

Pemodelan Cluster Loyalitas Customer Menggunakan Algoritma K-Means Dengan Parameter LRIFMQ

by Anik Vega Vitianingsih

Submission date: 26-Jun-2020 11:00PM (UTC+0800)

Submission ID: 1239491990

File name: 2691-8250-1-SM.docx (2.66M)

Word count: 3199

Character count: 19435

Pemodelan *Cluster* Loyalitas Customer Menggunakan Algoritma *K-Means* Dengan Parameter LRIFMQ

Aloysius Matz

Institut Sains dan Teknologi Terpadu Surabaya

aloyusmatz@gmail.com

Abstract—Loyal customers are one of the factors that determine the development of a business. Therefore, we need a strategy to keep customers loyal, even making customers who were previously less loyal become more loyal. The strategy used must be right on target according to customer segmentation. In this paper, customer segmentation is done using the k-means algorithm with LRIFMQ (length, recency, interval, frequency, monetary, quantity) as parameters, and the CLV (customer lifetime value) of each cluster is calculated. Data obtained from PT. XYZ (a company engaged in food processing) for 1 year (1 January 2019 - 31 December 2019), with 337.739 transactions, and 26.683 customers. AHP (analytical hierarchy process) method is used for LRIFMQ weighting. Silhouette coefficient is used to calculate the cluster quality and determine the optimal number of clusters. The best results are obtained with the silhouette coefficient value of 0,71616 with the number of clusters 2, and the best CLV of 0,77837.

Keywords—customer analysis, k-means, LRIFMQ, analytical hierarchy process, silhouette coefficient.

Abstrak—Pelanggan yang loyal adalah salah satu faktor yang menentukan kemajuan suatu bisnis. Oleh karena itu, perlu suatu strategi untuk membuat pelanggan tetap loyal, bahkan membuat pelanggan yang sebelumnya kurang loyal menjadi lebih loyal. Strategi yang digunakan harus tepat sasaran sesuai dengan segmentasi pelanggan. Pada paper ini, segmentasi pelanggan dilakukan menggunakan algoritma *k-means* dengan LRIFMQ (*length, recency, interval, frequency, monetary, quantity*) sebagai parameter, dan dihitung nilai CLV-nya (*customer lifetime value*) dari tiap kluster. Data didapatkan dari PT. XYZ (perusahaan yang bergerak di bidang pengolahan makanan) selama 1 tahun (1 Januari 2019 - 31 Desember 2019), sebanyak 337.739 transaksi, dan 26.683 pelanggan. Metode AHP (*analytical hierarchy process*) digunakan untuk pembobotan LRIFMQ. *Silhouette coefficient* digunakan untuk menghitung nilai kualitas kluster dan menentukan jumlah kluster optimal. Didapatkan hasil terbaik dengan nilai *silhouette coefficient* 0,71616 dengan jumlah kluster 2, dan nilai CLV terbaik 0,77837.

Kata kunci—analisa customer, *k-means*, LRIFMQ, *analytical hierarchy process*, *silhouette coefficient*.

I. PENDAHULUAN

Salah satu aset tidak berwujud yang dimiliki suatu perusahaan adalah pelanggan yang loyal. Dengan persaingan bisnis yang begitu ketat, sebagai pemilik perusahaan harus mengetahui kebutuhan customer dengan baik agar pelanggan tidak berpaling. Untuk menghindari itu, perusahaan perlu melakukan strategi khusus agar terjalin relasi yang baik antara perusahaan dan pelanggan. Strategi yang dijalankan juga harus sesuai dengan kebutuhan pelanggan tersebut, karena jika tidak maka usaha yang dilakukan akan sia-sia.

Langkah yang dapat dilakukan oleh perusahaan adalah dengan memberikan nilai kepada pelanggan berdasarkan kriteria tertentu yang menguntungkan perusahaan. Konsep ini yang dikenal dengan CLV (*customer lifetime value*) [1]. CLV merupakan nilai yang bersifat estimasi. Meskipun begitu, nilai ini dapat digunakan untuk mengevaluasi masa depan pelanggan terhadap perusahaan dengan teknik *data mining* dalam mendeteksi pola dan hubungan menggunakan histori data [2]. Salah satu model yang bisa mengukur CLV adalah RFM (*recency, frequency, monetary*) [3].

RFM pertama kali diperkenalkan oleh Hughes (1994), yang mana merupakan metode segmentasi paling umum digunakan untuk mengidentifikasi nilai pelanggan pada perusahaan berdasarkan 3 variabel: *recency, frequency, dan monetary*. *Recency* merupakan perhitungan berapa hari sejak transaksi terakhir pelanggan sampai dengan hari ini. *Frequency* merupakan perhitungan berapa banyak jumlah transaksi yang dilakukan selama periode perhitungan, sedangkan *monetary*

adalah berapa jumlah uang yang sudah ditransaksikan oleh pelanggan [4].

Model RFM ini kemudian dikembangkan oleh Chang dan Tasy (2004) dengan penambahan variabel *length*, dan dikenal dengan model LRFM (*length, recency, frequency, monetary*) [5]. *Length* merupakan perhitungan jarak hari antara transaksi pelanggan pertama kali dan terakhir kali. Pada paper ini, penulis mencoba menambahkan variabel *interval* dan *quantity*. *Interval* adalah rata-rata jarak hari dari setiap transaksi seorang pelanggan, sedangkan *quantity* adalah jumlah dari semua barang yang pernah ditransaksikan. Penambahan dua variabel ini disisipkan pada teori sebelumnya menjadi LRIFMQ (*length, recency, interval, frequency, monetary, quantity*). Tiap parameter tersebut mempunyai bobot yang akan ditentukan menggunakan metode AHP (*analytical hierarchy process*).

Setelah menilai pelanggan dengan LRIFMQ, langkah yang dilakukan selanjutnya adalah melakukan segmentasi pelanggan. Segmentasi ini bisa dilakukan dengan bantuan teknik *data mining* yang dikenal dengan istilah *clustering*. Secara garis besar, *clustering* dalam *data mining* adalah proses pembentukan segmen atau kluster dengan melihat kemiripan antar data berdasarkan parameter yang diberikan.

Terdapat banyak metode *clustering* yang mempunyai kelebihan dan kekurangan masing-masing. Dapat dilihat pada Tabel I bahwa mayoritas penelitian terdahulu menggunakan algoritma *k-means*, yang mana algoritma ini banyak digunakan karena performanya yang cukup efektif dan efisien [6]. Algoritma lain yang cukup populer adalah *fuzzy c-means*, yang mana algoritma ini mempunyai unsur kesamaran (*fuzziness*)

yang sulit diselesaikan menggunakan algoritma *k-means*. Oleh karena itu, paper ini menggunakan kedua algoritma tersebut.

Untuk menentukan jumlah cluster yang terbaik, digunakan metode lengan atau yang biasa dikenal dengan *elbow method* dan *silhouette coefficient*. Selain itu metode *silhouette coefficient* ini juga bisa digunakan untuk menentukan kualitas dari suatu proses *clustering*.

TABEL I
PENELITIAN TERDAHULU

Penelitian	Model	Metode Clustering
Hughes(1994)	RFM	-
Miglautsch(2000)	RFM	-
Shih and Liu(2003)	RFM	<i>K-means clustering</i>
Chang and Tsay(2004)	LRFM	<i>Self-organizing maps(SOM)</i>
Hu and Jing(2008)	RFM	<i>K-means clustering</i>
Bin Peiji and Dan(2008)	RFM	<i>K-means clustering</i>
Wu et al.(2009)	RFM	<i>K-means clustering</i>
Chang et al.(2010)	RFM	<i>K-means clustering</i>
Li et al.(2011)	LRFM	<i>Two-Step clustering</i>
Wei et al.(2012)	LRFM	<i>SOM</i>
Chen(2012)	RFM	<i>C-means clustering</i>
Kafashpoor and Alizadch(2012)	RFM	<i>Hicrarchical Clustering</i>

II. METODOLOGI PENELITIAN

Dalam menyelesaikan penelitian ini terdapat beberapa langkah yang harus dilakukan, yaitu sebagai berikut

A. Pengumpulan Data

Sumber data didapatkan dari PT. XYZ (perusahaan yang bergerak dalam bidang pengolahan makanan). Data yang diambil adalah transaksi penjualan yang terdiri dari 2.629.261 transaksi dan 64.239 pelanggan.

B. Pembobotan Menggunakan AHP

Analytical hierarchy process (AHP) merupakan salah satu metode yang digunakan dalam menentukan bobot. Dalam kasus ini yang ditentukan bobotnya adalah LRIFMQ. Langkah-langkah yang dilakukan adalah sebagai berikut:

- Melakukan prioritas menggunakan *pairwise comparison matrix* dengan *pairwise comparison index* yang ditunjukkan pada Tabel II [3].
- Melakukan normalisasi *pairwise comparison matrix*.
- Menghitung bobot dari *pairwise comparison matrix* dari nilai rata-rata tiap variabel. Dimana dari perhitungan ini akan dihasilkan nilai $W_L, W_R, W_i, W_F, W_M, W_Q$.
- Menghitung *consistency index* (CI), dengan rumus:

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (1)$$

- Menghitung *consistency ratio* (CR), dengan rumus

$$CR = \frac{CI}{RI} \quad (2)$$

Dimana RI untuk jumlah $n=6$ adalah 1,24. Diperlukan *consistency ratio* dibawah 0,1 untuk bobot dianggap valid dan bisa digunakan.

TABEL II
PAIRWISE COMPARISON INDEX

Definition	Index	Definition	Index
<i>Equally important</i>	1	<i>Equally important</i>	1/1
<i>Equally or slightly more important</i>	2	<i>Equally or slightly less important</i>	1/2
<i>Slightly more important</i>	3	<i>Slightly less important</i>	1/3
<i>Slightly to much more important</i>	4	<i>Slightly to way less important</i>	1/4
<i>Much more important</i>	5	<i>Way less important</i>	1/5
<i>Much to far more important</i>	6	<i>Way to far less important</i>	1/6
<i>Far more important</i>	7	<i>Far less important</i>	1/7
<i>Far more important to extremely more important</i>	8	<i>Far less important to extremely less important</i>	1/8
<i>Extremely more important</i>	9	<i>Extremely less important</i>	1/9

C. Pengolahan Data

Beberapa hal yang dilakukan dalam tahap pengolahan data ini adalah:

- Menghapus data yang duplikat.
- Menghapus data yang inkonsisten.
- Menghapus data pelanggan yang hanya melakukan transaksi 1 kali, karena parameter *interval* wajib mempunyai minimal 2 transaksi.

D. Pembentukan LRIFMQ

Transformasi data menjadi bentuk LRIFMQ.

- L = Tanggal transaksi akhir – tanggal transaksi awal
- R = 31/12/2019 – tanggal transaksi terakhir
- I = Rata-rata jarak antar transaksi
- F = Jumlah transaksi yang pernah dilakukan
- M = Jumlah nilai uang dari semua transaksi
- Q = Jumlah *quantity* dari semua transaksi

E. Normalisasi Menggunakan Min-Max

Normalisasi data menggunakan metode *min-max*, dengan range 0-1. Rumus normalisasi *min-max* adalah sebagai berikut:

$$V_1 = \frac{V_{\min}}{\max - \min} (D - C) + C \quad (3)$$

Dimana:

- V_1 = nilai *min-max*
- *min* = nilai terkecil dalam atribut
- *max* = nilai terbesar dalam atribut
- D = nilai *max* baru (1)
- C = nilai *min* baru (0)

F. Proses Clustering

Metode clustering utama yang digunakan adalah *k-means*, sedangkan *fuzzy c-means* digunakan sebagai perbandingan. Langkah-langkah clustering menggunakan *k-means* adalah sebagai berikut [7]:

- Menentukan jumlah kluster k
- Menentukan titik *centroid* secara random sebanyak jumlah kluster k .
- Menghitung jarak data terhadap *centroid* menggunakan formula *euclidian distance*, seperti berikut:

$$D_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

- Memperbaharui titik *centroid* berdasarkan nilai rata-rata pada tiap cluster.
- Ulangi poin c - d sampai semua titik *centroid* *convergen*/ tidak bergerak.
- Algoritma *k-means* dijalankan beberapa kali untuk mencari *global optima*.

Langkah-langkah clustering dengan *fuzzy c-means* adalah sebagai berikut:

- Memilih titik pusat sebanyak k secara random (jumlah k ditentukan sebelumnya).
- Menghitung *fuzzy membership* μ_{ij} , dengan rumus:

$$\mu_{ij} = \frac{1}{\sum_{c=1}^k \left(\frac{d_{ij}}{d_{cj}}\right)^{\frac{2}{m-1}}} \quad (5)$$

Dimana:

- $d_{ij} = \|x_i - v_j\|$
- $d_{cj} = \|x_i - v_c\|$

- Memperbaharui *centroid* pada tiap cluster v_j , dengan rumus:

$$v_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m} \quad (6)$$

- Ulangi sampai:

$$\varepsilon > \max_{ij} \{ |u_{ij}^{(k+1)} - u_{ij}^k| \} \quad (7)$$

Dimana ε merupakan batas yang ditentukan antara 0 dan 1.

G. Menentukan Jumlah Kluster Optimal

Untuk menentukan jumlah yang optimal, digunakan metode *silhouette coefficient* dan *elbow method*. Langkah-langkah dalam *silhouette coefficient* adalah sebagai berikut:

- Hitung rata-rata jarak objek data dengan objek data lain dalam 1 kluster. Akan didapatkan nilai $a(o)$, dimana rumusnya adalah sebagai berikut:

$$a(o) = \frac{\sum_{o' \in C_i, o' \neq o} \text{dist}(o, o')}{|C_i| - 1} \quad (8)$$

- Hitung rata-rata jarak antara objek data dengan objek data di kluster lain. Dari jarak rata-rata tiap kluster, ambil yang paling minim, dan akan didapatkan nilai $b(o)$, dimana rumusnya adalah sebagai berikut:

$$b(o) = \min_{C_j, 1 \leq j \leq k, j \neq i} \left\{ \frac{\sum_{o' \in C_j} \text{dist}(o, o')}{|C_j|} \right\} \quad (9)$$

- Hitung *silhouette coefficient* dengan rumus sebagai berikut:

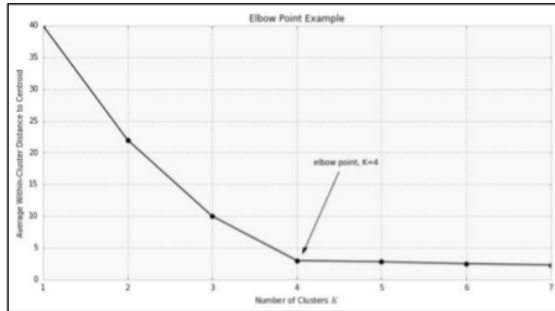
$$S_i = \frac{b(o) - a(o)}{\max\{a(o), b(o)\}} \quad (10)$$

Sedangkan *elbow method* adalah metode yang menentukan jumlah kluster optimal menggunakan visual, seperti yang ditunjukkan pada Gambar 1. Pada gambar tersebut terlihat terjadi perubahan yang cukup signifikan di titik kluster 4, yang mana artinya adalah jumlah kluster optimal adalah 4. Rumus yang digunakan dalam *elbow method* adalah sebagai berikut [8]:

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} |x_i - c_k|^2 \quad (11)$$

Dimana:

- K = jumlah kluster
- x_i = data ke- i
- c_k = *centroid* kluster k



Gambar 1. Ilustrasi Elbow Method

H. Menghitung CLV

1 Untuk menghitung nilai CLV, digunakan rumus sebagai berikut:

$$CLV = L * W_L + R * W_R + I * W_I + F * W_F + M * W_M + Q * W_Q \quad (12)$$

1 Dimana:

- L, R, I, F, M, Q = Rata-rata nilai pelanggan
- $W_L, W_R, W_I, W_F, W_M, W_Q$ = Bobot masing-masing LRIFMQ.

1 I. Evaluasi dan Analisa Hasil

Dalam tahap ini akan dilakukan evaluasi dan analisa hasil dari semua uji coba. Jika hasil yang didapatkan tidak sesuai karena kesalahan uji coba, maka langkah bisa diulang kembali ke titik terjadinya kesalahan. *clustering* yang terbaik adalah hasil dari uji coba dengan *silhouette coefficient* yang paling mendekati satu. Dilakukan juga analisa tipe pelanggan untuk tiap klusternya.

III. HASIL DAN PEMBAHASAN

Pada tahap persiapan awal, proses AHP dilakukan untuk menentukan bobot masing-masing parameter uji coba. Pemberian bobot menggunakan *pairwise comparison index* dilakukan secara subjektif karena sangat tergantung pada perusahaan bagaimana melihat kepentingan masing-masing parameter LRIFMQ. Bobot yang didapatkan untuk masing-masing $W_L, W_R, W_I, W_F, W_M, W_Q$ adalah 0,0685, 0,0408, 0,2582, 0,2913, 0,1899, 0,1513 dengan nilai *consistency ratio* 0,0156. Dengan nilai *consistency ratio* dibawah 0,1 maka bobot dianggap valid dan bisa digunakan.

Hasil pengolahan data, didapatkan jumlah transaksi 337.739 data dan 26.683 pelanggan. Data ini ditransformasi kedalam bentuk LRIFMQ, yang bisa dilihat pada Tabel III. Kemudian dilakukan normalisasi menggunakan metode *min-max* dengan range 0-1, dan hasilnya ditunjukkan pada Tabel IV.

TABEL III
DATA LRIFMQ

Cust No.	L	R	I	F	M	Q
1	357	3	2,73	132	11129041	35805,4
2	357	0	2,2	163	4993819	83651,6
3	62	298	1,35	47	454513	1410
4	62	299	1,35	47	178907,5	559
5	62	299	1,35	47	580652,5	1769
6	62	298	1,38	46	540590	1656
7	62	298	1,35	47	158350,5	495
8	359	3	0,67	533	36590709	16717,8
...						
25683	62	299	1,35	47	179106,5	560

TABEL IV
DATA LRIFMQ HASIL NORMALISASI

Cust No.	L	R	I	F	M	Q
1	0,9623	0,0083	0,0075	0,1037	0,0252	0,0069
2	0,9623	0	0,0061	0,1284	0,0113	0,0161
3	0,1671	0,8232	0,0037	0,0359	0,0010	0,0003
4	0,1671	0,8260	0,0037	0,0359	0,0004	0,0001
5	0,1671	0,8260	0,0037	0,0359	0,0013	0,0003
6	0,1671	0,8232	0,0038	0,0351	0,0012	0,0003
7	0,1671	0,8232	0,0037	0,0359	0,0004	0,0001
8	0,9677	0,0083	0,0019	0,4234	0,0828	0,0032
...						
25683	0,1671	0,8260	0,0037	0,0359	0,0004	0,0001

12 A. Uji Coba Menggunakan Algoritma K-Means

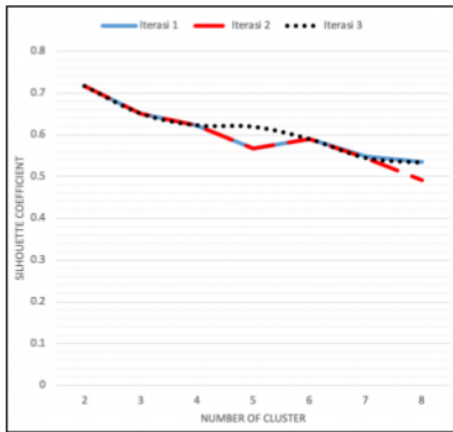
Uji coba menggunakan algoritma *k-means* dilakukan berulang sebanyak tiga kali. Masing-masing perulangan dilakukan *clustering* mulai dari $k=2$ sampai dengan $k=8$. Alasan diberhentikan pada angka 8 dikarenakan nilai *silhouette coefficient* semakin turun seiring bertambahnya jumlah kluster k .

Silhouette coefficient digunakan untuk mengukur kualitas proses *clustering*, sekaligus untuk menentukan berapa jumlah kluster yang optimal. Nilai *silhouette coefficient* memiliki range antara -1 sampai dengan 1, jika nilai mendekati 1 maka kualitas kluster yang dihasilkan semakin baik. Sedangkan mendekati -1 maka kualitas kluster semakin buruk [9]. Tabel V menunjukkan perbandingan nilai *silhouette coefficient* dari tiap perulangan, dan dari tiap jumlah kluster k yang berbeda. Didapatkan hasil bahwa jumlah kluster 2 memiliki nilai *silhouette coefficient* yang paling mendekati 1, yaitu 0,71616, dalam setiap perulangan yang dilakukan.

Dari uji coba yang dilakukan, diketahui pula bahwa perulangan penting dilakukan karena dapat dilihat pada iterasi ketiga $k=5$, dan $k=6$, nilai *silhouette coefficient* berubah menjadi lebih baik. Gambar 2 menunjukkan grafik dari nilai *silhouette coefficient* tiap perulangan yang dilakukan.

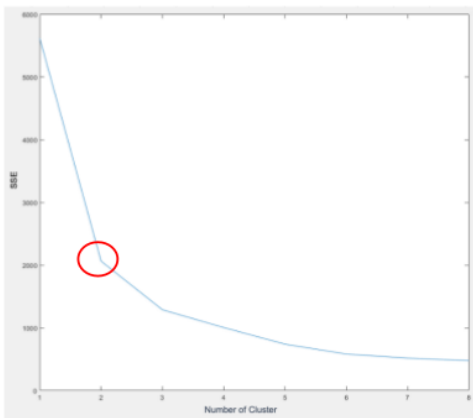
TABEL V
PERBANDINGAN SILHOUETTE COEFFICIENT MENGGUNAKAN K-MEANS

Number Of Cluster	Silhouette Coefficient		
	Iterasi 1	Iterasi 2	Iterasi 3
2	0,71616	0,71616	0,71616
3	0,6505	0,6505	0,6505
4	0,62222	0,62222	0,62222
5	0,56622	0,56622	0,61896
6	0,58945	0,58945	0,58956
7	0,54796	0,54448	0,54448
8	0,5338	0,49012	0,53293



Gambar 2. Grafik Silhouette Coefficient Menggunakan K-Means

Sementara itu, untuk menentukan jumlah kluster yang optimal, digunakan pula *elbow method* sebagai perbandingan. Dari Gambar 3 dapat dilihat garis yang mengalami patahan membentuk siku adalah pada titik jumlah kluster sama dengan 2.



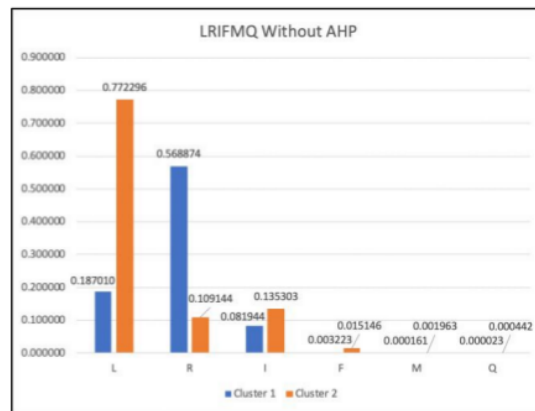
Gambar 3. Elbow Method

Dari kedua perbandingan metode yang digunakan untuk menentukan jumlah kluster, diambil kesimpulan bahwa jumlah kluster yang paling tepat untuk kondisi data PT. XYZ adalah 2. Hal ini tentunya dapat berbeda apabila menggunakan dataset lain.

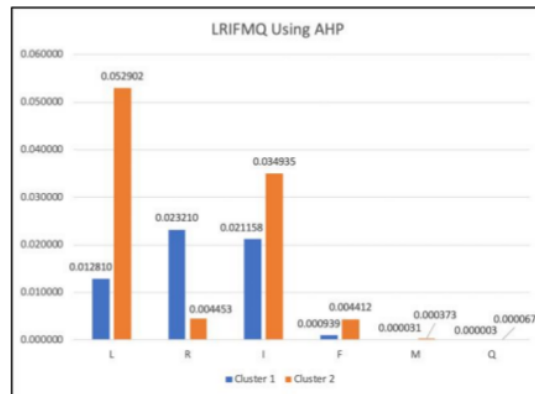
TABEL VI
PERBANDINGAN CLV DENGAN 2 KLUSTER (K-MEANS)

Weight Cluster Member	No Weight		AHP Weight	
	1	2	1	2
Avg. Silh	11831	13852	11831	13852
L	0,65395	0,77837	0,65395	0,77837
R	0,187010	0,772296	0,012810	0,052902
I	0,568874	0,109144	0,023210	0,004453
F	0,081944	0,135303	0,021158	0,034935
M	0,003223	0,015146	0,000939	0,004412
Q	0,000161	0,001963	0,000031	0,000373
CLV	0,000023	0,000442	0,000003	0,000067

Number of Cluster = 2, Average Silhouette = 0.716160



Gambar 4. Perbandingan Nilai LRIFMQ Tanpa AHP (K-Means)



Gambar 5. Perbandingan Nilai LRIFMQ Menggunakan AHP (K-Means)

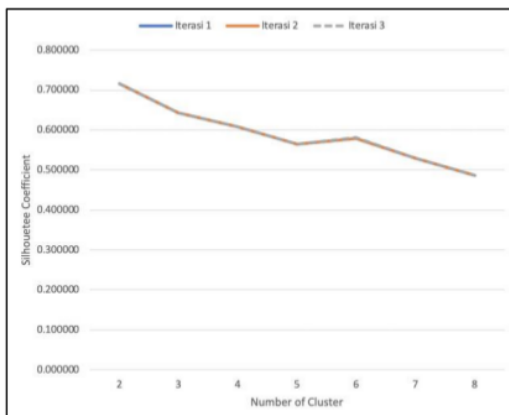
Pada Tabel VI dibandingkan nilai *customer lifetime value* (CLV) dan nilai LRIFMQ jika menggunakan AHP dan jika tidak. Perbandingan tersebut bisa dilihat dengan lebih jelas dengan grafik yang ditunjukkan pada Gambar 4 dan Gambar 5. Terlihat bahwa kluster 2 lebih unggul dari semua parameter LRIFMQ, dan mempunyai nilai CLV yang lebih tinggi dibandingkan kluster 1. Efek dari AHP terlihat pada parameter *interval* (I) dan *frequency* (F) melonjak cukup tinggi, sedangkan nilai *recency* (R) menjadi lebih kecil. Terlihat juga bahwa kluster 2 mempunyai nilai $L \uparrow R \downarrow I \uparrow F \uparrow M \uparrow Q \uparrow$ dimana ini merupakan karakter dari pelanggan loyal dan berpotensi loyal [10]. Kluster 1 mempunyai nilai kebalikan, dimana merupakan karakter dari segmen pelanggan baru yang tidak pasti dan berpotensi hilang.

B. Uji Coba Menggunakan Algoritma Fuzzy C-Means

Uji coba yang dengan menggunakan algoritma *fuzzy c-means* dilakukan dengan perlakuan yang sama dengan percobaan algoritma *k-means*, yaitu 3 perulangan dan jumlah $k=2$ sampai dengan $k=8$. Perbandingan *silhouette coefficient* tiap perulangan dapat dilihat pada Tabel VII.

TABEL VII
PERBANDINGAN SILHOUETTE COEFFICIENT MENGGUNAKAN FUZZY C-MEANS

Number Of Cluster	Silhouette Coefficient		
	Iterasi 1	Iterasi 2	Iterasi 3
2	0,71640	0,71640	0,71640
3	0,64245	0,64245	0,64242
4	0,60752	0,60752	0,60752
5	0,56478	0,56478	0,56476
6	0,57968	0,57925	0,58184
7	0,52982	0,52978	0,53001
8	0,48644	0,48645	0,48648



Gambar 6. Grafik Silhouette Coefficient Menggunakan K-Means

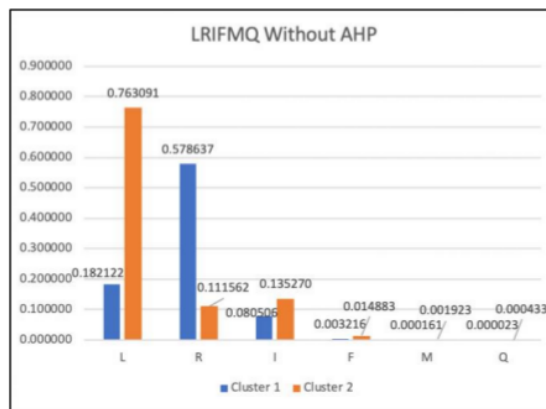
Dapat dilihat pada tabel tersebut bahwa ada perbedaan namun tidak signifikan, nilai yang tertinggi diberi warna berbeda. Perbedaan terlihat mulai dari $k=5$, dan seterusnya. Hal ini mengartikan *clustering* menggunakan *Fuzzy C-Means* cukup stabil di titik $k=2$ sampai dengan $k=4$. Pada Gambar 6 dapat dilihat grafik *silhouette coefficient* yang menumpuk, dikarenakan perbedaan nilai yang sangat kecil. Nilai *silhouette coefficient* tertinggi yang didapatkan adalah 0.7164 di jumlah kluster $k=2$.

Perbandingan CLV dan masing-masing variabel LRIFMQ ditunjukkan pada Tabel VIII. Dapat dilihat bahwa CLV pada kluster 2 lebih besar, yang menunjukkan pelanggan yang lebih loyal berada pada kluster tersebut. Jika dibandingkan dengan uji coba menggunakan *k-means*, member pada kluster 2 sebesar 14.171, lebih banyak 319 member.

TABEL VIII
PERBANDINGAN CLV DENGAN 2 KLUSTER (FUZZY C-MEANS)

Weight	No Weight		AHP Weight	
	1	2	1	2
Cluster				
Member	11512	14171	11831	13852
Avg. Silh	0,67021	0,76259	0,65395	0,77837
L	0,182122	0,763091	0,012475	0,052272
R	0,578637	0,111562	0,023608	0,004552
I	0,080506	0,135270	0,020787	0,034927
F	0,003216	0,014883	0,000937	0,004335
M	0,000161	0,001923	0,000031	0,000365
Q	0,000023	0,000433	0,000003	0,000065
CLV	0,844664	1,027162	0,057841	0,096516

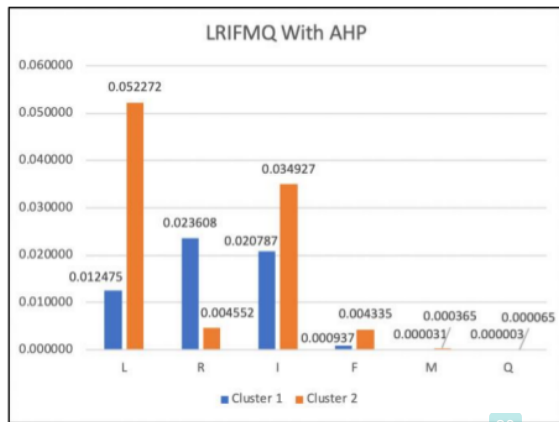
Number of Cluster = 2, Average Silhouette = 0.716398



Gambar 7. Perbandingan Nilai LRIFMQ Tanpa AHP (Fuzzy C-Means)

Perbandingan parameter LRIFMQ antara kluster 1 dan 2 ditunjukkan pada Gambar 7 dimana kluster 2 mempunyai nilai $L \uparrow R \downarrow I \uparrow F \uparrow M \uparrow Q \uparrow$ jika dibandingkan dengan kluster 1. Nilai tersebut menunjukkan karakter pelanggan yang loyal dan

berpotensi loyal. Gambar 8 menunjukkan parameter LRIFMQ menggunakan pembobotan AHP, yang efeknya juga sama seperti yang sudah dibahas pada uji coba *k-means*.



Gambar 8. Perbandingan Nilai LRIFMQ Menggunakan AHP (*Fuzzy C-Means*)

IV. KESIMPULAN

Penelitian ini menghasilkan 2 kluster yang dinilai menggunakan *silhouette coefficient* dan *elbow method*. Kedua kluster ini dihitung nilai CLV-nya, yang digunakan untuk menilai profitabilitas seorang pelanggan. Berdasarkan nilai CLV, diketahui bahwa kluster 2 adalah segmen pelanggan yang loyal dan berpotensi loyal dengan nilai CLV 0,097142. Sementara itu, kluster 1 merupakan pelanggan baru yang tidak pasti dengan nilai CLV 0,058151.

Kluster 2 dikatakan loyal dan berpotensi loyal karena memiliki nilai *length* tinggi, *recency* rendah, *interval* tinggi, *frequency* tinggi, *monetary* tinggi, dan *quantity* tinggi. Dikatakan berpotensi loyal karena terdapat nilai *interval* yang relatif tinggi dibanding kluster 1. Analisa ini bisa digunakan perusahaan untuk meningkatkan pelanggan yang berpotensi loyal menjadi loyal. Selain itu, menjaga pelanggan yang sudah loyal tetap pada posisinya.

V. REFERENSI

- [1] S. Monalisa, "Klasterisasi Customer Lifetime Value Dengan Model LRFM Menggunakan Algoritma K-Means," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 5, no. 2, pp. 247-252, 2018.
- [2] F. A. Buttle and S. Maklan, *Customer Relationship Management: Concepts and Technologies*, 2016.
- [3] A. Parvaneh, H. Abbasimehr and M. J. Tarokh, "Integrating AHP and Data Mining for Effective Retailer Segmentation Based on Retailer Lifetime Value," *Journal of Optimization in Industrial Engineering*, vol. 11, pp. 25-31, 2012.
- [4] A. J. Christy, A. Umamakeswari, L. Priyatharsini and A. Neyaa, "RFM ranking – An effective approach to customer segmentation," *Journal of King Saud University – Computer and Information Sciences*, 2018.

- [5] A. A. Zoeram, "A 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.
- [6] D. Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*, Informatika, 2019.
- [7] A. T. Rahman, Wiranto and R. Anggrainingsih, "Coal Trade Data Clustering Using K-Means," *ITSMART: Jurnal Ilmiah Teknologi dan Informasi*, vol. 6, no. 1, 2017.
- [8] N. P. E. Merliana, Ernawati and A. J. Santoso, "ANALISA PENENTUAN JUMLAH CLUSTER TERBAIK PADA METODE K-MEANS CLUSTERING (SENDI_U)," *PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU & CALL FOR PAPERS UNISBANK*, p. 978-979-3649-81-8, 2008.
- [9] B. Santoso, I. Cholissodin and B. D. Setiawan, "Optimasi K-Means untuk Clustering Kinerja Akademik Dosen Menggunakan Algoritme Genetika," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, pp. 1652-1659, 2017.
- [10] D.-C. Li, W.-L. Dai and W.-T. Tseng, "A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business," *Expert Systems with Applications*, pp. 7186-7191, 2011.

Pemodelan Cluster Loyalitas Customer Menggunakan Algoritma K-Means Dengan Parameter LRIFMQ

ORIGINALITY REPORT

24%

SIMILARITY INDEX

11%

INTERNET SOURCES

10%

PUBLICATIONS

21%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Universitas Brawijaya

Student Paper

6%

2

mdpi.com

Internet Source

3%

3

Submitted to Universitas 17 Agustus 1945

Surabaya

Student Paper

1%

4

Submitted to Reva University

Student Paper

1%

5

Submitted to University of Liverpool

Student Paper

1%

6

Deny Jollyta, Syahril Efendi, Muhammad Zarlis, Herman Mawengkang. "Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index", Prosiding Seminar Nasional Riset Information Science (SENARIS), 2019

Publication

1%

7	Submitted to Victoria University Student Paper	1%
8	jtiik.ub.ac.id Internet Source	1%
9	Jiayuan Ling, Gangmin Li. "A two-level stacking model for detecting abnormal users in Wechat activities", 2019 International Conference on Information Technology and Computer Application (ITCA), 2019 Publication	1%
10	Dhendra Marutho, Sunarna Hendra Handaka, Ekaprana Wijaya, Muljono. "The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News", 2018 International Seminar on Application for Technology of Information and Communication, 2018 Publication	1%
11	www.neliti.com Internet Source	1%
12	Submitted to Udayana University Student Paper	1%
13	Submitted to Universitas Islam Indonesia Student Paper	<1%
14	www.scribd.com Internet Source	<1%

15

Submitted to UIN Maulana Malik Ibrahim Malang

Student Paper

<1%

16

Serhat Peker, Altan Kocyigit, P. Erhan Eren.
"LRFMP model for customer segmentation in
the grocery retail industry: a case study",
Marketing Intelligence & Planning, 2017

Publication

<1%

17

ejournal-s1.undip.ac.id

Internet Source

<1%

18

Submitted to UIN Sultan Syarif Kasim Riau

Student Paper

<1%

19

www.authorstream.com

Internet Source

<1%

20

Submitted to Universitas Diponegoro

Student Paper

<1%

21

Rena Nainggolan, Resianta Perangin-angin,
Emma Simarmata, Astuti Feriani Tarigan.
"Improved the Performance of the K-Means
Cluster Using the Sum of Squared Error (SSE)
optimized by using the Elbow Method", Journal
of Physics: Conference Series, 2019

Publication

<1%

22

pt.scribd.com

Internet Source

<1%

Submitted to Universiti Putra Malaysia

23

Student Paper

<1%

24

jurnal.uns.ac.id

Internet Source

<1%

25

Submitted to Binus University International

Student Paper

<1%

26

publikasi.dinus.ac.id

Internet Source

<1%

27

Pradnya Paramita Pramono, Isti Surjandari, Enrico Laoh. "Estimating Customer Segmentation based on Customer Lifetime Value Using Two-Stage Clustering Method", 2019 16th International Conference on Service Systems and Service Management (ICSSSM), 2019

Publication

<1%

28

Submitted to Cedar Valley College

Student Paper

<1%

29

Reksa Suhud Tri Atmojo. "Analisis Data E-Absensi untuk Menganalisis Perbandingan Pola Disiplin Kerja menggunakan Algoritma Clustering K-Means", Electrician, 2019

Publication

<1%

30

Submitted to KYUNG HEE UNIVERSITY

Student Paper

<1%

31	media.neliti.com Internet Source	<1%
32	www.ijcaonline.org Internet Source	<1%
33	Kandeil, Dalia AbdelRazek, Amani Anwar Saad, and Sherin Moustafa Youssef. "A Two-Phase Clustering Analysis for B2B Customer Segmentation", 2014 International Conference on Intelligent Networking and Collaborative Systems, 2014. Publication	<1%
34	Submitted to Massey University Student Paper	<1%
35	Submitted to University of Maryland, University College Student Paper	<1%
36	Submitted to Sriwijaya University Student Paper	<1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography On