International Proceeding Conference on Information Technology,
Multimedia, Architecture, Design, and E-Business (IMADE). Vol.2, August 2022
p-ISSN: 2747-1764, e-ISSN: 2747-1756 http://eprosiding.idbbali.ac.id/index.php/imade

CLUSTERING CUSTOMER FOR DETERMINE MARKET STRATEGY USING K-MEANS AND TOPSIS: CASE STUDY

Made Agung Raharja[1)], I Kadek Ari Surya [2)], I Komang Ari Mogi [3]

Department of Informatics, Faculty of Mathematics and Natural Sciences, Udayana University
[1)]made.agung@unud.ac.id
[2)]ari.surya@student.unud.ac.id
[3)] arimogi@unud.ac.id

ABSTRACT

At this time the characteristics of customer groups are widely used as a reference in determining marketing strategies by companies in the midst of intense market competition. The research stage starts from preprocessing the data and then calculating the Customer Lifetime Value (CLV) through weighting the RFM (Recency, Frequency, Monetary) and LRFM (Length, Recency, Frequency, Monetary) models. The weights are given by experts with the Analytic Hierarchy Process (AHP) calculation. Then K-Means is used for the clustering process and then evaluated with the Silhouette Coefficient (SC). The best clustering results are ranked using the Technique for Orders Preference by Similarity to Ideal Solution (TOPSIS) method to produce alternative decisions in cluster selection. This research takes a case study in one of the handicraft shops in Bali which experienced a decline in product demand due to the impact of intense market competition. The results show that the SC value tends to increase along with the addition of variations in the percentage of data. The weighting of the data model from the case study expert weight resulted in the best SC value at k = 2 with the value of the RFM model of 0.665 (medium structure) and the LRFM value of 0.641. The cluster ranking validation test and Rank Consistency on the best clustering results get valid calculation results and there is no rank reversal. The recommendation for customer clusters is that Cluster 2 in the first place has 158 customers with cluster type R↓F↑M↑ with Enforced Strategy and Cluster 1 in the second place has 817 customers with cluster type R↑F↓M↓ with Let-go Strategy.

Keywords: *clustering, customer, K-Means, TOPSIS, silhouette coefficient, craft, RFM, LRFM*

INTRODUCTION

At this time the customer becomes one of the important aspects in the business world. Intense market competition makes companies compete to attract and retain customers. This is because establishing good relationships with customers will keep customers from switching to using competitors products. The tight market competition certainly also affects the business sector in the industry, especially those based on a regional basis.

The handicraft industry is one sector that makes a major contribution to the economy in the province of Bali. Regional economic recovery must be supported by the active role of the community and government in order to produce better economic conditions (Raharja *et al.*, 2021). Data from the Bali Province Trade and Industry Office (Disperindag) in 2019 shows that the percentage contribution of the export value of industrial and handicraft products of Bali's total non-oil exports reached 68.29%. However, in 2019 the export value of handicraft production decreased by 13% (Disperindag Bali, 2019). In 2020 there was also a larger decline of 29.44% (Disperindag Bali, 2020). The causes include declining market demand, inappropriate designs, uncompetitive prices, the trade war between America and

China (Disperindag Bali, 2020). This has an impact on the handicraft business sector in Bali, one of which is the handicraft shop oleh2bali.com which has experienced a decline in order demand in recent years.

Grouping of customers is needed to find out consumer behavior so that it will help in implementing the right marketing strategy to increase company revenue and the use of data mining techniques is one solution to this problem (Adiana, Soesanti and Permanasari, 2018). The clustering method can group customers who have the same Customer Lifetime Value (CLV). CLV has been used in many studies and has been applied for many purposes such as evaluating customers, customer segmentation, product recommendations, marketing and sales strategies (Azadnia *et al.*, 2011). Dewi et al (2013) said the RFM (recency, frequency, monetary) model can be used to calculate the CLV value using clustering. In addition, there is also the development of another model, namely the LRFM model (length, recency, frequency, monetary) which can be used to identify CLV values in developing strategies for customer groups (Ditendra *et al.*, 2020).

Research conducted by Sembiring Brahmana, Mohammed and Chairuang (2020) on the comparison of the K-Means, K-Medoids, and DBSCAN algorithms shows that the K-Means algorithm has the best accuracy. According to Azadnia *et al.* (2011) after obtaining customer clusters, CLV rating is one of the important issues for companies to develop the right strategy to retain customers, identify and compare market segments. Azadnia *et al.* (2011) continued, in his research clusters were ranked using the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method.

Based on the explanation above, this study aims to determine the characteristics of the recommendation of the best customer cluster in the research case study through the clustering process using K-Means and cluster ranking using TOPSIS. In addition, to determine the level of accuracy of the results of cluster testing using the Silhouette Coefficient and determine the level of accuracy of customer cluster ratings with RFM and LRFM data models. The limitation of this research is that the evaluation of the marketing strategy is not used as an evaluation value in this study.

RESEARCH METODOLOGY

Data Collection

This study uses secondary data obtained from the handicraft shop oleh2bali.com in Bali. The data is the history of customers in placing orders in 2014 - 2020 as many as 975 customer data records. Name and email data are anonymous (because of the code of ethics). The data format is comma separated values (.csv). In addition, it also uses 2463 order data.

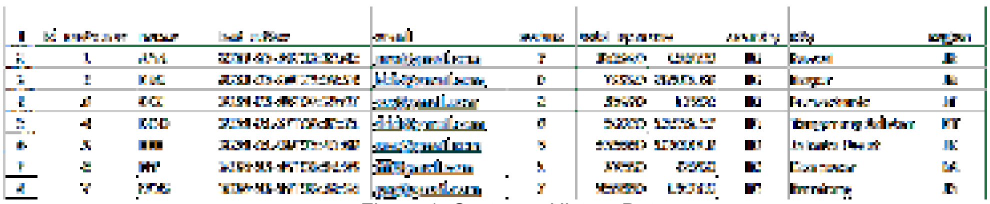
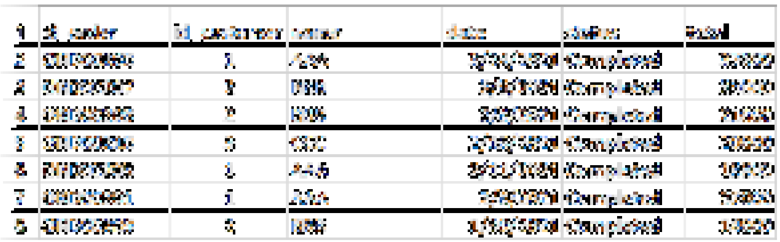


Figure 1. Customer History Data
Source: Author, 2022



	id_order	id_customer	name	date	status	total
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Completed	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Completed	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Completed	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Completed	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Completed	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Completed	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Completed	[illegible]

Figure 2. Order Data
Source: Author, 2022

Data Preprocessing

In this study, preprocessing data consists of data cleaning, data selection and data transformation. In data cleaning, data records are deleted that have null or missing values. In the data selection, several data attributes that are needed in the study are selected, especially to obtain the RFM and LRFM data models. Recency (R) is the duration or period of the last purchase with the current time. Frequency (F) is the total number of product purchases made within a certain period of time. Monetary (M) is the total value of purchases spent during a certain period. Length (L) states the length of the interval from the initial transaction and the last customer transaction in a certain period. After getting the values for each data model, preprocessing is continued on the data transformation. In the data transformation, the range of attribute values is equated between 0 – 1 using the Min Max Normalization method.

Calculation of Model Weight Using AHP

Analytic Hierarchy Process (AHP) is a decision support method developed by Thomas L. Saaty. Although AHP was developed to solve complex MCDM problems. However, due to its flexibility and practicality, AHP is applied by various researchers in various fields including planning and development, selection, allocation, decision making, and ranking or prioritizing. The weight assessment in this study was given by an expert, namely the owner of a case study craft shop who has a background in technology and business management. The final weights generated on the RFM model are R: 0.0593, F: 0.4507, M: 0.4899. While the LRFM model is L: 0.0405, R: 0.0747, F: 0.3155, M: 0.5691.

Calculating CLV Using RFM and LRFM Weighting

Customer Lifetime Value (CLV) is an understanding of the present value of all future benefits derived from customers (Angelie, 2017). CLV can be calculated by RFM and LRFM weighting. The following is the equation for calculating CLV with RFM weighting (Angelie, 2017):

$$CLV_{ci} = NR_{ci}\ x\ WR_{ci} + NF_{ci}\ x\ WF_{ci} + NM_{ci}\ x\ WM_{ci} \qquad (1)$$

Where:
$NRci$, $NFci$, $NMci$: is the normalized value of recency, frequency, and monetary.
$WRci$, $WFci$, $WMci$: is the weight value of recency, frequency, and monetary.

Then next is the CLV calculation equation with LRFM weighting (Monalisa, 2018):

$$C^j = W_L\ C_L + W_R\ C_R + W_F\ C_F + W_M\ C_M \qquad (2)$$

Where:
C^j : CLV of customer j
C_L, C_R, C_F, C_M : Normalization of L, R, F and M
W_L, W_R, W_F, W_M : Weight of L,R,F and M

To determine the effect of using the weights of the data model on the cluster results, this study uses three weights. Two weights are weights that come from previous research that has the same business field as the research case study. The following are the weights used in this study.

Table 1. Weight RFM Model

Weight Name	Weight Source	Data Attribute		
		Recency (Wr)	Frequency (Wf)	Monetary (Wm)
Weight 1	(Azadnia et al., 2011)	0.0970	0.3446	0.5583
Weight 2	(Dewi et al., 2013)	0.059	0.28	0.67
Weight 3	Case Study Expert	0.0593	0.4507	0.4899

Source: Author, 2022

Table 2. Weight LRFM Model

Weight Name	Weight Source	Data Attribute			
		Length (Wl)	Recency (Wr)	Frequency (Wf)	Monetary (Wm)
Weight 1	Parvaneh, Abbasimehr and Tarokh in (Monalisa, 2018)	0.238	0.088	0.326	0.348
Weight 2	(Pramono, Surjandari and Laoh, 2019)	0.222	0.182	0.305	0.292
Weight 3	Case Study Expert	0.0405	0.0747	0.3155	0.5691

Source: Author, 2022

Clustering Customer Using K-Means

K-Means is a non-hierarchical clustering method that partitions data into one or more clusters (Raharja and Supriana, 2019). K-Means has the ability to group large amounts of data with fast and efficient computation time. The following are the steps of the clustering method using the K-Means algorithm:

1. Enter the value of the number of clusters as many as k clusters, and a dataset that you want to group.
2. A total of k data is selected as the initial centroid. In this study, the initial centroid was chosen randomly.
3. Calculate the distance of each data object point to each centroid. Distance calculations can be done using the Euclidian Distance measure equation:
$$d(x,y) = |x-y| = \sqrt{\sum_{i=1}^{n} d\,(x_i - y_i)^2} \quad (3)$$
4. Allocate each data object into a cluster with the minimum distance.
5. Calculate the average of all data contained in the cluster as the center of the new cluster.
6. Repeat steps 3, 4, and 5 until no more objects have changed in a cluster.

Cluster Evaluation Using Silhouette Coefficient

The Silhouette Coefficient was introduced by Rousseeuw in 1987. The Silhouette Coefficient is used to determine the quality of the data clusters generated by the clustering algorithm by combining the concepts of cohesion and separation. The calculation stage is as follows:

1. Find the average distance of an object i to all objects other than the object in a cluster a(i).
2. Then find the average distance from the object to all other objects in different clusters, then select the smallest average distance value b(i).
3. After the a(i) and b(i) values are obtained, then the silhouette coefficient can be calculated from the object to i.

$$s(i) = \frac{b(i)-a(i)}{(a(i),b(i))} \quad (4)$$

Based on the category of the average value of the silhouette coefficient according to Rousseeuw, the results of the calculation of the silhouette coefficient value which is close to the value of 1 indicates that the quality of the cluster is getting better.

Cluster Ranking Using TOPSIS

TOPSIS is a multi-criteria decision-making method that uses the principle that the chosen alternative must have the closest distance from the positive ideal solution and the farthest from the negative ideal solution. TOPSIS method stages (Hastuti, Utami and Luthfi, 2013):

1. Build a decision matrix referring to the alternatives that will be evaluated based on the criteria.
2. Build a normalized decision matrix.
3. Create a weighted normalized decision matrix.
4. Determine the positive ideal solution matrix and the negative ideal solution matrix.
5. Determine the distance between the values of each alternative with the positive ideal solution matrix and the negative ideal solution matrix.
6. Determine the preference value for each alternative (Vi). Alternatives can be sorted by order of Vi. A larger Vi value indicates that the alternative is preferred.

Cluster Rank Validation and Rank Consistency

The cluster rank obtained from the TOPSIS calculation was tested for validation and consistency. The validation of the cluster ranking calculation can be done by comparing the cluster ratings generated by the system with manual calculations (Usman Adi Nugroho, Ratih Kartika Dewi, 2019). Then the consistency of customer cluster ratings is measured through Rating Consistency. Rodrigues et al. in (Usman Adi Nugroho, Ratih Kartika Dewi, 2019) Rank Consistency is a test used to determine the level of consistency of an algorithm by seeing whether the output results remain consistent or different after a change in alternative data or criteria. The test is carried out by comparing the cluster rankings generated by the system with the results of cluster rankings that have added an alternative originating from a duplicate of one of the clusters. The test is said to be successful if no rank reversal (change of position in the first and last rank) is generated.

Customer Cluster Evaluation

The results of the customer cluster are evaluated based on the attributes of the data model. From the centroid of each cluster, the type of customer cluster can be evaluated by checking the average attribute values of length, recency, frequency, monetary are above average (L↑, R↑, F↑, M↑) or below average (L , R↓, F↓, M↓). Based on the type of each cluster, program references and suggestions for marketing strategy activities based on frequency and monetary values (Dewi *et al.*, 2013) can be determined which consist of Enforced Strategy (F↑ M↑), Offensive Strategy (F↑ M↓), Defensive Strategy (F M↑), Let-go Strategy (F M↓). In addition, an analysis is also carried out on the highest percentage of customers on geographical conditions, order time and transactions to strengthen the marketing strategy that can be carried out.

RELATED RESEARCH

The customer grouping process can be done with several algorithms. In a study conducted by (Sembiring Brahmana, Mohammed and Chairuang, 2020) comparing the K-Means, K-Medoids, and DBSCAN algorithms to group customers using the RFM model, the results show that K-Means has the best validity level with the DBI result being 0.33009058 and the silhouette index is 0.912671056. In addition, in research (Monalisa, 2018) K-Means is also used to classify customer lifetime value by using the LRFM model. In evaluating the quality of the cluster there are several methods that can be applied. In research (Dewi and Pramita, 2019) the results of the silhouette method resulted in a better evaluation with a DBI value of 1.06 compared to the elbow method with a DBI value of 1.10 in the case of grouping handicraft products in Bali. After getting the final results of customer clusters, several studies tried to integrate them with decision-making methods. Research conducted by (Azadnia *et al.*, 2011) proposes a customer value assessment model that is integrated with the multi-criteria decision-making method and the fuzzy clustering method. Fuzzy Analytical Hierarchy Process (FAHP) is used to calculate the weight of the RFM variable. Then Fuzzy C-Means (FCM) is used to group customers and then TOPSIS is used to rank customer lifetime value. In addition, similar types of research were also carried out by (Dewi *et al.*, 2013) but the weighting of the RFM model using AHP and generating customer segments was mapped into 4 marketing strategies, namely Enforced Strategy, Offensive Strategy, Defensive Strategy, and Let-go Strategy. This marketing strategy is based on the frequency and monetary values of each cluster.

RESULTS AND DISCUSSION

Silhouette Coefficient Based on Data Percentage Variation

Testing at this stage is carried out by using variations in the percentage of RFM and LRFM data from 10% to 100%. The aim is to determine the effect of using a variation in the percentage of data on the resulting Silhouette Coefficient (SC). The following is a graph of the test results:

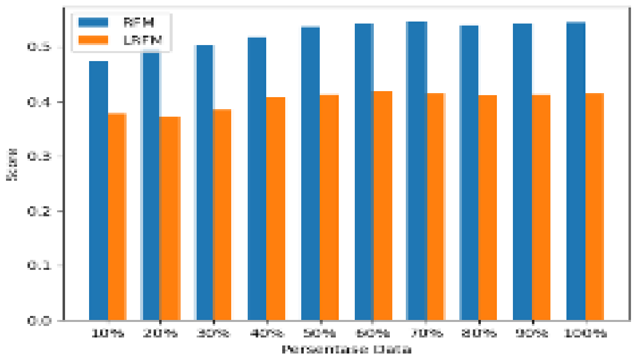


Figure 3. Silhouette Coefficient on Variation in Data Percentage
Source: Author, 2022

The graph shown shows that the RFM data model always produces a higher SC score than the LRFM data model. Along with the increasing variation in data volume, the largest SC scores generated in the RFM and LRFM data models both tend to increase. So that the next process uses all RFM and LRFM data. The next step is to test the weighting of the data model.

Silhouette Coefficient Based on Weighted Data Model

This test is carried out using three weights of the data model and with the number of k = 2 to k = 6. The aim is to determine the weight of the model and the best number of k.

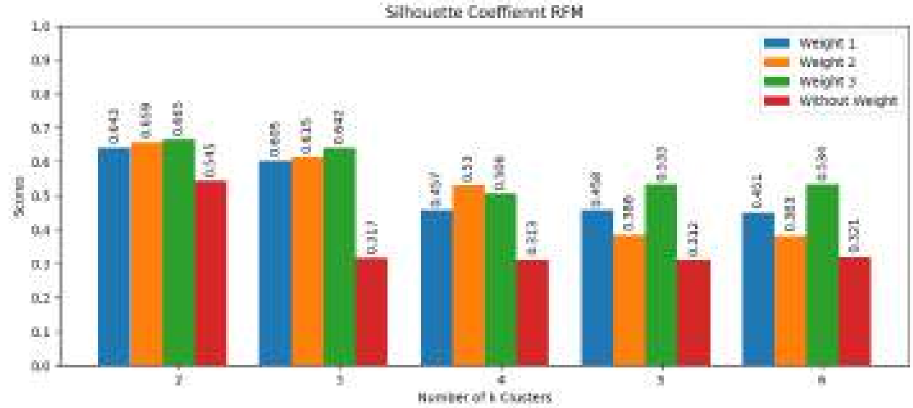


Figure 4. Silhouette Coefficient on Weighting RFM Model
Source: Penulis, 2022

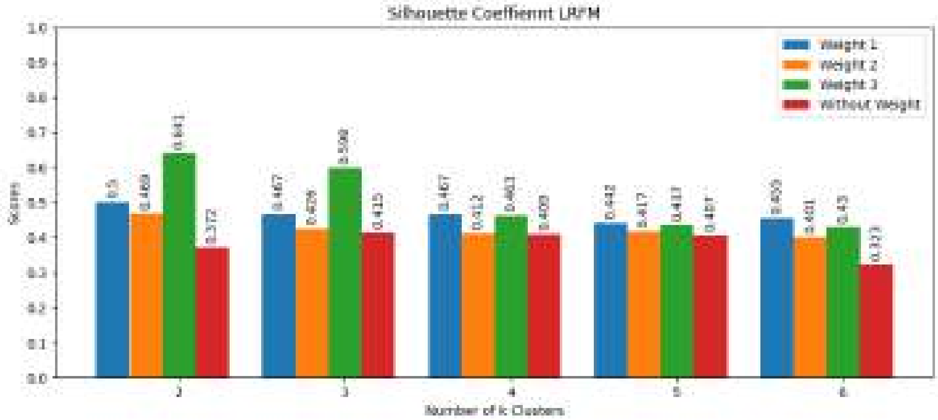


Figure 5. Silhouette Coefficient on Weighting LRFM Model
Source: Author, 2022

The weight of the data attribute model that produces the highest SC score is weight 3, namely the weight given by the expert at the case study site. The highest SC score in the RFM model is 0.665 which states that the results of the clustering evaluation are in the medium structure category. While the SC score on the LRFM model is 0.641 with the results of the clustering evaluation also being included in the medium structure category. Based on this comparison, it can be said that the best number of customer clusters in the case study craft shop were two clusters (k=2) which were obtained when using the RFM data attribute model and with the weight given by the case study expert. In the next stage, the final centroid of the best clustering results on the RFM and LRFM data attribute models is used for the TOPSIS process to determine the selection of the best alternative cluster through cluster ranking.

Cluster Ranking with TOPSIS

Table 3. Cluster Ranking

Model	Alternative	Preference	Ranking
RFM	Cluster 1	0.0	2
	Cluster 2	0.1	1
LRFM	Cluster 1	0.00202	2
	Cluster 2	0.81286	1

Source: Author, 2022

The preference value shows the ranking results of the alternatives. Based on the above calculations, the cluster in the RFM data model, namely cluster 2, is ranked first and cluster 1 is ranked second. Meanwhile, the cluster in the LRFM data model, namely cluster 2, also ranks first and cluster 1 ranks second.

Cluster Rank Validation Results and Rank Consistency

Table 4. Validation Cluster Rangking

Models	Rangking	Calculation Results		Status
		System	Manual	
RFM	1	Cluster 2	Cluster 2	Valid
	2	Cluster 1	Cluster 1	Valid
LRFM	1	Cluster 2	Cluster 2	Valid
	2	Cluster 1	Cluster 1	Valid

Source: Author, 2022

In the validation test results above, it can be seen that the cluster ranking results obtained by the system and manual calculations are all the same, both in the RFM and LRFM data attribute model clusters. Next is continued with Rank Consistency testing.

Table 5. RFM Consistency Rank Test Results

Alternative Addition	Test Ranking Results	Initial Ranking Results	Rank Reversal
Alternative Cluster 1	First : Cluster 2	First : Cluster 2	No
	Last : Cluster 1	Last : Cluster 1	No
Alternative Cluster 2	First : Cluster 2	First : Cluster 2	No
	First : Cluster 1	Last : Cluster 1	No

Source: Author, 2022

Table 6. LRFM Consistency Rank Test Results

Alternative Addition	Test Ranking Results	Initial Ranking Results	Rank Reversal
Alternative Cluster 1	First : Cluster 2	First : Cluster 2	No
	Last : Cluster 1	Last : Cluster 1	No
Alternative Cluster 2	First : Cluster 2	First : Cluster 2	No
	First : Cluster 1	Last : Cluster 1	No

Source: Author, 2022

Rank consistency testing is done by adding a new alternative to the existing alternative data. Evaluation of rank consistency on RFM and LRFM data is a step to check whether a rank reversal occurs or not in the ranking results generated from the test. Rank reversal is a change in the first and last rank of the alternative. The evaluation results show that there is no rank reversal when adding alternative cluster 1 and alternative cluster 2 when compared to the results of the initial cluster ranking.

Characteristics of Cluster Recommendation Results

Based on several tests that have been carried out, the characteristics of the best customer recommendations generated by the RFM model because it produces the highest SC value using all customer data and weights from experts in the case study and all cluster ranking calculations are valid and there is no ranking reversal.

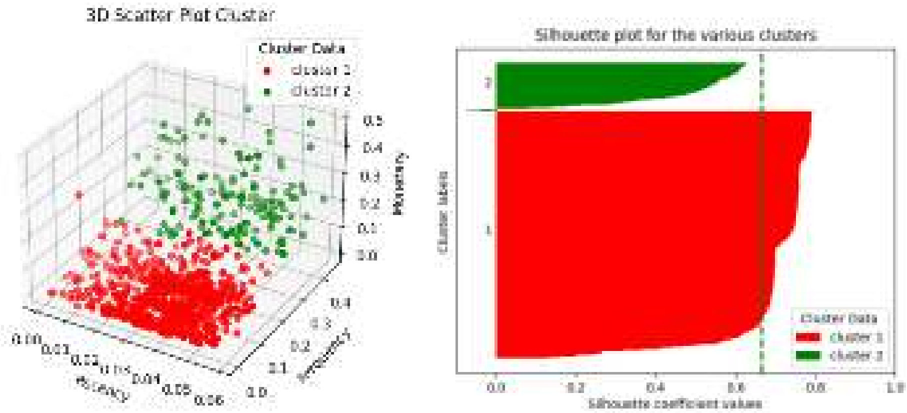


Figure 6. Visualization and Cluster Membership
Source: Author, 2022

The picture above is a visualization and cluster membership on the best clustering results. It can be seen that there are not too many overlapping data objects and some data objects have small SC values below zero. The results of the recommendation of the best number of clusters when the value of k = 2 by producing Cluster 2 as the first rank (the best cluster) and Cluster 1 as the second rank. Then the following is an evaluation of the characteristics of each cluster:

a. Cluster 2 is a recommendation for the first rank customer cluster that has priority to be considered. Cluster 2 has a cluster type R↓F↑M↑ with 158 customer members. The R↓F↑M↑ cluster type makes transactions at the latest time and has above average frequency and monetary values. The strategy applied based on research (Dewi *et al.*, 2013) is the Enforced Strategy (F↑M↑) with a marketing strategy program to maintain communication with customers, maintain long-term interactiveness, design customer loyalty programs, and understand customer needs and habits. Meanwhile, suggestions for marketing activities are sending promotional information by telephone, fax and email, and providing discounts.
   Based on geographic analysis, one of the promos that can be given is in the form of discounted shipping costs, especially for orders that are mainly aimed at the DKI Jakarta, West Java, and Banten areas. Then on the analysis of order time, and transaction time, discounts on certain products can also be given to customers, especially in October, November and December to be precise on the middle of the month. Then a good time to give promos is in the morning around 8 am to 10 am.

b. Cluster 1 as the second ranked customer cluster has a cluster type R↑F↓M↓ with 817 customer members. The R↓F↓M↓ cluster type makes transactions for a long time and has below average frequency and monetary values. The strategy applied based on research (Dewi *et al.*, 2013) is the Let-go Strategy (F↓ M↓) with a marketing strategy program that does not require companies to pay attention to this segment, and choose products with the main focus needed by customers. Suggestions for marketing activities separate new and old customers and communicate only with new customers. For new customers, occasional promos and discounts can also be given.
   Based on geographic analysis, the promo discount on shipping costs is mainly for orders that are mainly aimed at the DKI Jakarta, West Java, and East Java areas. Then based on the analysis of the time of orders and transactions that discounts on certain products can also be given to customers, especially in October, August and July to be exact on the middle of the month. A good time to give promos is during the day around 11 am to 2 pm.

CONCLUSION

Based on several tests of the Silhouette Coefficient (SC) and cluster rankings, the results are fairly good. The SC value tends to increase along with the addition of variations in the percentage of data. The weighting of the data model from the weight of the expert where the case study produces the best SC value at k = 2 with the value of RFM of 0.665 (medium structure) and the value of LRFM with a value of 0.641. RFM becomes the data model that produces the best cluster in the case study craft shop. The cluster ranking validation test and Rank Consistency get the results of all valid calculations and there is no rank reversal.

There are two customer clusters from the results of the K-Means and TOPSIS processes with their respective characteristics. Cluster 2 as the first rank (the best cluster) has a membership of 158 customers with cluster type R↓F↑M↑ and applies the Enforced Strategy. Most of the customers in this cluster come from DKI Jakarta, West Java, Banten with the habit of placing orders in October, November, December around the morning. Then Cluster 1 as the second rank has 817 customers with cluster type R↑F↓M↓ and applies the Let-go Strategy. Most of the customers in this cluster come from East Java, West Java, and DKI Jakarta with the habit of placing orders in July, August, October around noon.

REFERENCE

Adiana, B.E., Soesanti, I. and Permanasari, A.E. (2018) ‘Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model Dan Teknik Clustering’, *Jurnal Terapan Teknologi Informasi*, 2(1), pp. 23–32. doi:10.21460/jutei.2018.21.76.

Angelie, A.V. (2017) *Segmentasi Pelanggan Menggunakan Clustering K-Means Dan Model Rfm (Studi Kasus : Pt . Bina Adidaya Surabaya)*. ITS. Available at: https://core.ac.uk/download/pdf/291463031.pdf.

Azadnia, A.H. *et al.* (2011) ‘Integration model of Fuzzy C means clustering algorithm and TOPSIS method for customer lifetime value assessment’, in *IEEE International Conference on Industrial Engineering and Engineering Management*. doi:10.1109/IEEM.2011.6117870.

Dewi, D.A.I.C. and Pramita, D.A.K. (2019) ‘Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali’, *Matrix : Jurnal Manajemen Teknologi dan Informatika*, 9(3), pp. 102–109. doi:10.31940/matrix.v9i3.1662.

Dewi, D.P. *et al.* (2013) ‘Evaluasi dan Rekomendasi Peningkatan Nilai Perilaku Pembelian Pelanggan Berdasarkan Analisis RFM Perilaku Pembelian Pelanggan : Studi Kasus PT. XYZ’, 2(1).

Disperindag Bali (2019) ‘Laporan Kinerja Instansi Pemerintah (LKjIP) Dinas Perdagangan dan Perindustrian Provinsi Bali 2019’, *Dinas Perdagangan dan Perindustrian Provinsi Bali* [Preprint]. Available at: https://disperindag.baliprov.go.id/lkjip-2019-dinas-perindustrian-dan-perdagangan-provinsi-bali/.

Disperindag Bali (2020) ‘Laporan Kinerja Instansi Pemerintah (LKjIP) Dinas Perdagangan dan Perindustrian Provinsi Bali 2020’, *Dinas Perdagangan dan Perindustrian Provinsi Bali* [Preprint]. Available at: https://disperindag.baliprov.go.id/wp-content/uploads/2021/03/LKJiP-Disperindag-2020.pdf.

Ditendra, E. *et al.* (2020) ‘KLASTERISASI CLV DENGAN MODEL LRFM MENGGUNAKAN ALGORITMA FUZZY C-MEANS (Studi Kasus: Pangeran Gym Pekanbaru)’, *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, 6(1), pp. 109–116.

Hastuti, A.B., Utami, E. and Luthfi, E.T. (2013) ‘Implementasi Metode Fuzzy C-Means Dan Topsis Dalam Membangun Sistem Pendukung Keputusan (Studi

Kasus : Penentuan Jurusan Di Sma Negeri 1 Wonosari)', *Jurnal Dasi*, 14(2).

Monalisa, S. (2018) 'KLASTERISASI CUSTOMER LIFETIME VALUE DENGAN MODEL LRFM MENGGUNAKAN ALGORITMA K-MEANS', *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 5(2), pp. 247–252. doi:10.25126/jtiik.201852690.

Pramono, P.P., Surjandari, I. and Laoh, E. (2019) 'Estimating customer segmentation based on customer lifetime value using two-stage clustering method', in *2019 16th International Conference on Service Systems and Service Management, ICSSSM 2019*. IEEE, pp. 1–5. doi:10.1109/ICSSSM.2019.8887704.

Raharja, M.A. *et al.* (2021) 'Analysis of membership function in implementation of adaptive neuro fuzzy inference system (ANFIS) method for inflation prediction', *Journal of Physics: Conference Series*, 1722(1). doi:10.1088/1742-6596/1722/1/012005.

Raharja, M.A. and Supriana, I.W. (2019) 'Analisis Klasifikasi Tinggkat Kesehatan Lembaga Perkreditan Desa (Lpd) Menggunakan Metode K-Means Clustering', *Jurnal Teknologi Informasi dan Komputer*, 5(1), pp. 83–90. doi:10.36002/jutik.v5i1.641.

Sembiring Brahmana, R.W., Mohammed, F.A. and Chairuang, K. (2020) 'Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods', *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 11(1). doi:10.24843/lkjiti.2020.v11.i01.p04.

Usman Adi Nugroho, Ratih Kartika Dewi, M. (2019) 'Implementasi TOPSIS pada Sistem Rekomendasi Tempat Pembelian Sayuran Organik di Malang Berbasis Lokasi', *Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(5), pp. 4680–4686.