristic R↓F↓M↓L↓; ii) cluster 2 and 11 with characteristic R↓F↑M↑L↓; iii) cluster 3 and 9 with characteristic R↑F↑M↑L↑; and iv) cluster 12 and 13 with characteristic R↑F↓M↓L↑.

Hence, clusters with same characteristics were combined into new cluster, which yielded the same characteristic as the initial clusters. The clusters are now remained 8 clusters, and the details are depicted on Table 2. Then, each new clusters's RFM-Location mean value is recalculated to be used for calculating its CLV. The characteristics of the 8 clusters, its respective CLV, and rank based on CLVs are depicted on Table 3. The clusters RFM-location characteristics is then mapped to identify which characteristic stood out in each cluster. Using mapping matrix developed from [20], [21], the generated clusters are mapped and depicted on Figure 5.

From the matrix, there are five retailer segments identified to be used for developing tailored retention strategy. Those five segments are:

a. High value loyal customer: this group consisted of the most loyal retailer and having contributed the most profit for the distributor company. The retailers are also in close distance, and still actively buying in large nominal scale.
b. High value customer: this group consisted of retailers who are far away and did not make transaction recently. But these retailers often made large order and contributed big profit for the company, shown by high monetary and frequency value.
c. Low value loyal customer: this group consisted of retailers with high loyalty but yielded low profits. They are in close distance from the company, and still actively buying, but not in large scale.
d. Consumption resource customer: this group consisted of retailers who often spends big resource on shipping. It is due to the low monetary value in each order, but the location of these retailers is too far

away from the company. It is considered churn-prone, therefore the company needed to consider if this group is worth the retention or not.

e. Lost customer: this group consisted of retailers who are not actively buying anymore. The company should not bother to apply a retention strategy to this group.

Table 2. Cluster combination

Initial Cluster	Combined Cluster
1, 4, and 6	A
2 and 11	B
3 and 9	C
5	D
7	E
8	F
10	G
12 and 13	H

Table 3. Cluster CLV rankings

Cluster	Characteristic				CLV	Rank
	R	F	M	L		
A	↓	↓	↓	↓	0.041	6
B	↓	↑	↑	↓	0.061	3
C	↑	↑	↑	↑	0.088	1
D	↑	↓	↑	↑	0.050	5
E	↓	↓	↓	↑	0.010	8
F	↑	↓	↓	↓	0.050	4
G	↑	↑	↓	↑	0.064	2
H	↑	↓	↓	↑	0.014	7

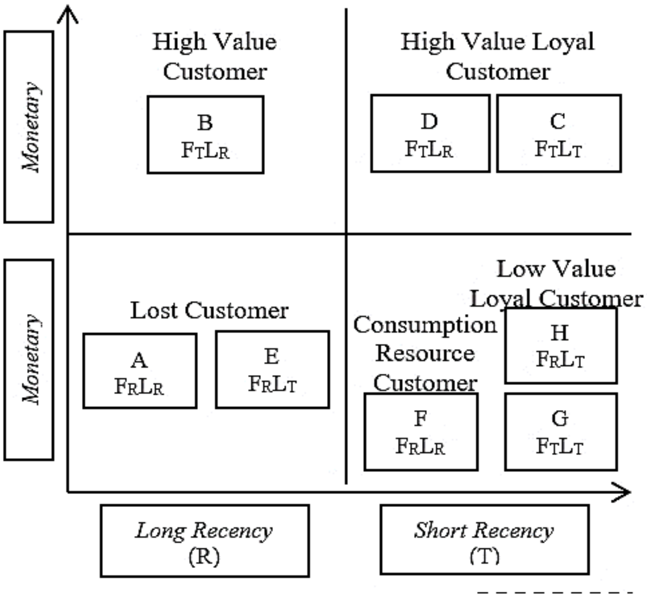

Figure 5. Retailer mapping matrix

Using unsupervised learning method for clustering retailers yielded fairly good result. Combined with ant optimization, the clustering with K-means resulted in good clusters, verified through Dunn index. From ant K-means, 13 clusters of retailers were generated with total of 1138 retailers in the sales data of the distributor company. However, using RFM-location model, these 13 clusters still have some similar characteristics with each other. Therefore, some clusters were combined and resulted in 8 clusters remaining with each unique characteristics per clusters. These findings leveraged the advantage of using ant clustering algorithm and the simplicity of K-means, so the initial cluster points can be determined optimally [15], [28]. While these two algorithms worked perfectly, the company's policy regarding how these generated clusters should be treated is the important thing.

The previous studies also used RFM model to determine the characteristics of retailer segment, but often neglect the important variable of shipping cost [10], [18], [22]. While in Indonesia, a vast country with broad sea in between and often under-developed inter-provincial roads, shipping cost can be the biggest threat to distributor's profit. Many retailers also consider shipping cost before placing an order and most of the time, shipping cost is the deciding factor of a buying order [29]. Thus, the addition of location factor in RFM-location method is essential [30], [31].

4. CONCLUSION

Using RFM-Location, some identical characteristics were identified from the generated 13 clusters and needed to be combined. After combining, 8 new clusters can become the focus of the marketing strategy development which is tailored to each segment of the retailers. Finally, referring to the customer segmentation matrix developed in the previous research, we determined 5 segments: i) high value loyal customer, ii) high value customer, iii) low value loyal customer, iv) consumption resource customer, and v) lost customer. Using these 5 segments, the distributor company can make tailored strategy for each segment and decide the priority of the strategy deployment.

ACKNOWLEDGEMENTS

In this section we wanted to say thank you to the data provider, PT. (undisclosed) the Indonesian pharmacy company.

REFERENCES

[1] BKPM, "Winning potentials in Indonesian pharma and medical industries," *Ministry of Investment/BKPM*, 2022. https://www.bkpm.go.id/en/publication/detail/news/winning-potentials-in-indonesian-pharma-and-medical-industries (accessed: Mar. 22, 2022).
[2] A. Secretariat, "ASEAN Statistical Yearbook 2020," *ASEAN*. 2020, Accessed: Aug. 20, 2020. [Online]. Available: https://www.aseanstats.org/publication/asyb_2020/
[3] D. Bateh, "Machine impact in supply chain management," *International Journal of Business Marketing and Management*, vol. 4, no. 5, pp. 32–37, 2019.
[4] N. Nikaein and E. Abedin, "Customers' segmentation in pharmaceutical distribution industry based on the RFML model," *Customers' segmentation in pharmaceutical distribution industry based on the RFML model*, vol. 37, no. 1, pp. 29–44, 2021.
[5] R. M. Cortez, A. H. Clarke, and P. V. Freytag, "B2B market segmentation: A systematic review and research agenda," *Journal of Business Research*, vol. 126, pp. 415–428, Mar. 2021, doi: 10.1016/j.jbusres.2020.12.070.
[6] S. Antwi, "'I just like this e-Retailer': Understanding online consumers repurchase intention from relationship quality perspective," *Journal of Retailing and Consumer Services*, vol. 61, Jul. 2021, doi: 10.1016/j.jretconser.2021.102568.
[7] A. Kompalla, A. Geldmacher, A. Just, and S. Lange, "Tailored automotive business strategies in the context of digitalization and service-oriented models," *Calitatea*, vol. 18, no. 156, pp. 77–84, 2017.
[8] V. Kumar and O. G. Ayodeji, "E-retail factors for customer activation and retention: An empirical study from Indian e-commerce customers," *Journal of Retailing and Consumer Services*, vol. 59, p. 102399, Mar. 2021, doi: 10.1016/j.jretconser.2020.102399.
[9] L. Leone, L. Haynes-Maslow, C. Kasprzak, S. Raja, and L. H. Epstein, "The WIC shopping experience: A qualitative study examining retail-based strategies to increase WIC retention and redemption rates," *Journal of Hunger & Environmental Nutrition*, vol. 17, no. 4, pp. 460–474, Jul. 2022, doi: 10.1080/19320248.2021.1915906.
[10] R. Gustriansyah, N. Suhandi, and F. Antony, "Clustering optimization in RFM analysis based on k-Means," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 1, pp. 470–477, Apr. 2020, doi: 10.11591/ijeecs.v18.i1.pp470-477.
[11] M. Pradana, "Maximizing strategy improvement in mall customer segmentation using K-means clustering," *Journal of Applied Data Sciences*, vol. 2, no. 1, pp. 19–25, Jan. 2021, doi: 10.47738/jads.v2i1.18.
[12] R. Vohra, J. Pahareeya, A. Hussain, F. Ghali, and A. Lui, "Using self organizing maps and K-means clustering based on RFM model for customer segmentation in the online eetail business," *In book: Intelligent Computing Methodologies*, pp. 484–497, 2020.
[13] M. S. Kumar and J. Prabhu, "A hybrid model collaborative movie recommendation system using K-means clustering with ant colony optimisation," *International Journal of Internet Technology and Secured Transactions*, vol. 10, no. 3, pp. 337–354, 2020.
[14] Z. Khan, S. Fang, A. Koubaa, P. Fan, F. Abbas, and H. Farman, "Street-centric routing scheme using ant colony optimization-based clustering for bus-based vehicular ad-hoc network," *Computers & Electrical Engineering*, vol. 86, Sep. 2020, doi: 10.1016/j.compeleceng.2020.106736.
[15] N. Kusumahardhini, G. F. Hertono, and B. D. Handari, "Implementation of K-means and crossover ant colony optimization algorithm on multiple traveling salesman problem," *Journal of Physics: Conference Series*, vol. 1442, 2017.
[16] A. K. Elfergany and A. Adl, "Identification of Telecom Volatile customers using a particle swarm optimized K-means clustering on their personality traits analysis," *International Journal of Service Science, Management, Engineering, and Technology*, vol. 11, no. 2, pp. 1–15, Apr. 2020, doi: 10.4018/IJSSMET.2020040101.
[17] M. J. Baker and S. J. Hart, *The marketing book*. Oxford; Boston: Butterworth-Heinemann, 2008.
[18] B. Rizki, N. G. Ginasta, M. A. Tamrin, and A. Rahman, "Customer loyality segmentation on point of sale system using recency-frequency-monetary (RFM) and K-means," *Jurnal Online Informatika*, vol. 5, no. 2, pp. 130–136, 2020.
[19] M. A. Rahim, M. Mushafiq, S. Khan, and Z. A. Arain, "RFM-based repurchase behavior for customer classification and segmentation," *Journal of Retailing and Consumer Services*, vol. 61, p. 102566, Jul. 2021, doi: 10.1016/j.jretconser.2021.102566.
[20] C. Marcus, "A practical yet meaningful approach to customer segmentation," *Journal of Consumer Marketing*, vol. 15, no. 5, pp. 494–504, Oct. 1998, doi: 10.1108/07363769810235974.

[21] H. H. Chang and S. F. Tsay, “Integrating of som and K-mean in data mining clustering: An empirical study of Crm and profitability evaluation,” *Journal of Information Management*, vol. 11, pp. 161–203, 2004.
[22] İ. Kabasakal, “Customer segmentation based on recency frequency monetary model: A case study in E-retailing,” *Bilişim Teknolojileri Dergisi*, vol. 13, no. 1, pp. 47–56, Jan. 2020, doi: 10.17671/gazibtd.570866.
[23] X. Lu, B. Rai, Y. Zhong, and Y. Li, “Cluster-based smartphone predictive analytics for application usage and next location prediction,” *International Journal of Business Intelligence Research*, vol. 9, no. 2, pp. 64–80, Jul. 2018, doi: 10.4018/IJBIR.2018070104.
[24] G. Johnson, R. A. Lewis, and D. Reiley, “Location, location, location: Repetition and proximity increase advertising effectiveness,” in *Available at SSRN 2268215*, 2016, pp. 1–41.
[25] J. Lu and R. Hu, “A new hybrid clustering algorithm based on K-means and ant colony algorithm,” *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*, 2013, doi: 10.2991/iccsee.2013.430.
[26] S. Dou, Z. Yao, X. Shi, and G. Liu, “Research on location and path planning of distribution center based on improved K-means clustering algorithm and improved ant colony algorithm,” in *LISS 2020*, Singapore: Springer Singapore, 2021, pp. 797–809.
[27] C. Srisa-an, “Location-based mobile community using ants-based cluster algorithm,” *International Journal for Applied Information Management*, vol. 1, no. 1, pp. 36–41, Apr. 2021, doi: 10.47738/ijaim.v1i1.6.
[28] H. Jiang, S. Yi, J. Li, F. Yang, and X. Hu, “Ant clustering algorithm with K-harmonic means clustering,” *Expert Systems with Applications*, vol. 37, no. 12, pp. 8679–8684, Dec. 2010, doi: 10.1016/j.eswa.2010.06.061.
[29] H.-B. Youn, S. Choo, and M. Park, “Segmentation of electronic commerce delivery service based on a stated preference survey,” *Journal of International Logistics and Trade*, vol. 12, no. 2, pp. 81–95, Aug. 2014, doi: 10.24006/jilt.2014.12.2.004.
[30] G. Balci and I. B. Cetin, “Market segmentation in container shipping services: a qualitative study,” *Management Research Review*, vol. 40, no. 10, pp. 1100–1116, Oct. 2017, doi: 10.1108/MRR-01-2017-0012.
[31] A. Parvaneh, M. J. Tarokh, H. Abbasimehr, and M. J. Tarokh, “Combining data mining and group decision making in retailer segmentation based on LRFMP variables,” *International Journal of Industrial Engineering & Production Researc*, vol. 25, no. 3, pp. 197–206, 2014.

BIOGRAPHIES OF AUTHORS

Ghea Sekar Palupi received her master’s degree in Information Management from National Taiwan University of Science and Technology. She is interested in data mining, specifically in supply chain management and customer relationship management. Her works included segmentation, forecasting, and gig economy worker-employer relationship. She currently teaches in the Department of Informatics at Universitas Negeri Surabaya, Indonesia. He can be contacted at email: gheapalupi@unesa.ac.id.

Muhammad Noor Fakhruzzaman was born and raised in Surabaya. He holds an interdisciplinary Master’s Degree in Human-Computer Interaction and Journalism & Mass Communication from Iowa State University. His current research interests fall between Data Science and Mass Communication, mainly automated media monitoring using Natural Language Processing. He currently teaches at Data Science Technology Study Program in Universitas Airlangga. His research includes Brain-Computer Interface, Indonesian Natural Language Processing, and Media Monitoring. He is a member of Kappa Tau Alpha, an honor society of Journalism and Mass Communication studies. He also loves to train in grappling sport: Wrestling, Brazilian Jiu-jitsu (2nd-degree blue belt) and fanatically watch pro-wrestling shows. He can be contacted at email: ruzza@stmm.unair.ac.id; ruzcmc@gmail.com.