

Implementasi Metode *Clustering* Sebagai Penunjang Strategi dalam Manajemen Pelanggan

Nurmilayanti¹, Nahrun Hartono²

^{1,2} Sistem Informasi, Sains dan Teknologi, UIN Alauddin Makassar
¹60900118014@uin-alauddin.ac.id, ²nahrun.hartono@uin-alauddin.ac.id

Abstract

Customer management is the most crucial thing because customers are one of the most important assets for a company or business. With the implementation of effective customer management, a company or business is able to achieve customer satisfaction which leads to the creation of loyal customers. To achieve this, businesses must recognize the characteristics of each customer so that the company can adjust the right strategy according to the characteristics of the customer. Using the K-Medoids clustering algorithm and analysis of the RFM (Recency, Frequency, Monetary) value can be done for customer segmentation so that business owners, in this case the Rahmadani Cosmetic Shop, can recognize the characteristics of their customers. RFM value analysis is measured based on the latest transaction value, number of transactions and total money during the transaction. The amount of transaction data used was 1327 entries from 2021 to 2022. Testing the optimum cluster value is done using the silhouette coefficient method which produces an optimum value of 0.496 for 4 clusters. The results of this study are 4 clusters of 1327 data that have been processed through data preprocessing and RFM scoring, obtained data of 178 customers, in cluster 1 is ranked second, cluster 2 is ranked third, cluster 3 is ranked first and cluster 4 is ranked.

Keywords: customer segmentation, k-medoids, clustering, RFM analysis, silhouette coefficient

Abstrak

Manajemen pelanggan merupakan hal yang paling krusial karena pelanggan adalah salah satu aset yang terpenting bagi sebuah perusahaan atau bisnis. Dengan penerapan manajemen pelanggan yang efektif, perusahaan atau bisnis mampu mencapai kepuasan pelanggan yang menyebabkan terciptanya pelanggan yang loyal. Untuk mencapai hal itu maka pelaku usaha harus mengenali karakteristik setiap pelanggan sehingga perusahaan dapat menyesuaikan strategi yang tepat sesuai dengan karakteristik pelanggan tersebut. Dengan menggunakan algoritma *K-Medoids clustering* serta analisis dari nilai RFM (*Recency, Frequency, Monetary*) dapat dilakukan untuk segmentasi pelanggan sehingga pemilik usaha dalam hal ini adalah toko kosmetik Rahmadani dapat mengenali karakteristik pelanggannya. Analisis nilai RFM diukur berdasarkan nilai terbaru transaksi, jumlah transaksi dan total uang selama transaksi. Jumlah data transaksi yang digunakan sebesar 1327 entri dari tahun 2021 sampai 2022. Pengujian optimum nilai *cluster* dilakukan dengan menggunakan metode *silhouette coefficient* yang menghasilkan nilai optimum 0.496 untuk 4 *cluster*. Hasil dari penelitian ini adalah 4 buah *cluster* dari 1327 data yang telah diolah melalui *preprocessing* data serta RFM scoring didapatkan data sebesar 178 pelanggan, pada *cluster* 1 berada pada peringkat kedua, *cluster* 2 peringkat ketiga, *cluster* 3 peringkat pertama dan *cluster* ke 4 berada pada peringkat terakhir.

Kata kunci: segmentasi pelanggan, *k-medoids*, *clustering*, analisis RFM, *silhouette coefficient*

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International

1. Pendahuluan

Pentingnya manajemen pelanggan dalam mencapai kesuksesan bisnis di era yang semakin maju akan teknologi dan persaingan yang semakin ketat. Dengan penerapan manajemen pelanggan yang efektif, pelaku bisnis dapat mencapai kepuasan pelanggan yang menyebabkan terciptanya loyalitas pelanggan [1]. Pencapaian dalam hal kepuasan dan loyalitas pelanggan yang tinggi tentunya akan memberikan dampak positif bagi pelaku usaha, hal ini dikarenakan perolehan keuntungan dan peningkatan pangsa pasar yang dapat di dapatkan. Untuk mencapai hal itu maka pelaku usaha harus mengenali karakteristik setiap pelanggan, sehingga perusahaan dapat menyesuaikan strategi yang tepat sesuai dengan karakteristik pelanggan tersebut. Toko kosmetik Rahmadani merupakan salah satu usaha

yang bergerak dibidang kecantikan baik pada *skin care* maupun *body care*. Usaha yang telah berdiri dari tahun 2018 dan masih mampu menjaga eksistensinya sampai sekarang, bahkan mampu memperluas pangsa pasarnya keluar provinsi. Namun, toko kosmetik Rahmadani dalam hal manajemen pelanggan mereka masih belum mampu menentukan informasi data pelanggan yang memberikan keuntungan yang maksimal. Data pelanggan yang ada hanya dikumpulkan tanpa proses lebih lanjut untuk mendapatkan *insight* yang lebih, akibatnya mereka kurang memahami karakteristik pelanggannya untuk memberikan strategi pemasaran yang sesuai dengan segmen pelanggan. Dalam mengidentifikasi pelanggan dengan karakteristik beragam diperlukan segmentasi pelanggan untuk mengelompokkan pelanggan yang memiliki kesamaan karakteristik. Segmentasi pelanggan bertujuan untuk

mengetahui perilaku konsumen sehingga membantu penerapan strategi pemasaran yang sesuai dan tepat sasaran [2]. Dalam hal segmentasi pelanggan dapat dilakukan dengan penerapan salah satu algoritma *clustering* dalam *data mining*, yang diharapkan dalam proses pengambilan keputusan yang tepat. *Data mining* dapat menggali pengetahuan berharga dari kumpulan data yang disusun berdasarkan pemahaman manusia untuk pengembangan suatu bisnis.

Clustering merupakan teknik *data mining* yang mengelompokkan data ke dalam himpunan-himpunan tertentu, data yang termasuk dalam satu kelompok memiliki kesamaan yang lebih dominan dari pada kesamaannya dengan data-data yang berada di kelompok-kelompok lainnya. *Clustering* juga dapat diartikan sebagai metode pembelajaran mesin yang digunakan untuk mengelompokkan data menjadi beberapa kelompok yang sesuai yang bertujuan untuk menemukan struktur yang terdapat dalam data yang tidak diketahui sebelumnya [3]. Teknik *clustering* merupakan teknik pengelompokan data berdasarkan kemiripannya menjadi beberapa kelompok yang memiliki karakteristik khusus yang membedakan dengan *cluster* yang lainnya [4].

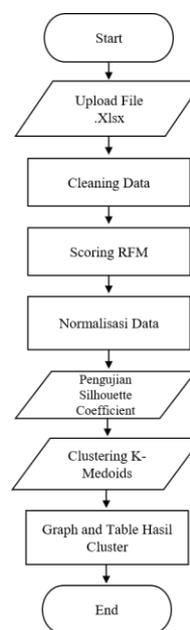
Salah satu metode *clustering* yang sangat populer dan banyak dipelajari untuk meminimalkan kesalahan *clustering* adalah *k-medoids* yang juga dikenal sebagai *Partitioning Around Medoids*. Algoritma *k-medoids* membentuk suatu *cluster* dengan cara menghitung jarak kemiripan yang dimiliki antar medoid dengan objek non medoids [5]. *K-Medoids* melakukan pembagian data dengan cara mengurangi jumlah perbedaan antara setiap objek spesifik dan objek yang mewakili kelompok terdekat, melalui total kesalahan yang bersifat mutlak [6]. Menurut Firda dalam riset yang dilakukan oleh Sihombing menjelaskan kelebihan *K-Medoids* yang menjadi landasan penggunaan algoritma *K-Medoids* pada penelitian ini, yaitu algoritma *K-Medoids* mempunyai keunggulan untuk menangani kelemahan *K-Means* berupa *outlier*, selain itu algoritma *K-Medoids* lebih efektif dalam mengelompokkan data yang memiliki skala relatif kecil [7], hal ini sejalan dengan skala data pada penelitian ini yang termasuk dalam *small data* atau data dengan skala yang kecil. Dalam penelitian ini, untuk melakukan segmentasi pelanggan berdasarkan profil dari masing-masing pelanggan dan selanjutnya menganalisis hasil segmentasinya untuk diketahui pelanggan yang *profitable* maupun sebaliknya. Untuk melakukan *profiling* pada pelanggan maka diperlukan model terhadap aktivitas pelanggan. Model yang kerap dipakai dalam pengelompokan pelanggan adalah model RFM (*Recency, Frequency, Monetary*), yang mengelompokkan pelanggan berdasarkan interval waktu transaksi terakhir, frekuensi transaksi dan besaran nilai yang dikeluarkan sebagai *royalty* perusahaan [8]. Perangkat lunak yang digunakan pada penelitian ini berdasarkan R studio.

Penelitian ini bertujuan untuk menerapkan metode

clustering dengan algoritma *K-Medoids* untuk memperoleh segmentasi pelanggan pada toko kosmetik Rahmadani. Segmentasi adalah proses membagi pelanggan menjadi beberapa *cluster* dengan kategori loyalitas pelanggan untuk membangun strategi pemasaran [9]. Segmentasi pelanggan yang dihasilkan diharapkan dapat memberikan wawasan yang berharga bagi pelaku bisnis dalam merancang strategi pengolahan manajemen pelanggan.

2. Metode Penelitian

Metode penelitian adalah protokol yang diikuti sebagai pedoman saat melakukan penelitian agar dapat memberikan justifikasi ilmiah terhadap temuan-temuan. Metode penelitian yang di implementasikan dalam penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Flow Chart RFM Analyst

2.1 Preprocessing Data

Tabel 1. Data Mentah

tgl	no.hp	tot_harga
1/1/2021	897*****6	Rp. 925000
...
9/6/2022	853*****0	Rp. 4715000
9/6/2022	812*****9	Rp.9465000
9/7/2022	852*****1	Rp. 10500000
9/7/2022	813*****9	Rp. 4540000
...
12/31/2022	823*****6	Rp. 1720000

Tabel 1 yang berisikan data mentah pada Toko Kosmetik Rahmadani sebesar 1327 entri data *customer* yang berupa agen dan *reseller* pada Toko Kosmetik Rahmadani. Tahapan *preprocessing* data pada penelitian ini terdiri dari pemilihan atribut, menyesuaikan tipe data dan mengatasi *missing value*, Pemilihan atribut yang relevan pada data berupa tanggal, no. hp dan total harga didasarkan karena

pemilihan atribut tanggal memungkinkan untuk mengukur seberapa baru pelanggan melakukan pembelian, semakin baru transaksi maka semakin besar kemungkinan pelanggan masih aktif dan berpotensi untuk melakukan pembelian lebih lanjut, selain itu tanggal transaksi juga memungkinkan untuk mengetahui berapa kali pelanggan melakukan transaksi. Sedangkan atribut no. hp akan dijadikan id bagi setiap pelanggan yang digunakan sebagai pembeda antar setiap pelanggan, untuk atribut total harga akan digunakan sebagai atribut yang dapat mencerminkan berapa jumlah uang yang telah dikeluarkan pelanggan selama transaksi.

Tahapan selanjutnya adalah mengubah tipe data pada setiap atribut untuk menyesuaikan dengan kebutuhan *clustering*. no.hp akan dijadikan sebagai id pelanggan yang bersifat unik, atribut tgal akan diformat menjadi Y/m/d (*years, month, day*) dan tot_harga akan diubah menjadi tipe data numerik. Tahap selanjutnya yaitu mengatasi *missing value* atau nilai yang hilang. Cara untuk mengatasi *missing value* bisa dikatakan sebagai *imputation* atau nilai pengganti, salah satu caranya dapat diisi dengan nilai *mean, median* ataupun *mode* [10]. Pada penelitian ini ada beberapa baris data yang *missing* pada kolom id dan tot_harga, untuk kolom id yang memiliki *missing value* akan diisi dengan nilai pada baris sebelumnya yang ditambah dengan 1, sementara untuk mengatasi *missing value* pada kolom tot_harga dapat dilakukan dengan mengisi median dari data tot_harga karena data pada kolom tot_harga memiliki distribusi yang tidak simetris, data yang tidak simetris dapat dilihat pada tabel 1 kolom tot_harga yang mena selisih nilai dari setiap data yang jauh, sebaliknya jika data memiliki distribusi yang simetris maka untuk mengatasi *missing value* lebih baik dengan menggunakan metode mean.

2.2. RFM Scoring

2.2.1 Nilai Recency (R)

Nilai *recency* yaitu mengukur nilai pelanggan berdasarkan pembelian paling akhir yang dilakukan pelanggan. Semakin kecil nilai R, maka semakin baru interaksi pelanggan yang berarti bahwa pelanggan tersebut semakin loyal [11]. *Recency* merupakan salah satu dimensi yang digunakan untuk mengategorikan pelanggan berdasarkan perilaku pembelian mereka. Pada penelitian ini data transaksi yang digunakan adalah data transaksi agen dan *reseller* dalam rentang waktu Januari 2021 hingga Desember 2022. Berikut adalah rumus untuk mencari nilai *recency*:

$$R = \text{tanggal terakhir transaksi keseluruhan} - \text{tanggal terakhir pelanggan transaksi} \quad (1)$$

Tanggal terakhir keseluruhan pelanggan transaksi pada penelitian ini adalah 31-12-2022, sedangkan pada tanggal transaksi terakhir pelanggan masing-masing melakukan transaksi terdiri dalam rentang waktu antara 2021 januari hingga Desember 2022. Berikut adalah

kode untuk mencari nilai *recency* pada R studio dengan menggunakan fungsi MAX seperti berikut:

Program Jurnal

```
rfm_recency <- df %>%  
  group_by (id) %>%  
  summarise(recency= as.numeric (as.Date  
(tgl)-max(tgal)))
```

Perintah `rfm_recency <- df %>%` dalam R studio adalah untuk membuat variabel `rfm_recency` sebagai objek data frame baru menggunakan data dari data frame `df`, perintah `group_by(id) %>%` berfungsi untuk mengelompokkan data berdasarkan id menggunakan fungsi `group_by(id)`. Perintah `summarise(recency = as.Date(tgl)-max(tgal))`, menghitung *recency* menggunakan fungsi `summarise()`, *recency* berfungsi sebagai kolom baru sementara perintah `as.numeric(as.Date(tgl) - (tgal))` digunakan untuk menyimpan hasil pengurangan `tgl` dengan `tgal` yang mana hasil pengurangan tersebut akan menjadi tipe `difftime` yang merupakan perbedaan dalam satuan waktu yang kemudian selanjutnya akan diubah menjadi tipe data numerik.

2.2.2 Nilai Frequency (F)

Frequency merupakan tingkat keseringan pelanggan transaksi [12]. Berikut adalah kode untuk mendapatkan nilai *frequency* pada R studio menggunakan fungsi *summarise*.

Program Jurnal

```
rfm_frequency <- df %>%  
  group_by(id) %>%  
  summarise (frequency = n())
```

Pada kode di atas `df` merupakan dataframe yang berisi data yang akan dianalisis, sementara `group_by(id)` berfungsi untuk mengelompokkan data berdasarkan kolom id sehingga kode akan dijalankan pada setiap kelompok id berdasarkan nilai id. Untuk kode `summarise(frequency=n())` adalah untuk melakukan ringkasan kelompok `i` atau dapat dikatakan sebagai langkah agregasi seperti menghitung jumlah, rata-rata, nilai max dan min, dan sebagainya, untuk menghitung frekuensi kemunculan setiap kelompok. Hasilnya akan disimpan dalam kolom baru dengan nama *frequency*.

2.2.3 Nilai Monetary (M)

Monetary adalah jumlah total nilai pembelian yang dilakukan pelanggan dalam periode tertentu [13]. Berikut adalah code untuk menghitung *monetary* dalam R studio dengan menggunakan fungsi SUM.

Program Jurnal

```
rfm_monetary <- df %>%  
  group_by(id) %>%  
  summarise (Monetary = sum(tot_harga))
```

Fungsi `df` pada kode di atas merupakan dataframe yang berisi data yang ingin dianalisis, sementara `group_by(id)` berfungsi untuk mengelompokkan data berdasarkan kolom id. Sedangkan untuk potongan kode `summarise(Monetary=sum(tot_harga))` berfungsi untuk melakukan ringkasan terhadap data pada setiap

kolom id, untuk menghitung jumlah total dari kolom tot_harga menggunakan fungsi sum() hasilnya akan disimpan pada kolom baru dengan nama Monetary. Tabel 2 di bawah adalah potongan hasil perhitungan dari recency, frequency dan monetary yang dihitung menggunakan rumus RFM, data yang digunakan terdapat pada tabel 1:

Tabel 2. Atribut dan Tipe Data Mentah

R	F	M
315	12	21775000
...
23	17	95825000
58	16	108770000
57	9	62375000
24	7	30135000
.....
0	3	9300000

2.3 Normalisasi Data

Normalisasi merupakan proses untuk membuat data yang memiliki rentan nilai tinggi menjadi setara dari ketiga atribut RFM, sehingga lebih mudah dalam pengolahan [14]. Untuk dataset dengan variasi nilai-nilai yang berbeda di antara atribut yang ada maka sebelum dihitung jaraknya terlebih dahulu lakukan normalisasi data agar data seimbang [15]. Pada penelitian ini, menggunakan metode normalisasi Z-score.

$$Z = (x - \mu) / \sigma \tag{2}$$

Keterangan:

x = nilai yang diamati (skor mentah)

μ = rata-rata populasi

σ = standar deviasi populasi

Z = Z-score (Nilai Baku)

Berikut adalah tabel data yang berisi nilai rata-rata populasi dan standar deviasi pada keseluruhan atribut RFM sebagai berikut:

Tabel 3. Atribut dan Tipe Data Mentah

Perhitungan	R	F	M
Rata-Rata	214.8258	7.449438	32159682.6
Standar Deviasi	228.1629	7.470486	50881330.69

Tabel 3 yang ada di atas adalah hasil perhitungan nilai rata-rata populasi dan standar deviasi menggunakan fungsi yang terdapat dalam excel. Berikut adalah sample data ke 1 untuk perhitungan Z-score:

Normalisasi data ke 1 atribut Recency

$$Z.R = (315 - 214.8258) / 228.1629 = -0.439047$$

Normalisasi data ke 1 atribut Frequency

$$Z.F = (12 - 7.449438) / 7.470486 = 0.609139$$

Normalisasi data ke 1 atribut Monetary

$$Z.M = (21775000 - 32159682.6) / 50881330.69 = -0.20603$$

Hasil selengkapnya dari normalisasi Z-score RFM dapat dilihat dari tabel 4:

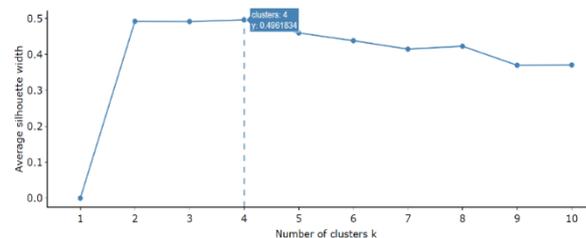
Tabel 4. Hasil Normalisasi Data ZScore

R	F	M
0.439047	0.609139	-0.20603
.....
-84074	1.278439	1.247378
-0.68734	1.144579	1.566783
-0.69172	0.207558	0.590844
-0.83636	-0.06016	-0.04194
.....
-0.94115	-0.5956	-0.45088

3. Hasil dan Pembahasan

3.1 Pengujian Optimal Cluster

Setelah proses normalisasi data RFM maka untuk menentukan jumlah optimum dari cluster dapat dilakukan menggunakan metode Silhouette Coefficient. Silhouette Coefficient adalah metode penilaian cluster yang bertujuan untuk memandang mutu objek dalam cluster [16]. Kinerja Silhouette Coefficient yang merupakan metrik evaluasi untuk mengukur sejauh mana titik-titik dalam cluster sesuai dengan cluster tersebut dan terpisah dari cluster lain. Nilai positif mendekati 1 menunjukkan bahwa titik dalam satu cluster sesuai dengan baik dan terpisah dengan baik dari cluster lainnya. Adapun plot Silhouette Coefficient dari hasil clustering yang telah dilakukan adalah pada gambar 2:



Gambar 2. Plot Silhouette Coefficient

Grafik pada gambar 2 menunjukkan nilai silhouette tertinggi berada pada angka 4 dengan nilai sebesar 0.4961 yang berarti bahwa jumlah cluster yang optimal berdasarkan hasil pengujian silhouette coefficient adalah sebanyak 4 cluster. Sehingga dalam penelitian ini akan dibentuk sebanyak 4 cluster.

3.2 Customer Segmentation Menggunakan K-Medoids

3.2.1 Inisialisasi Pusat Cluster Sebanyak K (Jumlah Cluster)

Berdasarkan visualisasi di atas maka diperoleh hasil bahwa jumlah cluster yang optimal adalah empat dengan nilai silhouette yang paling tinggi jika dibandingkan dengan nilai silhouette pada jumlah cluster sebanyak dua, tiga dan lima. Plot di atas menunjukkan bahwa dua cluster memiliki nilai silhouette 0.4922, tiga cluster memiliki nilai silhouette

0.4916 dan untuk empat *cluster* memiliki nilai *silhouette* sebesar 0.4966 sedangkan nilai *silhouette* untuk lima *cluster* adalah 0.459. Hasil *cluster* optimal ini selanjutnya akan digunakan untuk melakukan *cluster* pada *K-Medoids*.

3.3 Customer Segmentation Menggunakan *K-Medoids*

3.3.1 Inisiasi Pusat *Cluster* Sebanyak *K* (Jumlah *Cluster*)

Dalam penelitian ini akan dibentuk 4 *cluster* berdasarkan perhitungan *cluster optimum* menggunakan *silhouette coefficient* yang telah dilakukan pada R studio. Berikut adalah Tabel 5 yang berisi *medoids* awal yang akan digunakan pada tahap iterasi pertama.

Tabel 5. *Medoids* pada *Cluster* Iterasi Pertama

<i>Cluster</i>	R	F	M
1	-0.569005023	0.74299874	0.07616384
2	-0.547090857	-0.72946233	-0.53299869
3	-0.932780185	1.680019415	1.809609068
4	1.438332616	-0.863322421	-0.576000929

3.3.2 Hitung Setiap Objek ke *Cluster* Terdekat (Persamaan *Euclidean Distance*)

Setelah menentukan *medoids* secara random, maka selanjutnya adalah menghitung jarak terdekat (*cost*) dengan persamaan *Euclidean distance*. Contoh perhitungan jarak terdekat pada data:

Data Pertama ke *Cluster* 1:

$$C1 = \sqrt{(-0.569005023 - 0.439047)^2 + (0.742998739 - 0.609139)^2 + (0.073860027 - -0.20603)^2} = 1.0542$$

Data Pertama ke *Cluster* 2:

$$C2 = \sqrt{(-0.547090857 - 0.439047)^2 + (-0.729462325 - 0.609139)^2 + (-0.53299869 - -0.20603)^2} = 1.69513$$

Data Pertama ke *Cluster* 3:

$$C3 = \sqrt{(-0.932780185 - 0.439047)^2 + (1.680019415 - 0.609139)^2 + (1.809609068 - -0.20603)^2} = 2.6704$$

Data Pertama ke *Cluster* 4:

$$C4 = \sqrt{(1.438332616 - 0.439047)^2 + (-0.863322421 - 0.609139)^2 + (-0.576000929 - -0.20603)^2} = 1.81797$$

Setelah dilakukan perhitungan pada semua data, maka selanjutnya adalah menentukan *cluster* mana dari setiap data. Pada tabel 6 menampilkan hasil

perhitungan dari iterasi 1. Adapun iterasi yang dilakukan adalah 2 iterasi:

Tabel 6. *Medoids* pada *Cluster* Iterasi Pertama

<i>Cluster</i>	<i>Cost</i> 1	<i>Cost</i> 2	<i>Cost</i> 3	<i>Cost</i> 4
1	1.0542	1.69513	2.6704	1.81797
...
3	1.32166	2.7031	0.70336	3.62218
3	1.49176	2.77379	0.66826	3.58936
1	0.7564	1.47379	1.93209	2.65565
1	0.85409	0.88111	2.54976	2.47119
.....
2	1.48469	0.42518	3.21468	2.39824

Tabel 6 merupakan tabel penentuan *cluster*, dalam menentukan *cluster* sebuah data berada di *cluster* tertentu maka harus dilakukan perbandingan nilai minimum diantara keseluruhan *cost*, misalnya pada data pertama *cost* 1 nilainya 1.0542, *cost* 2 nilainya 1.69513, *cost* 3 nilainya 2.6704, dan *cost* 4 nilainya 1.81797. Sehingga nilai minimum (terkecil) dari *cost* 1 hingga *cost* 4 adalah *cost* 1 dengan nilai 1.0542 sehingga *clusternya* berada di *cluster* 1 yang memiliki nilai terkecil dibandingkan nilai *cost* lainnya. Proses ini berlaku untuk setiap data dalam tabel *cluster*.

3.3.3 Inisialisasi Pusat *Cluster* Baru

Setelah mendapatkan hasil dari iterasi 1 maka tahap selanjutnya dilakukan iterasi yang ke 2. Tentukan kembali *medoids* baru non *medoids*) yang dapat dilihat pada tabel 7.

Tabel 7. *Medoids* pada *Cluster* Iterasi Kedua

<i>Cluster</i>	R	F	M
1	-0.56462219	0.595602228	-0.49054697
2	-0.555856523	-0.863322421	-0.570143945
3	-0.840740686	1.947739609	1.12202878
4	2.025632275	-0.863322421	-0.386383037

3.3.4 Hitung Jarak Setiap Objek yang Berada pada Masing-masing *Cluster* dengan Kandidat Non *Medoids*.

Hitung jarak setiap data ke masing-masing non *medoids* menggunakan rumus *Euclidean distance* sehingga diperoleh hasil perhitungan jarak masing-masing *cluster*. Berikut merupakan pada data pertama.

Data Pertama ke *Cluster* 1:

$$C1 = \sqrt{(-0.56462219 - 0.439047)^2 + (-0.595602228 - 0.609139)^2 + (-0.49054697 - -0.20603)^2} = 1.59399$$

Data Pertama ke *Cluster* 2:

$$C2 = \sqrt{(-0.555856523 - 0.439047)^2 + (-0.863322421 - 0.609139)^2 + (-0.570143945 - -0.20603)^2} = 1.81438$$

Data Pertama ke *Cluster* 3:

$$C3 = \sqrt{(-0.840740686 - 0.439047)^2 + (1.947739609 - 0.609139)^2 + (1.12202878 - -0.20603)^2} = 2.27779$$

Data Pertama ke Cluster 4:

$$C4 = \sqrt{(2.025632275 - 0.439047)^2 + (-0.863322421 - 0.609139)^2 + (-0.386383027 - -0.20603)^2} = 2.17224$$

Setelah dilakukan perhitungan pada setiap data, maka selanjutnya menentukan *cluster*, penentuan *cluster* pada tabel ini sama dengan melakukan penentuan *cluster* pada tabel 6 yaitu membandingkan nilai minimum dari keseluruhan *cost*. Tabel 8 merupakan penentuan *cluster* iterasi ke 2.

Tabel 8. Medoids pada Cluster Iterasi Kedua

Cluster	Cost 1	Cost 2	Cost 3	Cost 4
1	1.59399	1.81438	2.27779	2.17224
...
3	2.57335	2.82591	0.68166	3.93511
3	2.65107	2.89101	0.9032	3.86933
1	1.3554	1.58748	1.82467	3.08085
1	0.75081	1.00251	2.31981	2.99269
.....
2	0.48469	0.42518	3.21468	2.39824

3.3.5 Hitung Total Simpangan (S)

Untuk menghitung total simpangan dapat dilakukan dengan menghitung total *cost* baru-total *cost* lama.

$$S = \text{total cost baru} - \text{total cost lama} \quad (3)$$

$$= 1519.68 - 1503.04$$

$$= 16.635$$

Perhitungan di atas didapatkan dari menjumlahkan keseluruhan nilai *cost* 1, 2, 3 dan 4. Total *cost* pada terasi pertama sebagai total *cost* lama dan total *cost* iterasi ke 2 dianggap sebagai total *cost* baru. Dikarenakan total simpangan > 0, maka perhitungan dihentikan sehingga didapatkan hasil *cluster* akhir sebagai berikut pada tabel 9:

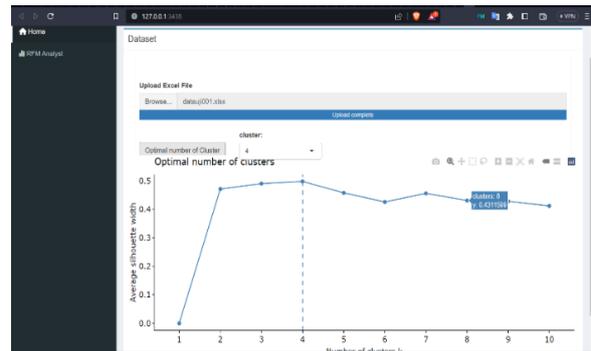
Tabel 9. Hasil Cluster

Cluster	Cost 1	Cost 2	Cost 3	Cost 4
1	1.0542	1.69513	2.6704	1.81797
...
3	1.32166	2.7031	0.70336	3.62218
3	1.49176	2.77379	0.66826	3.58936
1	0.7564	1.47379	1.93209	2.65565
1	0.85409	0.88111	2.54976	2.47119
.....
2	1.48469	0.42518	3.21468	2.39824

Hasil *cluster* akhir dengan nilai k=4 yaitu *cluster* 1 sebanyak 42 pelanggan, *cluster* 2 sebanyak 61 pelanggan, *cluster* 3 sebanyak 26 pelanggan, dan *cluster* 4 sebanyak 49 pelanggan.

3.4 Tampilan Rancangan Model

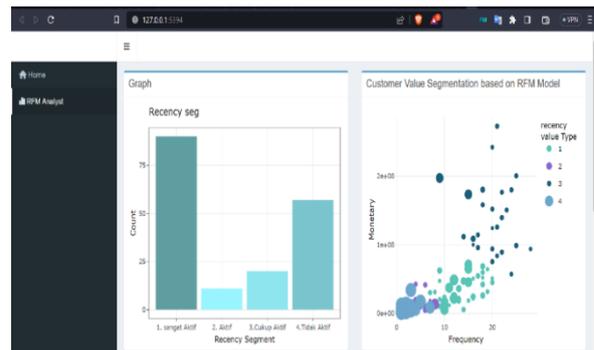
3.4.1 Tampilan Rancangan home page



Gambar 3. Dashboard Analisis home page

Gambar 3 merupakan *dashboard* pada halaman *home*. Tampilan terdiri *upload file* tipe *xlsx* kemudian terdiri dari inputan untuk melihat *optimum number of cluster* pada dataset, user juga dapat memilih jumlah *cluster* sesuai dengan hasil pengujian *optimum number of cluster* untuk menentukan segmentasi pelanggan.

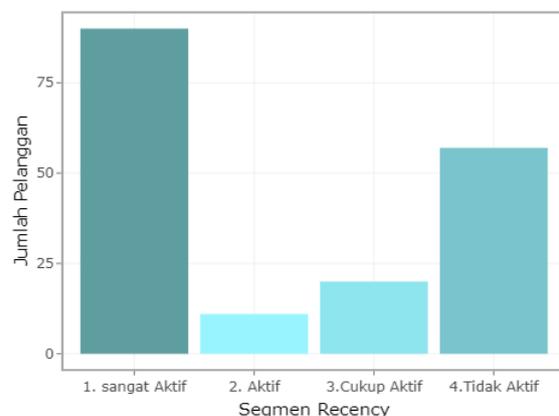
3.4.2 Tampilan Rancangan RFM Analyst



Gambar 4. Dashboard Analisis Halaman RFM Analyst

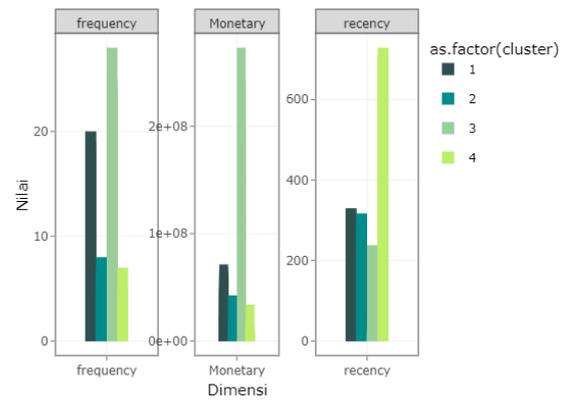
Gambar 4 merupakan tampilan *dashboard* pada halaman *RFM analyst*, yang berisi 3 diagram, diantaranya diagram batang *recency segment*, *scatter plot customer segmentation* dan diagram batang distribusi atribut RFM.

Jumlah Pelanggan berdasar Segmen Recency



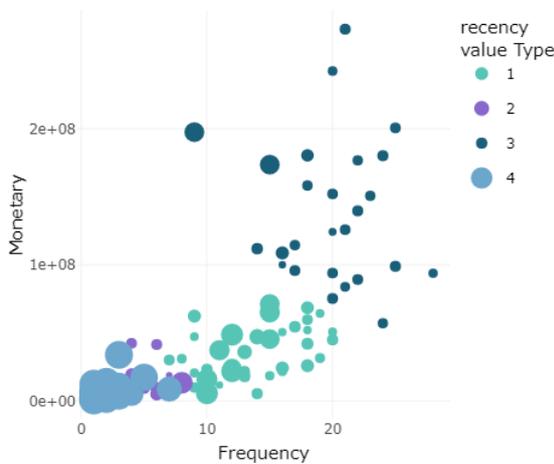
Gambar 5. Diagram Batang Recency Segment

Diagram batang *recency segment* yang menggambarkan pelanggan aktif pada 4 kuartal, batang 1 berada pada kuartal 1 yaitu dalam rentang 3 bulan terakhir pelanggan yang melakukan transaksi dan termasuk kategori pelanggan yang sangat aktif terdiri dari 90 pelanggan, untuk batang ke 2 menjelaskan bahwa pelanggan yang berada pada kuartal ke 2 atau 6 bulan terakhir yang melakukan transaksi, kategori pelanggan pada batang ke 2 adalah aktif dengan jumlah pelanggan sebanyak 11 orang, batang ke 3 berada pada rentang waktu 9 bulan terakhir pelanggan yang melakukan transaksi dan termasuk kategori cukup aktif dengan jumlah pelanggan pada kategori ini adalah 20 pelanggan. Untuk batang ke 4 dalam rentang waktu 12 bulan terakhir pelanggan yang melakukan transaksi sehingga kategori pelanggan pada batang ke 4 adalah tidak aktif dengan jumlah 57 pelanggan.



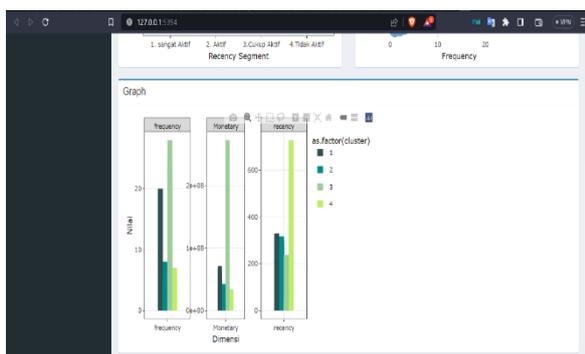
Gambar 8. Diagram Batang Distribusi Atribut RFM

Diagram Batang pada gambar 8 memberikan gambaran visual yang memungkinkan perbandingan nilai antara *cluster* dalam setiap atribut. Pada grafik batang di atas nilai *recency* yang paling rendah menandakan bahwa pelanggan yang baru melakukan transaksi, *cluster* 3 merupakan pelanggan yang memiliki nilai *recency* terendah kemudian disusul oleh pelanggan pada *cluster* ke 2, 1 dan 4. Untuk grafik batang bagian *frequency* nilai yang tertinggi menggambarkan bahwa pelanggan sering melakukan transaksi, *cluster* ke 3 menunjukkan bahwa nilai *frequency* yang paling tinggi atau pelanggan yang sangat sering melakukan transaksi, disusul pelanggan yang berada pada *cluster* 1, 2 dan 4. Pada grafik batang *monetary*, nilai tertinggi menunjukkan bahwa pelanggan melakukan transaksi dalam jumlah yang sangat besar. Nilai *monetary* yang paling tinggi pada grafik *monetary* adalah pada *cluster* ke 3 kemudian disusul 1, 2 dan 4, hal ini menunjukkan bahwa pelanggan yang berada pada *cluster* ke 3 merupakan pelanggan yang melakukan transaksi dalam jumlah yang sangat besar jika dibandingkan dengan *cluster* lain. Sehingga untuk peringkat 1 yaitu pelanggan yang *high loyalty* adalah *cluster* ke 3 yang berjumlah 26 pelanggan, peringkat ke 2 yaitu pelanggan yang dapat dikategorikan *medium loyalty* adalah pelanggan yang berada pada *cluster* ke 1 dengan jumlah 42 pelanggan. Peringkat ke 3 yaitu pelanggan yang dapat dikategorikan sebagai *low loyalty* adalah pelanggan yang berada pada *cluster* ke 2 dengan jumlah 61 pelanggan. Selanjutnya adalah peringkat ke 4 yaitu pelanggan yang dapat dikategorikan sebagai *regular* adalah pelanggan yang berada pada *cluster* ke 4 dengan jumlah 49 pelanggan.



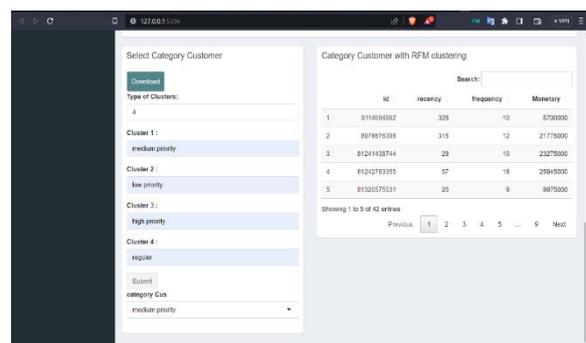
Gambar 6. Scatter Plot Customer Segmentation

Gambar 6 merupakan diagram *scatter plot* yaitu persebaran data atribut RFM, semakin kecil plotnya maka semakin baru interaksi pelanggan. Dapat dilihat bahwa pada *scatter plot* di atas nilai RFM terbaik berada pada *cluster* ke 3.



Gambar 7. Dashboard Analisis Halaman RFM Analyst

Gambar 7 di atas merupakan Tampilan *dashboard* yang terdiri dari grafik batang yang menggambarkan distribusi nilai dari 3 atribut RFM dalam *cluster*. Grafik batang dibagi menjadi beberapa *group* berdasarkan *cluster* yang ditampilkan dengan warna yang berbeda pada batang-batang tersebut.



Gambar 9. Dashboard Analisis Halaman RFM Analyst data table

Gambar 9 merupakan tampilan *dashboard* data tabel hasil segmentasi pelanggan yang dapat ditampilkan berdasarkan kategori pelanggan. Tabel hasil dari segmentasi pelanggan tersebut dapat di download dengan format file excel.

Gambar 10. *Dashboard* Halaman *RFM Analyst Category Customer*

Gambar 10 berupa tampilan untuk melakukan input peringkat *cluster* sesuai dengan hasil analisis plot sebelumnya, setelah melakukan submit maka hasil peringkat tersebut akan ditampilkan pada *data table* sebagai berikut:

Category Customer with RFM clustering

id	recency	frequency	Monetary
81341737595	27	20	44950000
85298738428	7	20	50765000
82190304649	15	19	31560000
85290950680	6	19	64340000
81242783355	57	18	25945000

Showing 1 to 5 of 42 entries

Previous 1 2 3 4 5 ... 9 Next

Gambar 11. *Dashboard* Halaman *RFM Analyst Data Table*

Gambar 11 berupa *data table* hasil *clustering* yang ditampilkan sesuai dengan peringkat yang di *select* pada gambar 10.

4. Kesimpulan

Dari pengujian optimum *number of cluster* terbentuk *cluster* optimal yaitu 4 buah *cluster* dengan nilai *silhouette* sebesar 0.4961. *Cluster* 1 terdiri dari 42 pelanggan, *cluster* 2 terdiri dari 61 pelanggan, *cluster* 3 terdiri dari 26 pelanggan dan *cluster* 4 terdiri dari 49 pelanggan. Berdasarkan analisis pada grafik *scatter plot* dan diagram batang yang ditampilkan maka peringkat pertama pada *cluster* ke 3 dengan nilai *recency* terkecil dan nilai *frequency* serta *monetary* terbesar sehingga tipe pelanggan berupa *high priority*, kemudian peringkat ke 2 berada pada *cluster* 1 dengan tipe pelanggan *middle priority*, peringkat ke 3 berada pada *cluster* 2 dengan tipe pelanggan *low priority* dan peringkat ke 4 pada *cluster* 4 dengan tipe pelanggan *regular*. Adapun grafik yang ditampilkan oleh model

mampu memberikan kemudahan oleh pihak Toko Kosmetik Rahmadani untuk mengetahui segmen pelanggan berdasarkan *cluster* sehingga memberikan kemudahan dalam memberikan pelayanan dan diskon sesuai dengan segmen pelanggan pada Toko Kosmetik Rahmadani.

Daftar Rujukan

- [1] N. Mirantika, T. S. Syamfithriani, and R. Trisudarmo, "Implementasi Algoritma K-Medoids Clustering Untuk Menentukan Segmentasi Pelanggan," vol. 17, pp. 196–204, 2023.
- [2] B. E. Adiana, I. Soesanti, and A. E. Permasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model Dan Teknik Clustering," *J. Terap. Teknol. Inf.*, vol. 2, no. 1, pp. 23–32, 2018, doi: 10.21460/jutei.2018.21.76.
- [3] M. R. Nahjan, N. Heryana, and A. Voutama, "Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko OJ Cell," vol. 7, no. 1, pp. 101–104, 2023.
- [4] B. N. Sari and A. Primajaya, "Penerapan Clustering Dbscan Untuk Pertanian Padi Di Kabupaten Karawang," *J. Inform. dan Komput.*, vol. 4, no. 1, pp. 28–34, 2019, [Online]. Available: www.mapcoordinates.net/en.
- [5] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
- [6] A. Wibowo and A. R. Handoko, "Segmentasi Pelanggan Ritel Produk Farmasi Obat Menggunakan Metode Data Mining Klasterisasi Dengan Analisis Recency Frequency Monetary (RFM) Termodifikasi," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, p. 573, 2020, doi: 10.25126/jtiik.2020702925.
- [7] R. E. Sihombing, D. Rachmatin, and J. A. Dahlan, "Program Aplikasi Bahasa R untuk Pengelompokan Objek Menggunakan Metode K-Medoids," *J. EurekaMatika*, vol. 7, no. 1, pp. 58–79, 2020.
- [8] A. D. Savitri, F. A. Bachtari, and N. Y. Setiawan, "Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 9, pp. 2957–2966, 2018.
- [9] N. W. Wardani, *Penerapan Data Mining dalam Analytic CRM*. 2020.
- [10] DQLab, "Teknik Pengolahan Data : Mengetahui Missing Values dan Cara-Cara Menanganinya," *New Media Tower*. p. 1, 2020, [Online]. Available: <https://www.dqlab.id/digital-transformation-pahami-teknik-pengolahan-ini-dalam-industri-data>.
- [11] R. Y. Firmansah, J. Dedy Irawan, and N. Vendyansyah, "Analisis Rfm (Recency, Frequency and Monetary) Produk Menggunakan Metode K-Means," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 5, no. 1, pp. 334–341, 2021, doi: 10.36040/jati.v5i1.3282.
- [12] F. Hadi, D. Octari Rahmadia, F. Hadi Nugraha, N. Putri Bulan, Mustakin, and S. Monalisa, "Penerapan K-Means Clustering Berdasarkan RFM Mofek Sebagai Pemetaan dan Pendukung Strategi Pengelolaan Pelanggan (Studi Kasus: PT. Herbal Penawar Alwahidah Indonesia Pekanbaru)," *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 15, no. 1, pp. 69–76, 2017, [Online]. Available: <http://ejournal.uin-suska.ac.id/index.php/sitekin/article/view/4575>.
- [13] A. A. D. Sulistyawati and M. Sadikin, "Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan," *Sistemasi*, vol. 10, no. 3, p. 516, 2021, doi: 10.32520/stmsi.v10i3.1332.
- [14] A. S. Hartati, "Penerapan Algoritma K-Means Dalam Pemetaan Pelanggan Potensial Menggunakan Model Recency Frequency Monetary," *J. Ekon. Vol. 18, Nomor 1 Maret 201*, vol. 2, no. 1, pp. 41–49, 2020.
- [15] S. Bahri, D. Marisa Midyanti, and P. Korespondensi,

- “Penerapan Metode K-Medoids Untuk Pengelompokan Mahasiswa Berpotensi Drop Out Application of K-Medoids Method for Dropout Potential Student Grouping,” vol. 10, no. 1, pp. 165–172, 2023, doi: 10.25126/jtiik.2023106643.
- [16] B. A. D. Daffa Rafif Agustian, “Analisis clustering demam berdarah dengue,” *JIKO (jurnal Inform. dan Komput.*, vol.

Vol. 6, no. no. e-ISSN : 2477-3964 — p-ISSN : 2477-4413, pp. 18–21, 2022, [Online]. Available: [ps://scholar.archive.org/work/q732ndje4vht7j4bwmoryyzuba/access/wayback/https://ejournal.akakom.ac.id/index.php/jiko/article/download/504/pdf](https://scholar.archive.org/work/q732ndje4vht7j4bwmoryyzuba/access/wayback/https://ejournal.akakom.ac.id/index.php/jiko/article/download/504/pdf).