

SEGMENTASI LOYALITAS PELANGGAN DENGAN MODEL RFM MENGUNAKAN ALGORITMA CLARANS

Novianto Heri Darmawan^{1*}, Yulison Herry Crisnanto¹ dan Puspita Nurul Sabrina¹

¹Jurusan Teknik Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani
Jl. Terusan Jend. Sudirman, Cibeer, Kec. Cimahi Sel., Kota Cimahi, Jawa Barat 40531.

*Email: noviantoherid17@if.unjani.ac.id

Abstrak

Dalam upaya menentukan kelompok-kelompok pelanggan yang berpotensi loyal perlu dilakukan pemeriksaan yang cermat berdasarkan karakteristik setiap pelanggan dalam bertransaksi. Dengan memiliki kesamaan karakteristik untuk membuat suatu pengelompokan pelanggan juga di perlukan dalam konsep segmentasi pelanggan. Hal tersebut diperlukan karena untuk mengetahui perilaku-pelaku pelanggan sehingga dapat membantu dalam penerapan strategi penjualan yang tepat untuk meningkatkan keuntungan bagi pihak perusahaan. Tujuan penelitian ini yaitu mengidentifikasi kelompok pelanggan - pelanggan dengan algoritma CLARANS berdasarkan model RFM. Metode CLARANS ialah metode klasterisasi yang berbasis partisi k-medoid, dengan menggunakan abstraksi graph untuk penentuan k-medoidsnya. Dalam penelitian ini membahas mengenai segmentasi loyalitas pelanggan menggunakan metode CLARANS dan berdasarkan model RFM. Berdasarkan hasil analisis interpretasi RFM yang dilakukan memiliki objek yang terbentuk sebanyak 3 cluster dengan 787 pelanggan berada di cluster 1, 1456 pelanggan berada di cluster 2, 2798 pelanggan berada di cluster 3. Adapun hasil pengujian akurasi terdapat beberapa nilai K, hasil nilai yang paling tinggi dengan mendekati 1 yaitu pada nilai K = 3 dengan nilai Silhouette Coefficient 0.766840732467501 dimana nilai K = 3 tersebut dikategorikan "Strong Classification". Dapat di simpulkan penelitian ini dapat mengelompokan tingkat loyalitas pelanggan menggunakan model RFM dan algoritma CLARANS.

Kata kunci: Segmentasi; Loyalitas Pelanggan; Clustering; RFM; CLARANS

1. PENDAHULUAN

1.1. Latar Belakang

Loyalitas pelanggan termasuk dalam aset yang berharga bagi tiap perusahaan, karena perusahaan akan mendapatkan profit dalam membuat pelanggan loyal, yang dimana pelanggan-pelanggan ini akan memiliki komitmen terhadap produk. Maka dengan meningkatkan loyalitas pelanggan perusahaan akan semakin berkembang dan dapat bertahan dari banyaknya saingan bisnis yang semakin ketat (Kristanty MN Nadapdap & Diren Pradiva, 2020).

Di Indonesia beberapa perusahaan lokal yang bergerak dalam bidang fashion telah menggunakan teknologi e-commerce dan situs web sebagai media pemasaran. Hal serupa yang dilakukan oleh perusahaan XYZ, yang dimana dalam memasarkan produknya dengan menggunakan e-commerce. Data yang digunakan pada penelitian ini di dapatkan pada perusahaan XYZ sebanyak 5093 record data dari bulan Januari hingga bulan Juni 2021

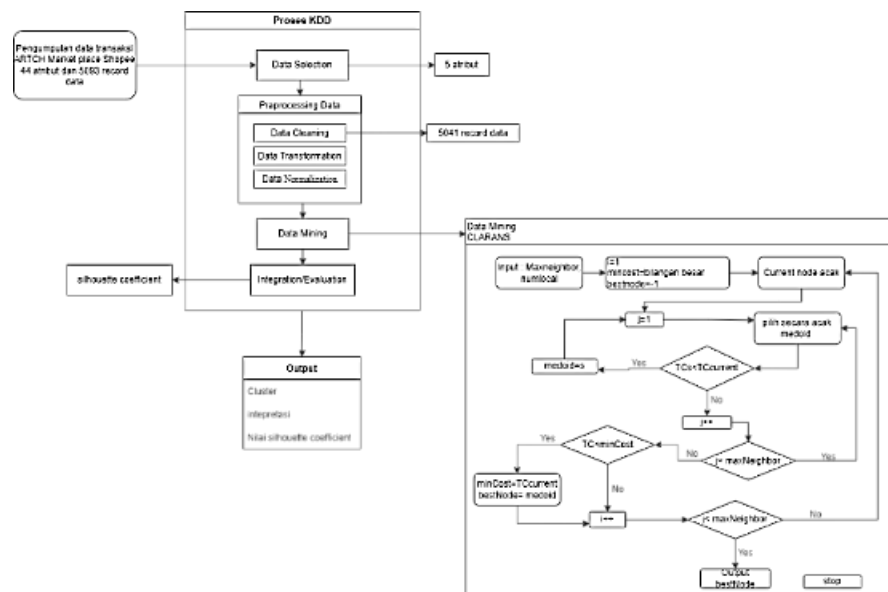
Karakteristik tiap pelanggan bisa dilihat dengan menerapkan model RFM. Model RFM ialah Recency merupakan waktu terakhir pelanggan melakukan transaksi, Frequency merupakan jumlah transaksi, dan Monetary merupakan jumlah uang yang dikeluarkan. Model RFM banyak digunakan karena mudah digunakan dan cepat diimplementasikan (Maryani et al., 2018). Klasterisasi dapat diimplementasikan dengan menggunakan teknik klasterisasi dalam data mining yaitu dengan metode CLARANS (*Clustering Algorithm based on Randomized Search*) (Mandrasi, 2018).

Metode CLARANS ialah metode klasterisasi yang berbasis partisi k-medoid, dengan menggunakan abstraksi graph untuk penentuan k-medoidsnya. Algoritma ini merupakan bagian dari algoritma PAM & CLARA. Pencarian graph secara random (acak) untuk mendapatkan beberapa medoid yang mewakili sejumlah cluster merupakan cara kerja dari algoritma ini (Schubert & Rousseeuw, 2021). Medoid ialah

data poin yang terletak ditengah dari sekumpulan data. Parameter dalam Algoritma CLARANS ini dengan menggunakan *maxneighbour* dan *numlocal*. Definisi dari *Maxneighbour* ini ialah nilai maksimum dari kumpulan node yang diuji. Sedangkan *Numlocal* ialah nilai maksimal dari lokal minimum (Setiyono & Mukhlash, 2005). Metode CLARANS memiliki keunggulan dalam segi keefisienan yaitu melakukan pengecekan terhadap sample tetangga dari suatu node saja, maka jika dilihat dari segi waktu pemrosesan metode CLARANS jauh lebih singkat. Sehingga dalam penelitian ini segmentasi pelanggan menggunakan metode CLARANS dengan menerapkan model RFM dengan menggunakan sumber data non-spasial. Sumber data yang digunakan ini ialah data transaksi dengan rentang waktu bulan Januari hingga Juni 2021 sebanyak 5093 record data.

Sehingga untuk menentukan karakteristik pelanggan dengan menggunakan algoritma CLARANS berdasarkan model RFM untuk segmentasi pelanggan maka diharapkan proses segmentasi akan menghasilkan kelompok pelanggan yang akurat sebagai acuan untuk menentukan strategi yang dapat mempertahankan hubungan pelanggan pada perusahaan XYZ.

2. METODOLOGI



Gambar 1. Metode Penelitian

1.1. Pengumpulan Data

Pengumpulan data yang dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian. Tujuan dari langkah pengumpulan data dan teknik pengumpulan data ini adalah demi mendapatkan data yang valid, sehingga hasil dan kesimpulan penelitian pun tidak akan diragukan kebenarannya. Data yang digunakan pada penelitian ini yaitu data penjualan dari bulan Januari hingga Juni 2021 sebanyak 5093 record, data yang diperoleh data penjualan online E-Commerce Shopee ARTCH. ARTCH merupakan perusahaan yang bergerak dibidang fashion yang berada di daerah Bandung.

1.2. Recency, Frequency, Monetary (RFM)

RFM merupakan proses tahapan menganalisis berdasarkan tiga variable yaitu, pelanggan belanja (*recency*), jumlah transaksi pelanggan (*frequency*), dan seluruh biaya yang dihabiskan selama melakukan transaksi dalam 1 priode (*monetary*) (Jamal & Yanto, 2019) (Angelie, 2017). *Recency* (R) diperoleh dari selisih hari sejak transaksi terakhir kali dilakukan oleh setiap pelanggan per tanggal 7 Februari 2022.

Frequency (F) diperoleh dari jumlah transaksi yang dilakukan oleh setiap pelanggan. *Monetary* (M) didapatkan dengan menjumlahkan setiap transaksi masing – masing pelanggan.

Menentukan score R, F dan M yang dikonversi ke dalam skor 1 sampai 5. Dimana bobot 1 menunjukkan nilai paling kecil, bobot 2 menunjukan nilai diantara nilai minimum sampai nilai rata – rata, bobot 3 menunjukkan nilai rata – rata, bobot 4 menunjukan nilai diantara nilai rata – rata sampai nilai maksimum, dan bobot 5 menunjukan nilai maksimum dari seluruh data transaksi (Jamal & Yanto, 2019)(Hidayatullah et al., 2018).

Tabel 1. Score RFM

Score	5	4	3	2	1
<i>Recency</i> (hari)	< 200	201 – 270	271 – 320	321 - 365	> 365
<i>Frequency</i> (kali)	> 20	10 – 15	6 – 9	2 - 5	1
<i>Monetary</i> (rb)	> 500	216 - 400	110 - 250	69 - 100	< 50

1.3. Algoritma CLARANS

Algoritma CLARANS (*Clustering Large Application based on RANdomized Search*) merupakan salah satu algoritma varian dari metode *K-medoid*. Algoritma CLARANS juga merupakan pengembangan algoritma-algoritma dari sebelumnya yang menerapkan metode *K-medoid*, seperti PAM (*Partitioning Around Medoid*) dan CLARA (*Clustering Large Application*) yang menangani jumlah data lebih dari 1000 *record* data . Berikut merupakan tahapan dari algoritma CLARANS:

1. Masukkan parameter *Numlocal* dan *maxneighbor*. Inisialisasi $i = 1$, dan *mincost* = bilangan yang besar.
2. Set *current* menjadi node yang acak pada, $G_{n,k}$
3. Set $j = 1$.
4. pilih secara *random neighbor* (*S*) dari *current node*. Hitung selisih *cost* dari 2 node (*current* dan *S*) berdasarkan persamaan (2.9).
5. jika *S* punya *cost* lebih rendah, maka *set current node* = *S*, dan balik ke langkah (3).
6. sebaliknya, jika $S > \text{current node}$, maka increment $j \rightarrow j = j + 1$. lalu cek jika $j \leq \text{maxneighbor}$, kembali ke langkah (4).
7. sebaliknya jika $j > \text{maxneighbor}$, bandingkan *cost* antara *S* dengan *mincost*. Jika $S \ll \text{mincost}$, maka *mincost* = *cost S* dan *bestnode* = *S*.
8. lalu increment $i \rightarrow i = i + 1$. jika $i > \text{numlocal}$, maka *output* = *best node* dan proses berhenti. Jika $i < \text{numlocal}$, kembali ke langkah (2). Proses berhenti jika nilai *numlocal* telah terpenuhi.

Pencarian node dilakukan dengan biaya lebih rendah dari langkah (3) hingga (6) di atas. Namun, jika node saat ini dibandingkan dengan jumlah maksimum peer (*maxneighbor*) dan masih memberikan biaya terendah, node saat ini dinyatakan sebagai minimum lokal. Kemudian, pada langkah (7), biaya minimum lokal ini dibandingkan dengan biaya terendah yang diperoleh selama ini. Biaya terendah dari dua biaya di atas disimpan dalam *mincost*. Algoritma CLARANS kemudian mengulangi pencarian lokal minimum lainnya sampai nilai *numlocal* terpenuhi (Ng & Han, 2002).

3. HASIL DAN PEMBAHASAN

3.1. Perolehan Data

Sebelum melanjutkan ke tahap Proses KDD, diperlukanya pengumpulan dataset sebagai objek untuk untuk mengekstrak informasi. Dataset yang digunakan dalam penelitian ini adalah data transaksi penjualan *E-commers Shopee* dari bulan Januari hingga bulan Juni 2021 yang terdiri dari 5093 record data, serta 44 atribut data.

3.2. Proses Knowledge Discovery in Database

Proses *Knowledge Discovery in Database* merupakan keseluruhan proses dalam data untuk mencari & identifikasi sebuah pola (Kurniawan, 2014). Proses ini dapat dilihat sebagai berikut :

1. *Data Selection*

Dalam tahap data *selection* dilakukan pengumpulan data yang diperlukan yaitu menggunakan data transaksi *E-commerce* Shopee dari bulan Januari – Juni 2021 memiliki 44 atribut dengan 5093 *record* data.

Tabel 2. Data Mentah

No. Pesanan	Status Pesanan	Status Pembatalan	No. Resi	Opsi Pengiriman	...	Waktu Pesanan Selesai
200101EVRNFQ68	Selesai		JP2862445378	J&T	...	01/01/2021 14.20
200101FQNXX4U7	Selesai		JP6005027015	J&T	...	01/01/2021 08.20
...
2007016RCXQYBA	Selesai		JP7243603448	J&T	...	30/06/2021

Dari 44 variabel akan dilakukan seleksi data untuk menseleksi data yang dibutuhkan untuk mengidentifikasi nilai RFM. Dari 44 variabel menjadi 5 variabel untuk dilakukan pengolahan data mining.

2. *Pra-processing* Data

Selanjutnya akan dilakukan *Praprocessing* data sebelum masuk ke *clustering*. persiapan data dilakukan agar data mentah yang telah dikumpulkan dapat diolah sesuai dengan kriteria yang dibutuhkan. Masukan tahap ini adalah data transaksi penjualan yang sebelumnya diperoleh.

a. *Data Cleaning*

Data cleaning, bertujuan untuk membersihkan data – data yang memiliki atribut kosong, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi) yang diperlukan untuk menentukan model RFM. Setelah dilakukan *Cleaning* data, di dapatkan 4065 *record* data dari 5093 *record*.

b. *Data Transformation*

Selanjutnya data akan ditransformasikan menjadi variabel RFM. Data transaksi pelanggan tidak dapat langsung dijadikan masukan dalam melakukan segmentasi. Maka dari itu perlu di lakukan transformasi data berdasarkan RFM.

Tabel 1 Data Transformation

NO	Username	Nama	Recency	Frequency	Monetary
1	septiariani0609	Septiariani Babellia	365	2	37000
2	farhanmoechtar	Farhan Basyirullah	365	1	45000
3	ajirestufauzi	Aji restu fauzi	365	1	45000
...
4065	pathinhaifakhn_	Pathin Haifa Khairunnisa	185	1	208000

c. *Data Normalization*

Setelah mengetahui masing – masing variabel R, F, dan M. maka selanjutnya data akan dinormalisasi dengan metode *min – max*. *NRecency* mewakili nilai normalisasi dari *recency*, *NFrequency* mewakili nilai normalisasi dari *frequency*, dan *NMonetary* mewakili nilai normalisasi dari *monetary*.

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Tabel 2 Data Normalization

NO	Username	Nama	NRecency	NFrequency	NMonetary
1	septiariani0609	Septiariani Babellia	1	0,028571	0,024787
2	farhanmoechtar	Farhan Basyirullah	1	0	0,030984
3	ajirestufauzi	Aji restu fauzi	1	0	0,030984
...
4065	pathinhaifakhn_	Pathin Haifa Khairunnisa	0	0	0,157242

3. Data Mining

Selanjutnya masuk dalam tahap data *Mining*, dimana pada tahap ini menggunakan teknik *clustering* dengan algoritma CLARANS. Algoritma CLARANS juga menerapkan metode *K-medoid* yang di kembangkan dari algoritma-algoritma sebelumnya, seperti PAM (*Partitioning Around Medoid*) dan CLARA (*Clustering Large Application*) algoritma CLARANS dapat menangani jumlah data lebih dari 1000 record (Schubert & Rousseeuw, 2021). Pada penelitian ini sendiri menggunakan 4065 record data transaksi penjualan pada *marketplace* shopee.

```

Set mincost to a maxneighbor;
For i = 1 to numlocal do
  Randomly select a node as the current node C in the graph;
  Initialize j to 1;
  Repeat
    Randomly select a neighbor N of C;
    If Cost(N, D) < Cost(C, D) Then
      Assign N as the current node C;
      Reset j to 1;
    Else increment j by 1;
    End-if;
  Until j > maxneighbor;
  If Cost(C, D) < mincost Then mincost = Cost(C, D);
  bestnode = C;
  End-if;
End-for;
Return bestnode;

```

3.3. Integration / Evaluation

Pada tahapan evaluation pengujian dilakukan terhadap data yang digunakan dengan memasukan beberapa nilai K kemudian dilakukan perhitungan nilai *silhouette coefficient* dari hasil *clustering* setiap nilai K yang dimasukkan.

Tabel 3 Hasil Pengujian Cluster

<i>Cluster</i>	<i>Silhouette Coefficient</i>
2	0.685783504571510
3	0.766840732467501
4	0.368239997819880
5	0.668239997819880

Didapatkan hasil nilai yang paling tinggi dengan mendekati 1 yaitu pada nilai K = 3 dengan nilai *Silhouette Coefficient* 0.766840732467501 dimana nilai K = 3 tersebut dikategorikan “*Strong Classification*”. Sedangkan jika nilai K mendekati nilai 0 maka semakin buruk cluster tersebut. Penilaian nilai K berdasarkan interpretasi *Kauffman*.

3.4. Interpretasi

Dalam tahap ini cluster yang telah diperoleh akan diimplementasikan dalam sebuah laporan agar lebih mudah dipahami oleh *User*. Hasil perhitungan pada tahap modeling menggunakan algoritma CLARANS. Hasil cluster ditunjukkan pada tabel.

Tabel 4 Intepretasi

Cluster	Jumlah data	R	F	M
Cluster 1	787	5	4	4
Cluster2	1456	3	1	2
Cluster3	2798	4	3	2

Setelah didapatkan nilai dari masing-masing atribut *Recency*, *Frequency* dan *Monetary* selanjutnya yang perlu dilakukan adalah memberikan mengidentifikasi karakter pelanggan yang telah di bahas pada tinjauan pustaka. Berdasarkan karakteristik pelanggan yang telah dibuat maka dapat di tampilkan pada tabel dibawah ini.

Tabel 5 Analisis Karakteristik

Cluster	Jumlah data	Karakteristik
1	787	<i>Golden Customer</i>
2	1456	<i>Typical Customer</i>
3	2798	<i>Everyday shopper</i>

4. KESIMPULAN

Hasil dari penelitian segmentasi loyalitas pelanggan dengan model RFM menggunakan algoritma CLARANS disini analisis dari interpretasi RFM yang dilakukan memiliki objek yang terbentuk sebanyak 3 cluster dengan 787 pelanggan berada di cluster 1 masuk dalam kelompok customer Golden Customer, 1456 pelanggan berada di cluster 2 dengan masuk dalam kelompok customer Typical Customer, 2798 pelanggan berada di cluster 3 dengan masuk dalam kelompok customer Everyday Shopper pada perusahaan ARTCH.

Adapun pengujian akurasi dalam sistem terdapat beberapa nilai K didapatkan hasil nilai yang paling tinggi dengan mendekati 1 yaitu pada nilai K = 3 dengan nilai *Silhouette Coefficient* 0.766840732467501 dimana nilai K = 3 tersebut dikategorikan "*Strong Classification*". Sedangkan jika nilai K mendekati nilai 0 maka semakin buruk cluster tersebut. Penelian nilai K berdasarkan interpretasi Kauffman.

Saran pengembangan penelitian diberikan agar penelitian ini dapat berkembang untuk pengembangan penelitian selanjutnya, berikut saran-saran yang diusulkan:

1. Penelitian ini dapat dikembangkan dengan menggunakan metode-metode *clustering* lainnya untuk melakukan perbandingan.
2. Variasi data dari variabel Frekuensi di harapkan data lebih variatif untuk data yang akan di mining.

DAFTAR PUSTAKA

- Angelie, A. V. (2017). Segmentasi Pelanggan Menggunakan Clustering K-Means Dan Model Rfm (Studi Kasus : Pt . Bina Adidaya Surabaya) Customer Segmentation Using K-Means Clustering and Rfm Model (Case Study : Pt . Bina Adidaya Surabaya Segmentasi Pelanggan Menggunakan Clusteri. 163.
- Hidayatullah, D. P., Rokhmawati, R. I., & Perdanakusuma, A. R. (2018). Analisis Pemetaan Pelanggan Potensial Menggunakan Algoritma K-Means dan LRFM Model Untuk Mendukung Strategi Pengelolaan Pelanggan (Studi Pada Maninjau Center Kota Malang). Jurnal Pengembangan

- Teknologi Informasi Dan Ilmu Komputer, 2(8), 2406–2415.
- Jamal, J., & Yanto, D. (2019). Analisis RFM dan Algoritma K-Means untuk Clustering Loyalitas Customer. *Energy*, 9(1), 1–8.
- Kristanty MN Nadapdap, & Diren Pradiva. (2020). Jurnal Pengaruh Segmentasi Pasar Dan Promosi Terhadap Loyalitas Pelanggan Pada Pt. Laris Sumut Makmur Medan . VI, 207–217.
- Kurniawan. (2014). Data Mining. *Journal of Chemical Information and Modeling*, 53(9), 8–24.
- Mandrasi, A. (2018). Aplikasi Data Mining Menggunakan Metode CLARANS Untuk Klasterisasi Gempa Bumi Signifikan di Indonesia.
- Maryani, I., Riana, D., Astuti, R. D., Ishaq, A., Sutrisno, & Pratama, E. A. (2018). Customer segmentation based on RFM model and clustering techniques with K-means algorithm. *Proceedings of the 3rd International Conference on Informatics and Computing, ICIC 2018*, 1–6. <https://doi.org/10.1109/IAC.2018.8780570>
- Ng, R. T., & Han, J. (2002). CLARANS: A method for clustering objects for spatial data mining. *IEEE Transactions on Knowledge and Data Engineering*, 14(5), 1003–1016. <https://doi.org/10.1109/TKDE.2002.1033770>
- Schubert, E., & Rousseeuw, P. J. (2021). Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. *Information Systems*, 101, 101804. <https://doi.org/10.1016/j.is.2021.101804>
- Setiyono, B., & Mukhlash, I. (2005). Kajian Algoritma GDBScan, Clarans dan Cure untuk Spatial Clustering. *Limits: Journal of Mathematics and Its Applications*, 2(2). <https://doi.org/10.12962/j1829605x.v2i2.1373>
- Wulandari, G. F. (2014). Segmentasi Pelanggan Menggunakan Algoritma K-Means Untuk Customer Relationship Management (CRM) Pada Hijab Miulan. *Industrial Marketing Management, I(segmentasi pelanggan)*, 7.