

KLASTERISASI *CUSTOMER LIFETIME VALUE* DENGAN MODEL LRFM MENGUNAKAN ALGORITMA K-MEANS

Siti Monalisa

Jurusan Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau
Email: siti.monalisa@uin-suska.ac.id

(Naskah masuk: 11 Maret 2018, diterima untuk diterbitkan: 28 Mei 2018)

Abstrak

Penelitian ini bertujuan menghasilkan nilai *Customer Lifetime Value* (CLV) pada setiap segmen pelanggan dengan menggunakan algoritma K-means dalam melakukan klusterisasi pelanggan menggunakan model LRFM (Length, Recency, Frequency dan Monetary). Pembentukan kluster menggunakan metode validasi *Dunn Index* dan *Silhouette Coefficient* dengan nilai 0.84 dan 0.54. Kluster yang dihasilkan pada penelitian ini berjumlah 3 kluster dengan nilai yang tertinggi pada masing-masing metode validasi *Dunn Index* dan *Silhouette Coefficient*. Nilai CLV akan dihasilkan dengan mengalikan nilai normalisasi LRFM dengan nilai bobot LRFM dan selanjutnya dijumlahkan. Penjumlahan nilai CLV tersebut dilakukan pada masing-masing kluster yang telah terbentuk. Berdasarkan penjumlahan tersebut maka dicari ranking CLV. Ranking CLV tertinggi dihasilkan dari nilai CLV yang terbesar diantara 3 kluster tersebut. Ranking CLV tertinggi pada penelitian ini adalah kluster ke 2 yaitu 0,208 dengan simbol LRFM $L \uparrow R \downarrow F \uparrow M \uparrow$ dengan arti bahwa kluster ini memiliki segmen pelanggan dengan nilai loyalitas yang tinggi. Ranking CLV ke 2 adalah kluster 3 dengan nilai CLV 0,082 dan selanjutnya kluster 1 dengan nilai CLV 0,038.

Keywords : CLV; *Dunn Index*; K-Means; LRFM dan *Silhouette*

CLUSTERING OF CUSTOMER LIFETIME VALUE BASED ON LRFM MODEL USING K-MEANS ALGORITHM

Abstract

This study aims to generate the value of Customer Lifetime Value (CLV) in every customer segment by using the K-means algorithm in customer clustering with model LRFM (Length, Recency, Frequency dan Monetary). The number of clusters using the Dunn Index validation method and Silhouette Coefficient with the value of 0.84 and 0.54. The resulting cluster is 3 cluster with the highest value on each of the Dunn Index validation methods and Silhouette Coefficient. The CLV value will be generated by multiplying the LRFM normalization value by the LRFM weights and then they are summed. The sum of CLV values is done on each cluster that has been formed. Based on the sum then specified CLV ranking. The highest CLV ranking results from the largest CLV values among 3 clusters. The highest CLV ranking in this study is the second cluster of 0.208 with the symbol LRFM $L \uparrow R \downarrow F \uparrow M \uparrow$ with the meaning that this cluster has a customer segment with high loyalty value. The second CLV ranking is cluster 3 with CLV value 0.082 and last rank is cluster 1 with value CLV 0,038.

Keywords: CLV; *Dunn Index*; K-Means; LRFM and *Silhouette Index*

1. PENDAHULUAN

Pelanggan merupakan aset bagi perusahaan. Saat ini, perusahaan berlomba-lomba dalam memenangkan hati para pelanggan. Karena masing-masing pelanggan memiliki kebutuhan, preferensi harapan dan perilaku yang berbeda-beda maka sebaiknya perusahaan tidak memperlakukan mereka dengan perlakuan yang sama. Pelanggan tidak hanya memiliki kebutuhan, preferensi dan harapan yang berbeda namun juga profil pendapatan dan biaya yang berbeda dan karenanya harus dikelola dengan cara yang berbeda pula (Buttle and Stan, 2015).

Perlakuan yang sama terhadap semua pelanggan akan menyebabkan pelanggan yang tidak begitu bernilai bagi perusahaan akan berakhir menjadi penghancur nilai daripada pencipta nilai bagi perusahaan (Buttle and Stan, 2015). Nilai pelanggan dapat dilihat dengan mengetahui nilai yang dihasilkan saat ini dari semua keuntungan untuk masa depan perusahaan. Konsep ini disebut dengan *Customer Lifetime Value* (CLV) (Khajvand *et al.*, 2011). CLV adalah ukuran perolehan laba pelanggan untuk perusahaan (Buttle and Stan, 2015). Estimasi model CLV ini dapat digunakan untuk mengevaluasi nilai masa depan pelanggan ke perusahaan dengan teknik data mining dalam mendeteksi pola dan

hubungan dalam histori data pelanggan (Buttle and Stan, 2015). Salah satu model yang bisa mengukur CLV adalah model RFM (Recency, Frequency dan Monetary) yang merupakan model penting dalam memperkirakan nilai hidup pelanggan (Parvaneh, Abbasimehr and Tarokh, 2012) dan model ini juga bisa mengukur nilai profitabilitas pelanggan (He and Li, 2016).

RFM merupakan model yang dikembangkan oleh Arthur Hughes (1994) dengan 3 variabel yaitu *recency*, *frequeuncy* dan *monetary* (Kandeil, Saad and Youssef, 2014) dan diperpanjang dengan tambahan satu variabel lain yaitu *length* oleh Chang dan Tasy (2004) yang disebut dengan model LRFM (Length, Recency, Frequency dan Monetary). Variabel *length* ini merupakan variabel yang mengukur lamanya transaksi pertama dan terakhir pelanggan pada satu periode waktu. Hal ini penting diukur dalam memperkuat nilai *recency* yang dihasilkan pelanggan berdasarkan lamanya transaksi terakhir dengan periode waktu tertentu.

Dalam mengukur nilai profitabilitas pelanggan maka hal yang perlu dilakukan adalah mengidentifikasi pelanggan dengan menggunakan pendekatan yang disebut segmentasi pelanggan (Parvaneh, Abbasimehr and Tarokh, 2012). Segmentasi pelanggan mengacu pada pelanggan yang ada dibagi menjadi beberapa kelompok pelanggan sesuai dengan standar tertentu (He and Li, 2016). Metode yang digunakan dalam melakukan segmentasi pelanggan adalah metode klustering. Klustering disebut juga sebagai segmentasi data pada beberapa aplikasi karena mengelompokkan data besar ke dalam kelompok memiliki kemiripan dengan kelompoknya (Han, Kamber and Pei, 2012). Klustering merupakan metode yang banyak digunakan dalam berbagai bidang termasuk didalamnya segmentasi pelanggan, perilaku pelanggan, profitabilitas pelanggan, pengelompokan data kebakaran hutan dan lain sebagainya. Berbagai algoritma yang digunakan dalam klustering yaitu K-Means, DBSCAN (Density Based Spatial Clustering of Application with Noise), SOM (Self Organizing Map) dan lainnya. Metode yang paling umum digunakan adalah algoritma *K-Means* (Berry and Linoff, 2008). Namun Algoritma ini memiliki sensitifitas terhadap *start point* (Hosseini, Maleki and Gholamian, 2010) sehingga akan mempengaruhi validasi jumlah kluster yang dihasilkan. Karena hasil klustering bergantung pada metode atau algoritma yang digunakan maka banyak peneliti yang telah meneliti mengenai metode validasi dalam penentuan jumlah kluster yang terbaik. Metode validasi klustering yang digunakan diantaranya *Dunn Index* dan *Silhouette Coefficient*. *Dunn Index* merupakan metode validasi kluster yang didasarkan pada perhitungan geometri perhitungan geometri dari kekompakan masing-masing kluster dan pemisahan antar kluster (Retno, 2013). Kluster yang

memiliki nilai *Dunn Index* terbesar lah yang menjadi kluster terbaik (Vendramin and Hruschka, 2009). Metode validasi *Dunn Index* digunakan pada penelitian ini dikarenakan penelitian ini menggunakan data numerik. Berdasarkan penelitian (Retno, 2013) metode *Dunn Index* ini memiliki nilai validasi tertinggi untuk data numerik dibandingkan metode validasi lain seperti *Hubert's Statistic* dan *Silhouette Coefficient*. Sedangkan *Silhouette Coefficient* dalam menentukan kluster terbaik yaitu dengan melihat nilai yang mendekati angka 1 (Retno, 2013).

Oleh karena itu penelitian ini bertujuan menghasilkan *Customer Lifetime Value* dengan algoritma K-Means menggunakan metode validasi *Dunn Index* dan *Silhouette Coefficient* sebagai penentuan jumlah kluster terbaik..

2. KERANGKA TEORI

2.1. Customer Lifetime Value (CLV)

CLV digunakan dalam menghitung nilai profitabilitas pelanggan. CLV dihitung setelah melakukan segmentasi pelanggan. Nilai CLV dihitung berdasarkan pada rangking CLV yang ditentukan untuk setiap segmen (Khajvand *et al.*, 2011). Perhitungan CLV menggunakan persamaan i.

$$C^j = W_L C_L^j + W_R C_R^j + W_F C_F^j + W_M C_M^j$$

(i)

Dimana :

C^j = Peringkat CLV pelanggan j

$C_L^j, C_R^j, C_F^j, C_M^j$ = Normalisasi L, R, F dan M dari kluster j

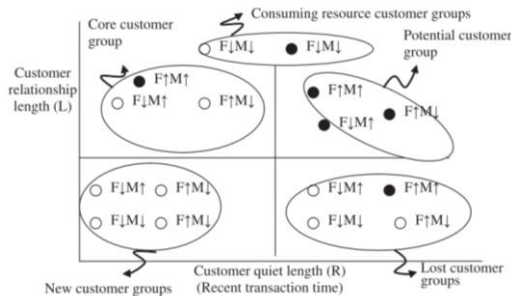
W_L, W_R, W_F, W_M = Bobot yang dihasilkan L,R,F dan M dari hasil AHP.

2.2 Model LRFM

Model LRFM merupakan pengembangan dari model RFM yang dikembangkan oleh Arthur Hughes (Kandeil, Saad and Youssef, 2014). RFM merupakan model yang banyak digunakan dalam melakukan segmentasi perilaku pelanggan. Model RFM ini terdiri dari *Recency*, *Frequency* dan *Monetary*. Selanjutnya Chang dan Tsay menambah satu variabel yaitu *Length* (Kandeil, Saad and Youssef, 2014).

- Length* yaitu lamanya hubungan antara pelanggan dan perusahaan yang diukur selama periode analisis
- Recency* yaitu tanggal terakhir transaksi yang dilakukan pelanggan pada perusahaan selama periode analisis.
- Frequency* adalah jumlah transaksi yang dilakukan oleh pelanggan pada perusahaan selama periode yang dianalisis
- Monetary* yaitu jumlah uang yang dikeluarkan pelanggan untuk perusahaan selama periode analisis

Metode segmentasi dengan LRFM memberi nilai sesuai dengan nilai pelanggan (Kandeil, Saad and Youssef, 2014). Nilai pelanggan akan disegmentasikan sesuai segmennya masing-masing pada 16 kluster yang terdapat pada Gambar 1.



Gambar 1 Matriks Loyalitas Pelanggan Chang dan Tsay (2004)

2.2. Algoritma K-Means

Metode K-Means adalah metode yang dikategorikan ke dalam metode clustering partisi (Kandeil, Saad and Youssef, 2014). Algoritma ini paling banyak digunakan dalam CRM dan pemasaran (Kandeil, Saad and Youssef, 2014). Hal ini disebabkan karena algoritma ini memiliki kesederhanaan dan kemudahan dalam menggunakannya serta pengguna bisa menentukan sendiri jumlah cluster nya. Tahapan dalam metode K-means adalah sebagai berikut :

- [1] Tentukan jumlah kluster
- [2] Pilih centroid awal secara acak sesuai jumlah kluster
- [3] Hitung jarak data ke centroid dengan rumus *euclidean distance*.

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

- [4] Perbarui centroid dengan menghitung nilai rata-rata nilai pada masing-masing kluster
- [5] Kembali ke tahapan ke 3 jika masih terdapat data yang berpindah kluster atau perubahan nilai centroid

Pada tahap 1 dan 2 dari tahapan K-Means, jumlah kluster ditentukan dengan metode dunn index dan *silhouette coefficient* yang merupakan metode validasi dalam menentukan jumlah kluster yang terbaik.

2.3 Dunn Index

Metode Dunn Index merupakan kriteria validitas yang didasarkan pada perhitungan geometri dari kekompakan setiap cluster dan pemisahan antar cluster (Retno, 2013). Dunn index tidak memiliki suatu rentang nilai, untuk mencari indeks Dunn terbaik dapat dilihat dari nilai terbesar yang dihasilkan (Retno, 2013). Kluster terbaik dilihat

dengan nilai DN yang terbesar (Vendramin et al. 2009). Dunn index di hitung berdasarkan persamaan berikut :

$$D_{nc} = \min_{i=1, \dots, nc} \left\{ \min_{j=i+1, \dots, nc} \left(\frac{d(c_i, c_j)}{\max_{k=1, \dots, nc} diam(c_k)} \right) \right\} \quad (3)$$

Dimana $d(c_i, c_j)$ adalah fungsi yang tidak sama antara kluster c_i dan c_j yang didefinisikan sebagai :

$$d(c_i, c_j) = \min_{x \in c_i, y \in c_j} d(x, y) \quad (4)$$

dan $diam(C)$ adalah diameter kluster yang mungkin dipertimbangkan sebagai ukuran dispersion kluster. Diameter kluster C bisa dedefinisikan sebagai :

$$diam(C) = \max_{x, y \in C} d(x, y) \quad (5)$$

2.4 Silhouette Coefficient

Metode validiy yang lain adalah *Silhouette Coefficient* didasarkan pada pertimbangan geometri tentang *cohesion* dan *Separation*. *Silhouette Index* dikembangkan pertama kali oleh Rousseeuw pada tahun 1986. Nilai *Silhouette Index* (SI) didapatkan dengan persamaan sebagai berikut :

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \quad (6)$$

Dimana:

- $s(i)$ = nilai *Silhouette coefficient*
- $a(i)$ = rata-rata jarak antara titik i dengan seluruh titik dalam A (*cluster* dimana titik i berada)
- $b(i)$ = rata-rata jarak antara titik i ke seluruh titik dalam *cluster* selain A.

Hasil perhitungan nilai *Silhouette Coefficient* berada pada range -1 hingga 1 (Kodinariya and Makwana, 2013). Nilai tertinggi yang memiliki nilai terbaik dalam pembentukan kluster. Nilai $0.7 < SC \leq 1$ (Strong Structure), Nilai $0.5 < SC \leq 0.7$ (Medium Structure), Nilai $0.25 < SC \leq 0.5$ (Weak Structure) dan Nilai $SC \leq 0.25$ (No structure).

3. METODOLOGI

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 :

$$B = \left(\frac{(A - \min A)}{(\max A - \min A)} \right) * (D - C) + C \quad (7)$$

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. Penentuan jumlah kluster optimal dengan metode validasi *Dunn Index* dan *Silhouette Index*
 - d. Melakukan klusterisasi dengan algoritma *K-Means* berdasarkan jumlah kluster yang terbaik
 - e. Melakukan pembobotan nilai LRFM dengan metode AHP.
 - f. Mengalikan nilai bobot LRFM dengan rata-rata nilai LRFM dan menjumlahkannya seperti pada persamaan 8 berikut :
- $$CLV = L * W_L + R * W_R + F * W_F + M * W_M \quad (8)$$

dimana :

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

L_w, R_w, F_w, M_w =Bobot masing-masing nilai LRFM

- g. Merangking dan menganalisis hasil klusterisasi dengan rank LRFM.

4. HASIL DAN PEMBAHASAN

Penelitian menggunakan data HPAI Cabang Pekanbaru dengan total pelanggan 1.867 pada periode analisis Februari 2017 sampai September 2017. 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 2).

Tabel 1. Data LRFM

No. Pelanggan	L	R	F	M
1	0	201	1	385,000
2	71	174	4	2,244,000
3	126	4	5	4,225,000
4	134	121	3	116,000
5	124	148	2	193,000
6	50	202	2	2,058,000
7	179	3	2	180,000
8	197	59	7	1,248,000
9	0	243	1	100,000
...
1867	125	121	7	1,155,000

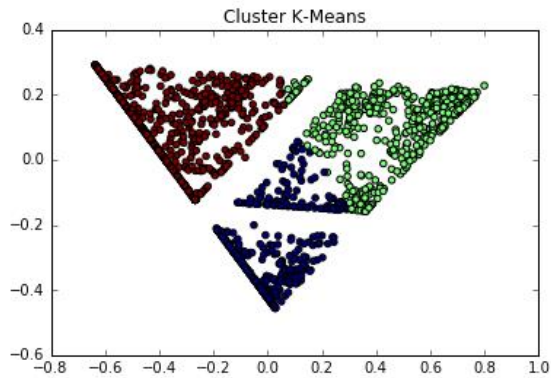
Tabel 2. Hasil Normalisasi Data

No. Pelanggan	L _N	R _N	F _N	M _N
1	0,0000	0,7390	0,0000	0,0007
2	0,2620	0,6397	0,0144	0,0041
3	0,4649	0,0147	0,0192	0,0078
4	0,4945	0,4449	0,0096	0,0002
5	0,4576	0,5441	0,0048	0,0004
6	0,1845	0,7426	0,0048	0,0038
7	0,6605	0,0110	0,0048	0,0003
8	0,7269	0,2169	0,0288	0,0023
9	0,0000	0,8934	0,0000	0,0002
...
1867	0,4613	0,4449	0,0288	0,0021

Setelah tahapan normalisasi dilakukan, data tersebut di kluster dengan metode *K-means* dengan metode validasi *Dunn Index* dan *Silhouette Coefficient* dalam menentukan kluster yang terbaik dengan menguji pembentukan kluster yang terbentuk dari kluster 3 sampai dengan kluster 9. Penentuan kluster dari kluster 3 hingga kluster 9 pada penelitian ini adalah karena semakin banyak kluster semakin kecil nilai *Dunn Index* dan *Silhouette Coefficient* sehingga diputuskan untuk menghentikan jumlah kluster sampai dengan 9. Nilai validasi kedua metode ini ditunjukkan pada Tabel 3 dan Gambar 2.

Tabel 3. Nilai Dunn Index dan Silhouette Coefficient

Jumlah cluster	Nilai Dunn Index	Nilai Silhouette Coefficient
3	0.84	0.54
4	0.71	0.52
5	0.61	0.49
6	0.53	0.50
7	0.52	0.49
8	0.52	0.50
9	0.36	0.50



Gambar 2 Plot Tiga Kluster

Selanjutnya kluster yang memiliki nilai Dunn Index dan Silhouette Index yang tertinggi akan dijadikan hasil kluster pada penelitian ini. Kluster yang telah dihasilkan tersebut akan dikalikan dengan bobot nilai LRFM dengan menggunakan metode pembobotan AHP (Analytical Hierarchy Process). Dalam hal ini, 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. Bobot ini digunakan karena setiap perusahaan memiliki pendapat masing-masing mengenai prioritas pada setiap atribut L, R, F dan M. Pelanggan pada penelitian (Parvaneh, Abbasimehr and Tarokh, 2012) ini adalah pelanggan retail. Hal ini sama dengan pelanggan yang diteliti pada penelitian ini yaitu pelanggan retail.

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

Kluster	Jumlah Pelanggan	L	R	F	M
1	517	0,094	0,145	0,007	0,001
2	571	0,724	0,111	0,059	0,019
3	779	0,080	0,683	0,006	0,002
Rata-Rata		0,299	0,313	0,024	0,007

Tabel 5 Ranking CLV pada Setiap Kluster pada Model LRFM

Kluster	Jumlah Pelanggan	$L * W_L$	$R * W_R$	$F * W_F$	$M * W_M$	CLV	Ranking CLV
1	517	0,022	0,013	0,003	0,0003	0,038	3
2	571	0,172	0,010	0,019	0,007	0,208	1
3	779	0,019	0,060	0,002	0,0007	0,082	2
Rata-Rata		0,071	0,028	0,01	0,003	0,109333	

Berdasarkan Tabel 4 dan Tabel 5 diatas maka segmen pelanggan yang memiliki nilai CLV yang tertinggi yaitu pada kluster 2 dengan nilai CLV 0,208 yang jauh berada di atas nilai rata semua segmen. Berdasarkan Gambar 1 dari penelitian Chang dan Tsay (2004) 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 rangking CLV ke dua adalah kluster 3 dengan nilai 0,082 dengan simbol LRFM $L \downarrow R \uparrow F \downarrow M \downarrow$ yaitu *uncertain lost customers* (segmen dengan pelanggan yang hilang tidak pasti). Rangking terakhir berada pada kluster 1 dengan simbol LRFM $L \downarrow R \downarrow F \downarrow M \downarrow$ yaitu *uncertain new customers* (segmen dengan pelanggan baru yang tidak pasti). Berdasarkan simbol LRFM tersebut maka perusahaan bisa membuat strategi dalam mempertahankan pelanggan dan mengakuisisi pelanggan agar menjadi pelanggan yang loyal dengan nilai profitabilitas yang tinggi.

5. KESIMPULAN

Penelitian ini menghasilkan 3 kluster yang terbaik dengan metode validasi kluster dengan metode dunn index dan Silhouette Coefficient. Kluster yang telah terbentuk tersebut akan dianalisis untuk mencari nilai CLV nya. CLV berguna dalam menentukan nilai profitabilitas dan loyalitas pelanggan dengan model LRFM. Berdasarkan nilai CLV maka rangking tertinggi diantara 3 kluster tersebut berada pada kluster ke dua dengan nilai CLV berada jauh tertinggi dari rata-rata kluster lainnya yaitu 0,208. Rangking kedua yaitu berada pada kluster 3 dengan nilai CLV 0,082. Rangking ke tiga yaitu pada kluster 1 dengan nilai CLV 0,038.

Kluster dengan nilai CLV yang tertinggi yaitu kluster 2 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.

UCAPAN TERIMA KASIH

Terimakasih penulis ucapkan sebesar-besarnya kepada keluarga, institusi dan perusahaan yang telah memberikan kesempatan dalam menyelesaikan penelitian ini. Semoga penelitian ini memberikan manfaat dan kontribusi yang besar terhadap organisasi dan masyarakat.

DAFTAR PUSTAKA

- BERRY, M. J. A. dan LINOFF, G. S., 2008. Data Mining Techniques : For Marketing, Sales, And Customer Relationship Management. second edition.
- BUTTLE, F. dan STAN, M., 2015. Customer Relationship Management. Third edit. Butterworth-Heinemann.
- DUNN, J. C., 1974. Well-separated clusters and optimal fuzzy partitions, *Journal of Cybernetics*, 4(1), pp. 95–104. doi: 10.1080/01969727408546059.
- HAN, J., KAMBER, M. dan PEI, J., 2012. Data Mining: Concepts and Techniques, San Francisco, CA, itd: Morgan Kaufmann. doi: 10.1016/B978-0-12-381479-1.00001-0.
- HE, X. dan LI, C., 2016. The Research and Application of Customer Segmentation on E-commerce Websites. doi: 10.1109/ICDH.2016.47.
- HOSSEINI, S. M. S., MALEKI, A. dan GHOLAMIAN, M. R., 2010. Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty, *Expert Systems with Applications*. Elsevier Ltd, 37(7), pp. 5259–5264. doi: 10.1016/j.eswa.2009.12.070.
- KANDEIL, D. A., SAAD, A. A. dan YOUSSEF, S. M., 2014. A two-phase clustering analysis for B2B customer segmentation, *Proceedings - 2014 International Conference on Intelligent Networking and Collaborative Systems*, IEEE INCoS 2014, pp. 221–228. doi: 10.1109/INCoS.2014.49.
- KHAJVAND, M., ZOLFAGHAR, K., ASHOORI, S. dan ALIZADEH, S., 2011. Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study, *Procedia Computer Science*. Elsevier, 3, pp. 57–63. doi: 10.1016/j.procs.2010.12.011.
- PARVANEH, A., ABBASIMEHR, H. dan TAROKH, M. J., 2012. Integrating AHP and data mining for effective retailer segmentation based on retailer lifetime value, *Journal of Optimization in Industrial Engineering*, 5(11), pp. 25–31.
- RETNO, D., 2013. Perbandingan Metode Cluster Validity Pada, *Skripsi Fakultas Matematika Dan Ilmu Pengetahuan Alam*, Institut Pertanian Bogor.
- VENDRAMIN, L. AND HRUSCHKA, E. R., 2009. On the Comparison of Relative Clustering Validity Criteria. pp. 733–744.