

KLASTERISASI *CUSTOMER LIFETIME VALUE* DENGAN MODEL LRFM MENGGUNAKAN ALGORITMA FUZZY C-MEANS

Fhadilatul Juniati¹, Revilia Zafa², Muslim³ Siti Monalisa⁴

Prodi Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau

Email: fhadilatuljuniati@gmail.com, rrefiliazafa20@gmail.com, 11653103698@students.uin-suska.ac.id,
siti.monalisa@uin-suska.ac.id

ABSTRAK

PT. Pan Baruna Cabang Pekanbaru merupakan suatu perusahaan bergerak dibidang distributor produk makanan dan sembako, pencacatan transaksi yang di lakukan pada PT Pan Baruna masih bersifat manual yaitu di catat dengan menggunakan excel. Data transaksi yang terjadi setiap pada PT Pan Baruna cukup banyak hal ini yang menyebabkan tumpukan yang seharusnya dapat di olah lebih lanjut agar menghasilkan informasi yang bermanfaat. Algoritma Fuzzy CMeans merupakan salah satu algoritma yang dapat digunakan untuk pengolahan data, kluster yang di hasilkan pada penelitian ini berjumlah 3 kluster untuk mendapat nilai CLV yaitu dengan mengalikan nilai normalisasi LRFM dengan nilai bobot LRFM dan selanjutnya di jumlahkan. Perjumlahan nilai CLV terssebut dilakukan pada masing-masing kluster yang telah terbentuk. Berdasarkan penjumlahan tersebut maka dicari rangking CLV. Rangking CLV tertinggi dihasilkan dari nilai CLV yang terbesar di antara 3 kluster tersebut. Rangking tertinggi pada penelitian ini adalah kluster ke 1 yaitu 0,433 dengan bahwa kluster ini memiliki segmen pelanggan nilai loyalitas yang tinggi. Rangking ke-2 adalah kluster 2 dengan nilai CLV 0,268 dan selanjutnya kluster 3 dengan nilai CLV 0,178.

Kata Kunci: Cluster, *Fuzzy-C-Means*, *LRFM*

1. PENDAHULUAN

Clustering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan (kemiripan) (Satriyanto, 2015). *Clustering* berbeda dengan group, kalau group berarti kelompok yang memiliki kondisi yang sama. Sedangkan cluster kelompoknya tidak harus sama, pengelompokan didasarkan pada kedekatan dari suatu karakteristik sampel yang ada, salah satunya dengan menggunakan rumus jarak Euclidean. Pengaplikasian cluster sangatlah banyak, hal ini dikarenakan dalam mengidentifikasi suatu permasalahan atau pengambilan keputusan selalu tidak sama persis akan tetapi cenderung memiliki kemiripan saja (Satriyanto, 2015).

Metode clustering pada dasarnya mengoptimalkan pusat cluster (Centroid) (Kusumadewi, 2010). Era globalisasi yang beberapa tahun lalu dimulai, menyebabkan perubahan besar bagi banyak negara. Terlebih setelah disepakatinya MEA pada awal tahun 2016, menyebabkan dampak buruk pada berbagai sektor, termasuk sektor industri (Tempo, 2016). Hal ini memaksa perusahaan-perusahaan yang bergerak di sektor industri di Indonesia untuk bersaing dengan negara-negara tetangga di ASEAN. Salah satu industri yang saat ini tengah bersaing adalah industri makanan. PT. PAN BARUNA merupakan cabang perusahaan yang bergerak di industri makanan. PT. PAN BARUNA selama bulan 1 Januari 2019 sampai 31 Oktober 2019, angka penjualan produk tersebut tidak mencapai target

yang ditetapkan setiap bulannya. Adapun tidak tercapainya target penjualan disebabkan berbagai faktor. Salah satunya adalah harga jual produk lebih rendah yang ditawarkan oleh kompetitor menyebabkan sebagian pelanggan beralih kepada kompetitor. Terkait dengan masalah ini, PT. PAN BARUNA telah melakukan berbagai kegiatan promosi untuk menarik kembali pelanggan. PT. PAN BARUNA menerapkan strategi marketing kepada pelanggan tanpa adanya segmentasi pelanggan. Oleh karena itu, PT. PAN BARUNA menerapkan strategi yang sama untuk semua pelanggan.

Penerapan strategi yang sama ini dapat menimbulkan beberapa dampak, yaitu strategi promosi yang tidak tepat sasaran sehingga strategi tersebut hanya bermanfaat pada pelanggan tertentu saja. Selain itu, waktu dan dana yang digunakan untuk promosi menjadi kurang efektif dan efisien. Untuk menyelesaikan masalah pelanggan tersebut, salah satu solusi yang dapat dilakukan adalah menentukan segmentasi pelanggan. Banyak metode yang dapat digunakan dalam melakukan segmentasi pelanggan, salah satunya adalah penggalian data terhadap riwayat transaksi pembelian pelanggan pada periode tertentu. Data transaksi tersebut kemudian ditentukan variabel-variabel RFM. Data transaksi yang digunakan adalah periode 1 Januari sampai 31 Oktober 2019. Algoritma clustering yang digunakan adalah Fuzzy C-Means. Fuzzy C-Means merupakan salah satu teknik untuk menentukan cluster optimal pada suatu ruang vektor yang didasarkan pada bentuk normal Euclidean untuk

jarak antar vektor. Fuzzy C-Means menggunakan derajat keanggotaan sebagai dasar penentuan suatu vektor masuk ke dalam cluster tertentu (Denny B. Saputra dan Edwin Riksakomara, 2018).

2. LANDASAN TEORI

2.1 Model LRFM

Untuk mengidentifikasi kebiasaan pelanggan, metode yang umum digunakan adalah *recency, frequency, and monetary* (RFM) *model* yang digunakan untuk merepresentasikan karakteristik kebiasaan pelanggan. Model RFM pertama kali diperkenalkan oleh Hughes dan saat ini banyak digunakan oleh industri termasuk manufaktur, *retailer*, dan industri jasa. Model LRFM melibatkan 4 variabel yaitu *Length of the purchases, Recency of the last purchases, Frequency of the purchases, dan Monetary value of the purchases*. Adapun penjelasan keempat variabel tersebut adalah:

- Lenght yaitu lamanya hubungan antara pelanggan dan perusahaan yang di ukur selama periode analisis.
- Recency* adalah jarak dari waktu transaksi terakhir kali dilakukan dengan waktu saat ini. Semakin kecil nilai jarak waktu, maka semakin besar nilai R.
- Frequency* adalah total jumlah transaksi yang dilakukan selama periode tertentu. Semakin besar jumlah transaksi, maka semakin besar nilai F.
- Monetary* adalah total nilai produk dalam bentuk uang dalam periode tertentu. Semakin besar nilai produk, maka semakin besar nilai M.

2.2 Algoritma Fuzzy C-Means

Konsep dasar FCM pertama kali adalah menentukan pusat cluster yang akan menandai lokasi rata-rata untuk tiap-tiap cluster. Pada kondisi awal, pusat cluster ini masih belum akurat. Tiap-tiap data memiliki derajat keanggotaan untuk tiap-tiap cluster (Redjeki.s, 2107). Dengan cara memperbaiki pusat cluster dan nilai keanggotaan tiap-tiap data secara berulang, maka dapat dilihat bahwa pusat cluster akan menuju lokasi yang tepat (Mas`udia. P. E., F. Arinie, L. D. Mustafa, 2018). Perulangan ini didasarkan pada minimasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan ke pusat cluster yang terbobot. Algoritma yang digunakan pada metode Fuzzy C-means adalah sebagai berikut:

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w \times x_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (1)$$

dengan $k=1,2,\dots,c$; dan $j=1,2,\dots,m$

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ((\sum_{j=1}^m (x_{ij} - V_{kj})^2) (\mu_{ik})^w) \quad (2)$$

$$\mu_{ik} = \frac{(\sum_{j=1}^m (x_{ij} - V_{kj})^2)^{\frac{1}{w-1}}}{\sum_{k=1}^c (\sum_{j=1}^m (x_{ij} - V_{kj})^2)^{\frac{1}{w-1}}} \quad (3)$$

dengan $i = 1,2,3,\dots,n$; dan $k = 1,2,3,\dots,c$.

Algoritma FCM

[1] Input data yang akan di cluster X, berupa matriks berukuran $n \times m$ (n = jumlah sampel data, m = atribut setiap data). X_{ij} = data sampel ke- i ($i = 1,2,3,\dots,n$), atribut ke- j ($j = 1,2,3,\dots,m$).

[2] Tentukan : jumlah cluster (c), pangkat (w), maksimum iterasi (MaxIter), error terkecil yang diharapkan (ξ), fungsi objektif awal ($P_0 = 0$) dan iterasi awal ($t=1$).

[3] Bangkitkan bilangan random μ_{ik} , dimana $i=1,2,3,\dots,n$; $k=1,2,\dots,c$; sebagai elemen - elemen matriks partisi awal U.

[4] Hitung pusat cluster ke- k dengan persamaan

[5] Hitung fungsi obyektif pada iterasi ke - t , P_t atau dengan cara menghitung hasil maksimal dari selisih matriks partisi akhir-dikurangi matriks partisi awal menggunakan persamaan (2)

[6] Hitung perubahan matriks partisi dengan persamaan (3)

[7] Cek kondisi berhenti :

a. Jika ($|P_t - P_{t-1}| < \xi$) atau ($t > \text{MaxIter}$) maka berhenti.

b. Jika tidak : $t=t+1$, ulangi langkah ke -4.

2.3 Customer Lifetime Value (CLV)

Menurut Cheng, mendefinisikan nilai pelanggan adalah model yang didasarkan pada penghitungan jarak antara nol dan cluster pusat sebagai nilai tinggi dan merujuk pada sebagian besar loyalitas pelanggan di dalamnya. CLV digunakan dalam menghitung nilai profitabilitas pelanggan. CLV dihitung setelah melakukan segmentasi pelanggan. Nilai CLV dihitung berdasarkan peringkat CLV yang ditentukan untuk setiap segmen. Perhitungan CLV menggunakan persamaan i.

$$C^j = W_L C_L^j + W_R C_R^j + W_F C_F^j + W_M C_M^j \quad (1)$$

Where:

C^j = Customer CLV rating j

$C_L^j, C_R^j, C_F^j, C_M^j$ = Normalization L, R, F, M from cluster j

W_L, W_R, W_F, W_M = result of weight L, R, F, M from AHP

3. METODOLOGI PENELITIAN

Penelitian ini terdiri dari beberapa tahapan

- a. Memilih data sesuai dengan atribut L, R, F dan M
- b. Melakukan normalisasi data disebabkan selisih data antara L, R, F dan M yang sangat jauh sehingga perlu dilakukan normalisasi dengan metode min-max menggunakan range 0-1. Persamaan metode min-max sebagai berikut :

$$B = \left(\frac{A - \min A}{(\max A - \min A)} \right) * (D - C) + C$$

Dimana :

A = objek data

min A = nilai terkecil dari atribut yang akan dinormalisasikan

max A = nilai terbesar dari atribut yang akan dinormalisasikan

C dan D = adalah range yang digunakan misalnya 0-1

- c. Melakukan klusterisasi dengan algoritma K-Means berdasarkan jumlah kluster yang terbaik

- d. Melakukan pembobotan nilai LRFM dengan metode AHP.

- e. Mengalikan nilai bobot LRFM dengan rata-rata nilai LRFM dan menjumlahkannya seperti pada persamaan 8 berikut :

$$CLV = L * WL + R * WR + F * WF + M * WM$$

Dimana :

L,R,F,M =Rata-rata nilai pelanggan

LW, RW, FW, MW =Bobot masing-masing nilai LRFM

- f. Meranking dan menganalisis hasil klusterisasi dengan rank LRFM.

4. HASIL DAN PEMBAHASAN

Penelitian menggunakan data PT. PAN BARUNA cabang Pekanbaru dengan total pelanggan yang di analisis pada periode 1 Januari 2019 sampai 31 Oktober 2019 sebanyak 100 data. Data tersebut terlebih dahulu akan dilakukan proses preprocessing. Tahap ini merupakan proses yang berguna dalam meningkatkan akurasi dan efisiensi pemodelan data. Tahap ini berguna untuk melihat apakah terdapat data yang ganda, kosong dan sebagainya yang nantinya akan merusak hasil penelitian. Selanjutnya data dipilih sesuai atribut L,R, F dan M, dapat lihat pada Tabel 1. Data yang telah dipilih sesuai L, R, F dan M selanjutnya dilakukan normalisasi dengan menggunakan persamaan 7. Normalisasi pada penelitian ini digunakan karena nilai M yaitu monetary memiliki selisih yang jauh berbeda dibandingkan dengan nilai L, R dan F. Hal ini dikarenakan nilai M merupakan jumlah uang yang

dikeluarkan pelanggan untuk perusahaan dengan satuannya yaitu Rupiah (Rp). Selisih yang sangat jauh inilah diperlukan normalisasi agar tidak mengganggu hasil penelitian. Range yang digunakan pada penelitian ini yaitu nilai antara 0-1 (lihat Tabel 3).

Tabel.1 Data Mentah

NO	Nama	Tgl. Faktur	Nett
1	FARHAN	2-Jan-19	227401
		29-Jan-19	113700
		12-Mar-19	227401
		9-Apr-19	113700
		25-Apr-19	113700
			22650
		8-May-19	227401
		20-Jun-19	341100
		30-Jul-19	113700
		27-Aug-19	172552
2	NABILA	24-Sep-19	39127
		28-Sep-19	227400
		9-Apr-19	113700
		7-May-19	56850
		3-Jul-19	113700
3	A.RAUF	27-Aug-19	56850
		8-Oct-19	113700
		12-Apr-19	68801
		4	EDDY JAYA
6-Aug-19	94751		
23-Aug-19	94751		
5	ABDAN	7-Sep-19	113700
		27-Apr-19	113700
		20-Aug-19	113700
		7-Sep-19	113700
6	ABUN	27-Sep-19	113700
		11-Oct-19	113700
		25-May-19	113700
		13-Jul-19	113700
		6-Sep-19	94751
7	ADE	19-Sep-19	189500
		1-Feb-19	37901
		16-Feb-19	37901
		16-Mar-19	113700
		27-May-19	37901
		22-Jun-19	113700

8	ADIBAH GROSIR	27-Apr-19	113700
		25-May-19	113700
...
100	MUTIA	2-Jan-19	113700
		15-Jan-19	227401
		26-Mar-19	113700
		9-Apr-19	113700
		25-Apr-19	113700
		7-May-19	113700
		22-May-19	113700
		20-Jun-19	227400
		13-Aug-19	113700
		8-Oct-19	113700

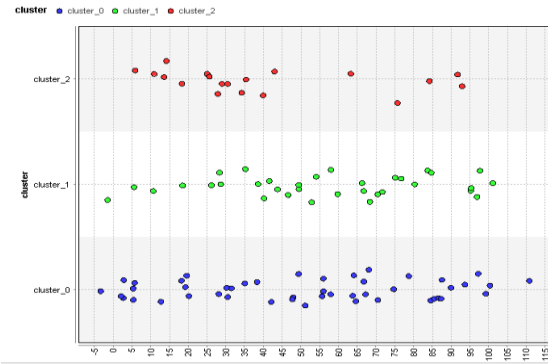
8	0.609	0.594	0.100	0.044
9	0.861	0.323	0.200	0.071
10	0.833	0.087	0.400	0.150
11	0.874	0.476	0.100	0.031
....
100	1.000	0.059	0.900	0.310

Tabel 2. Tranformasi Data

NO. Pelanggan	L	R	F	M
1	302	33	11	1939832
2	205	23	5	454800
3	202	202	1	68801
4	299	54	4	416902
5	187	20	5	568500
6	159	42	4	511651
7	272	131	5	341103
8	187	159	2	227400
9	261	90	3	341100
10	253	30	5	682201
11	265	129	2	170550
....
100	302	23	10	1364401

Tabel 3. Normalisasi Data LRFM

NO Pelanggan	L	R	F	M
1	1.000	0.098	1.000	0.444
2	0.670	0.059	0.400	0.097
3	0.660	0.764	0.000	0.007
4	0.990	0.181	0.300	0.088
5	0.609	0.047	0.400	0.124
6	0.514	0.134	0.300	0.111
7	0.898	0.484	0.400	0.071



Gambar 1. Hasil Klasterisasi pelanggan PT Pan Baruna

Setelah tahapan normalisasi dilakukan, data tersebut di kluster menggunakan metode Fuzzy C-Means. Data di bagi menjadi 3 kluster yang menghasilkan sebanyak 47 pelanggan pada kluster 1, 34 Pelanggan kluster 2 dan 19 pelanggan pada kluster 3. Selanjutnya kluster yang telah di hasilkan tersebut akan dikalikan dengan bobot nilai LRFM dengan menggunakan metode pembobotan AHP (*Analytical Hierarchy Process*) peneliti menggunakan bobot yang sama dengan penelitian (Parvaneh, Abbasimehr and Tarokh, 2012) dengan nilai W_L , W_R , W_F , W_M yaitu 0.238, 0.088, 0.326, dan 0.348.

Tabel 4. Nilai CLV pada Model LRFM (sebelum dikalikan dengan hasil bobot)

Cluster	Pelanggan	L	R	F	M
1	47	0.832	0.162	0.448	0.216
2	34	0.712	0.739	0.038	0.061
3	19	0.383	0.203	0.123	0.082
Rata-rata		0.642	0.368	0.203	0.120

Tabel 5 Ranging CLV pada setiap kluster pada Model LRFM

kluster	pelanggan	L*WL	R*WR	F*WF	M*WM	CLV	Ranging CLV
1	47	0.198	0.014	0.146	0.075	0.433	1
2	34	0.169	0.065	0.012	0.021	0.268	2
3	19	0.091	0.018	0.040	0.028	0.178	3
Rata-rata		0.153	0.032	0.066	0.042	0.293	

Berdasarkan Tabel 4 dan 5 hasil ranging CLV maka perusahaan bisa membuat strategi dalam mempertahankan pelanggan dan mengakuisisi pelanggan agar menjadi pelanggan yang loyal dengan nilai profitabilitas yang tinggi.

5. KESIMPULAN

Berdasarkan nilai CLV maka segmen pelanggan yang memiliki nilai CLV yang tertinggi yaitu pada kluster 1 dengan nilai CLV 0,433 yang jauh berada di atas nilai rata semua segmen. Maka segmen ini memiliki nilai loyalitas yang tinggi dengan simbol LRFM $L \uparrow R \downarrow F \uparrow M \uparrow$ yaitu *loyal customer* (segmen yang terbaik yang memiliki nilai loyalitas pelanggan yang tinggi). Selanjutnya segmen yang memiliki ranging CLV kedua adalah kluster 2 dengan nilai 0,268 dengan simbol LRFM $L \downarrow R \uparrow F \downarrow M \downarrow$ yaitu *uncertain lost customers* (segmen dengan pelanggan yang hilang tidak pasti). Ranging terakhir berada pada kluster 3 dengan nilai 0.178 dengan simbol LRFM $L \downarrow R \downarrow F \downarrow M \downarrow$ yaitu *uncertain new customers* (segmen dengan pelanggan baru yang tidak pasti).

Kluster dengan nilai CLV yang tertinggi yaitu kluster 1 tersebut memiliki segmen pelanggan berupa loyalitas yang tinggi dengan nilai *length* yang tinggi, *recency* yang rendah, *frequency* yang tinggi dan *Monetary yang tinggi*. Ini berarti pelanggan yang berada pada segmen ini sering melakukan pembelian dengan jumlah uang yang dikeluarkan tinggi untuk perusahaan sehingga segmen ini dikatakan pelanggan dengan profitabilitas yang tinggi. Oleh karena itu pelanggan pada segmen ini perlu dipertahankan dengan memberikan layanan yang terbaik agar mereka tidak menjadi target pemasaran bagi pesaing.

DAFTAR PUSTAKA

- Aniq Noviciatie Ulfah, Shofwatul dan Uyun. 2015. Analisis Kinerja Algoritma *Fuzzy C-Means* Dan *K-Means* Pada Data Kemiskinan, Jatisi, ISSN : 2407-4322, Vol. 1 No. 2.
- Beta Estri Adiana, Indah Soesanti dan Adhistya Erna Permasari, 2018. Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model Dan Teknik *Clustering*, *JUTEI*, ISSN 2579-3675, e-ISSN 2579-5538, DOI 10.21460/jutei.2017.21.76, Volume.2 No.1
- Denny B. Saputra dan Edwin Riksakomara, 2018. Implementasi *Fuzzy C-Means* Dan modelrfm untuk Segmentasi Pelanggan (Studi Kasus : PT. XYZ), *JURNAL TEKNIK ITS*, 2337-3520 (2301-928X Print), Vol. 7, No. 1.
- Fajar Agustini, 2017. Implementasi Algoritma Fuzzy C-Means Studi Kasus Penjualan Di Sushigroove Restaurant, *JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER*, E-ISSN: 2527-4864, VOL. 3. NO. 1.
- Fitri Marisa, Sarifah Shakinah Syed Ahmad, Zeratul Izzah Mohd Yusof, Fachrudin dan Tubagus Mohammad Akhriza Aziz, 2019. Segmentation Model of Customer Lifetime Value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM Mod, *International Journal Of Integrated Engineering*, VOL. 11 NO.3 (2019) 169-180.
- Jemaictry Tamaela, Eko Sedyono dan Adi Setiawan, *Cluster Analysis* Menggunakan Algoritma *Fuzzy C-Means* Dan *K-Means* Untuk Klasterisasi Dan Pemetaan Lahan Pertanian Di Minahasa Tenggara, *Jurnal Buana Informatika*, Volume 8, Nomor 3 151-160.
- Prima Resti Nastiti, Bedi Suprpty dan Bedi Suprpty, 2018. Perbandingan Hasil Algoritma Self Organizing Map (Som) Dan Fuzzy C-Means Clustering Untuk Kualifikasi Data Kinerja Dosen, Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi, e-ISSN 2540-7902 dan p-ISSN 2541-366X, Vol. 3, No. 2.
- Siti Monalisa, 2018. Klasterisasi *Customer Lifetime Value* Dengan Model Lrfm Menggunakan Algoritma K-Means, *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, p-ISSN: 2355-7699, e-ISSN: 2528-6579, Vol. 5, No. 2.
- Sudriyanto, 2017. Clustering Loyalitas Pelanggan Dengan Metode Rfm (Recency, Frequency, Monetary) Dan Fuzzy C-Means, Prosiding SNATIF Ke -4, ISBN: 978-602-1180-50-1.
- Wahidah Sanusi, Ahmad Zaky1, dan Besse Nur Afni, Analisis Fuzzy C-Means dan Penerapannya Dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Faktor-faktor Penyebab Gizi Buruk.