

Aplikasi Analisis Segmentasi Pelanggan untuk Menentukan Strategi Pemasaran Menggunakan Kombinasi Metode *k-Means* dan Model RFM

Ahmad Febri, Norma Ningsih*, Julianto Lemantara
Sistem Informasi, Fakultas Teknologi Informatika, Universitas Dinamika
Jl. Raya Kedung Baruk No.98, Kec. Rungkut, Surabaya
*e-mail : norma@dinamika.ac.id

(*received*: 15 November 2020, *revised*: 6 Desember 2020, *accepted*: 29 Desember 2020)

Abstrak

Selain harus bersaing dengan dealer lain, kenjeran Auto2000 juga harus bersaing dengan cabang Auto2000 lainnya. Salah satu strategi yang digunakan untuk memenangkan persaingan adalah melalui penerapan segmentasi pelanggan. Saat ini kenjeran Auto2000 belum memiliki mekanisme segmentasi pelanggan, permasalahan yang dihadapi adalah tidak dapat mengklasifikasikan pelanggan karena jumlah datanya yang besar dan belum ada alat yang mendukung segmentasi. Berdasarkan permasalahan tersebut, maka diusulkan solusi aplikasi segmentasi pelanggan yang dapat membantu kenjeran Auto2000 melakukan segmentasi. Untuk mengetahui karakteristik pelanggan digunakan model RFM. Sedangkan metode K-Means digunakan untuk melakukan segmentasi, namun metode ini memiliki kelemahan yaitu sulit untuk menentukan jumlah cluster yang terbaik. Untuk mengatasi masalah ini, metode Koefisien Silhouette digunakan. Metode ini digunakan untuk membantu metode K-Means dalam menentukan jumlah cluster terbaik yang akan digunakan. Berdasarkan hasil pengujian, aplikasi telah berjalan sesuai fungsinya yaitu mampu mengelompokkan pelanggan menjadi beberapa kelompok dan menemukan kelompok pelanggan potensial.

Kata Kunci: segmentasi pelanggan, *k-means clustering*, RFM, *silhouette coefficient*

Abstract

Besides having to compete with other dealers, Auto2000 kenjeran also had to compete with other Auto2000 branches. One of the strategies used to win the competition is through the implementation customer segmentation. Currently Auto2000 kenjeran does not have a customer segmentation mechanism, the problem faced is that it cannot classify customers because of the large amount of data and there are no tools that support segmentation. Based on these problems, a customer segmentation application solution is proposed that can help Auto2000 kenjeran perform segmentation. To determine the characteristics of the customer, the RFM model is used. While the K-Means method is used to perform segmentation, but this method has a weakness, that's difficult to determine the best number of cluster. To resolve this problem, the Silhouette Coefficient method is used. This method is used to assist the K-Means method in determining the best number of cluster that will be used. Based on the test results, the application has been running according to its function, which is able to group customers into several groups and to find groups of potential customers.

Keywords: customer segmentation, *k-means clustering*, RFM, *silhouette coefficient*

1. Pendahuluan

Auto2000 adalah jaringan jasa perawatan, perbaikan, penjualan dan penyediaan suku cadang mobil Toyota yang sudah berdiri sejak tahun 1975. Sampai saat ini Auto2000 memiliki 124 outlet yang telah diotorisasi, dalam penelitian ini hanya Auto2000 cabang kenjeran Surabaya saja yang menjadi subjek penelitian. Selain harus bersaing dengan diler-diler lainnya, Auto2000 kenjeran juga harus bersaing dengan cabang Auto2000 lain. Kondisi ini menjadikan persaingan antara semua diler semakin ketat, sehingga menyebabkan Auto2000 cabang kenjeran harus dapat bisa bersaing dalam memberikan

<http://sistemasi.ftik.unisi.ac.id>

pelayanan terbaik untuk konsumen. Salah satu strategi yang digunakan oleh Auto2000 kenjeran untuk bisa memenangkan persaingan yakni melalui penerapan strategi *Customer Relationship Management* (CRM). Salah satu konsep pada CRM adalah memperlakukan pelanggan yang berbeda secara berbeda pula, karena kebutuhan mereka berbeda, dan nilainya bagi perusahaan juga mungkin berbeda [1]. atau yang biasa disebut dengan strategi segmentasi pelanggan.

Dengan adanya segmentasi pelanggan bisa membantu pihak Auto2000 cabang kenjeran untuk dapat mengidentifikasi karakteristik para pelanggannya, sehingga pihak Auto2000 cabang kenjeran dapat memberikan pelayanan yang berbeda pada tiap kelompok pelanggannya, yang berdampak pada pelanggan yang loyal menjadi semakin loyal dan pelanggan yang tidak loyal bisa menjadi loyal. Manfaat lain yang bisa didapat dengan adanya segmentasi pelanggan ini yaitu strategi pemasaran yang digunakan bisa dilakukan secara tepat. Segmentasi pelanggan ini penting dilakukan agar pihak auto2000 dapat memperlakukan pelanggan dengan tepat sehingga meningkatkan tingkat loyalitas kepuasan pelanggan.

Saat ini pada Auto2000 cabang kenjeran masih belum memiliki mekanisme untuk melakukan segmentasi pelanggan, kendala yang dihadapi yaitu kesulitan dalam melakukan pengelompokan pelanggan karena banyaknya data yang harus diolah dan tidak adanya tool yang mendukung dalam melakukan pengelompokan pelanggan. Masalah lain yang muncul adalah hanya manajer saja yang bisa melakukan pengelolaan data pelanggan, hal ini dikarenakan pada Auto2000 tidak ada karyawan yang spesifik bertugas mengelola data pelanggan

Berdasarkan masalah yang terjadi, maka diusulkan sebuah solusi yaitu sebuah aplikasi segmentasi pelanggan yang dapat membantu Auto2000 cabang kenjeran dalam melakukan segmentasi pelanggan. Untuk melakukan identifikasi karakteristik dari pelanggan maka diperlukan sebuah model yang memberikan gambaran segala aktivitas pelanggan. Model yang umum digunakan dalam mengelompokkan pelanggan dan menentukan variabelnya adalah *model Recency, Frequency, Monetary* (RFM). Menurut [2] model RFM adalah model berbasis perilaku yang digunakan untuk menganalisis perilaku seorang pelanggan yang selanjutnya dilakukan pemetaan kelompok pelanggan berdasarkan perilaku pelanggan bertransaksi yang telah tercatat dalam basis data transaksi pelanggan.

Sedangkan untuk mengelompokkan pelanggan, metode yang digunakan yaitu metode K-Means *clustering*. Metode K-Means adalah metode yang mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam sebuah kelompok, data berkarakteristik sama akan dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain [3]. Tetapi metode ini juga memiliki kelemahan yaitu sulit menentukan jumlah *cluster* terbaik. *Clustering* adalah suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain [4]. Identifikasi jumlah *cluster* merupakan hal yang paling penting dan utama pada proses *clustering* dengan menggunakan metode K-Means. Sehingga jika jumlah *cluster* yang ditentukan tidak baik maka hasil *cluster* juga tidak akan sesuai dengan yang diharapkan dan tidak akan menghasilkan informasi yang dibutuhkan oleh pengguna. Untuk mengatasi hal tersebut maka digunakanlah metode *Silhouette Coefficient*, metode ini akan membantu metode K-Means dalam menentukan jumlah *cluster* terbaik yang akan digunakan dalam proses pengelompokan pelanggan. sehingga pada penelitian ini untuk mendapatkan variabel karakteristik pelanggan menggunakan model RFM, sedangkan untuk melakukan segmentasi pelanggan berdasarkan variabel yang telah ditentukan menggunakan *clustering* dengan metode K-Means. kelebihan metode K-Means adalah cukup mudah untuk diimplementasikan dan dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan [5].

Dengan adanya aplikasi segmentasi pelanggan dapat membantu perusahaan yaitu auto2000 cabang kenjeran mengelompokkan pelanggan dengan lebih efisien, sehingga dapat membantu perusahaan dalam menentukan strategi pemasaran yang tepat sesuai karakteristik dari tiap kelompok pelanggan. Serta dapat membantu perusahaan dalam menemukan kelompok pelanggan potensial sehingga dapat meningkatkan keuntungan dari perusahaan.

2. Tinjauan Literatur

Pada Penelitian [6], dilakukan analisis segmentasi pelanggan dengan RFM menggunakan Metode *Fuzzy C-Means Clustering*. *Fuzzy C-Means* merupakan salah satu metode pengelompokan yang dapat dilakukan untuk melakukan segmentasi. Metode *Elbow* digunakan untuk membantu dalam menentukan jumlah cluster dalam penerapan *Fuzzy C-Means*. *Partition Coefficient* dan *Euclidean Distance* merupakan metode validasi yang digunakan untuk mengetahui cluster terbaik. Pada penelitian ini didapatkan cluster 3 merupakan hasil terbaik. Hasil cluster divisualisasikan dengan dashboard dengan beberapa grafik yang memuat hasil segmentasi pelanggan berdasarkan nilai RFM dari transaksi yang dilakukan pelanggan.

Hasil penelitian [7], melakukan segmentasi pelanggan menggunakan model RFM dan teori *rough set* untuk memahami karakteristik pelanggan. Hasil dari ini adalah 4 segmen pelanggan yang dimiliki oleh perusahaan dan karakteristik masing-masing pelanggan. Karakteristik pelanggan ini akan membantu PT Abbott Indonesia cabang Malang untuk mengambil keputusan dalam memprioritaskan tenaga dan sumber dayanya ke pelanggan tertentu (potensial). Setelah itu juga dihasilkan 31 aturan klasifikasi dengan tingkat akurasi 95% yang digunakan untuk mengklasifikasikan pelanggan lama atau baru ke dalam segmen pelanggan yang terbentuk dengan cepat dan tepat.

Mengacu pada penelitian [8], melakukan pengelompokan member pada *alvaro fitness* dengan menggunakan metode *K-Means Clustering* dan *Silhouette Coefficient*. Hasil pengujian *Silhouette Coefficient* setiap distance measure nya, antara lain *Euclidean Distance* bernilai 0,232149, *Manhattan Distance* bernilai 0,240016, dan *Chebyshev Distance* bernilai 0.242821. Berdasarkan hasil dari pengujian *silhouette coefficient* yang dilakukan, distance measure paling optimal untuk kasus ini adalah *Chebyshev Distance*, yaitu dengan nilai *silhouette coefficient* paling mendekati 1 adalah 0.242821.

Pada Penelitian ini digunakan model RFM untuk menganalisa perilaku pelanggan yang ada di AUTO2000 kenjeran dengan menentukan beberapa variabel yaitu *Recency*, *Frequency*, dan *Monetary*. setelah karakteristik pelanggan dianalisa selanjutnya adalah mengelompokkan pelanggan menjadi beberapa kelompok berdasarkan karakteristiknya menggunakan metode *K-Means* dimana untuk menentukan jumlah klaster pada *K-Means* menggunakan metode *Silhouette Coefficient*. *Silhouette Coefficient* digunakan untuk melihat kualitas dan kekuatan dari *cluster*. Dengan pemodelan ini diharapkan dapat membantu pihak AUTO2000 untuk menentukan strategi pemasaran yang tepat kepada pelanggannya berdasarkan karakteristik yang sudah dianalisa sebelumnya.

3. Metode Penelitian

Pada Penelitian ini digunakan beberapa metode yang digunakan dalam pengembangan aplikasi segmentasi pelanggan, metode untuk proses segmentasi pelanggan yaitu metode *K-Means*, model RFM, dan *Silhouette Coefficient*.

3.1 RFM

Menurut Hughes pada [2] RFM adalah model yang digunakan untuk menganalisis perilaku seorang pelanggan kemudian dilakukan pemetaan kelompok pelanggan berdasarkan perilaku pelanggan bertransaksi yang telah tercatat dalam dalam basis data transaksi pelanggan. Model RFM sudah umum digunakan oleh banyak kalangan dan sudah digunakan untuk memprediksi perilaku pelanggan selama lebih dari 60 tahun dan termasuk dalam salah satu teknik paling kuat yang tersedia untuk menganalisa perilaku dari pelanggan. RFM terdiri dari 3 variabel yaitu *Recency*, *Frequency*, dan *Monetary* [9][10]. Menurut [11] berikut ini pembagian kategori pelanggan berdasarkan skor RFM yang ditampilkan pada Tabel 1 :

Tabel 1. Kategori Pelanggan Berdasarkan Skor RFM

| Kategori | Skor Recency | Skor Kombinasi Frequency & Monetary | Deskripsi |
|-----------------------------|---------------------|--|---|
| Champion | 4-5 | 4-5 | Melakukan servis baru-baru ini, sering melakukan servis dan sudah mengeluarkan banyak uang untuk melakukan servis. |
| Loyal Customers | 2-5 | 3-5 | Sering melakukan servis dan mengeluarkan banyak uang. Responsif terhadap promosi. |
| Potential Loyalists | 3-5 | 1-3 | Melakukan servis baru-baru ini. Sudah melakukan servis beberapa kali dan mengeluarkan uang dengan cukup. |
| Recent Customers | 4-5 | 0-1 | Melakukan servis baru-baru ini tapi tidak sering. |
| Promising Customers | 3-4 | 0-1 | Melakukan servis baru-baru ini, tetapi belum mengeluarkan banyak uang. |
| Customers Needing Attention | 2-3 | 2-3 | Pelanggan diatas rata-rata umum tapi belum melakukan servis baru-baru ini. |
| About To Sleep | 2-3 | 0-2 | Pelanggan dibawah rata-rata umum. Perlu mengajak servis kembali agar tidak kehilangan pelanggan. |
| At Risk | 0-2 | 2-5 | Sering melakukan servis dan sudah mengeluarkan banyak uang tapi sudah lama tidak melakukan servis. Perlu mengajak servis kembali. |
| Can't Lose Them | 0-1 | 4-5 | Sering melakukan servis dan sudah mengeluarkan banyak uang tapi sudah lama tidak kembali servis. Perlu mengajak servis kembali. |
| Hibernating | 1-2 | 1-2 | Servis yang dilakukan terakhir sudah lama sekali, tidak sering melakukan servis dan sedikit mengeluarkan uang. |

3.2 K-Means

Menurut [3], [12] dan [13] K-Means adalah metode yang mempartisi data yang ada kedalam bentuk dua atau lebih kelompok, metode ini mempartisi data ke dalam sebuah kelompok sehingga data berkarakteristik sama akan dimasukan kedalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain.

Sedangkan menurut [14] metode K-Means merupakan algoritma klusterisasi yang paling tua dan paling banyak digunakan dalam berbagai aplikasi kecil hingga menengah karena kemudahan implementasinya. Menurut [15], langkah-langkah algoritma K-Means adalah sebagai berikut:

1. Pilih secara acak nilai k dan pusat cluster awal. Dalam penelitian ini jumlah *cluster*(k) akan ditentukan melalui metode *Silhouette Coefficient*.
2. Jarak antara data dan pusat *cluster* dihitung menggunakan *Euclidian Distance*. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

dimana:

$D(i,j)$ = Jarak data ke i ke pusat cluster j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

3. Data ditempatkan dalam *cluster* terdekat, dihitung dari tengah *cluster*.
4. Pusat *cluster* baru akan ditentukan bila semua data telah ditetapkan dalam *cluster* terdekat.
5. Proses penentuan pusat *cluster* dan penempatan data dalam *cluster* diulangi sampai nilai centroid tidak berubah lagi.

3.3 Silhouette Coefficient

Menurut [16] *Silhouette Coefficient* adalah suatu metode yang digunakan untuk melihat kualitas dan kekuatan dari *cluster*, dimana *cluster* tersebut diukur seberapa baik suatu objek ditempatkan dalam suatu cluster. Metode ini merupakan gabungan dari metode *cohesion* dan *separation*. Metode ini memiliki tahapan dalam perhitungannya yaitu sebagai berikut:

1. Menghitung rata-rata jarak a_i dan b_i dari data yang terpilih dengan menggunakan Persamaan (2) dan Persamaan (3).

$$a_i^j = \frac{1}{m_j - 1} \sum_{r=1}^{m_j} d(x_i^j, x_r^j) \quad (2)$$

$i = 1, 2, \dots, m_j$

$$b_i^j = \min \left\{ \frac{1}{m_n} \sum_{r=1}^{m_n} d(x_i^j, x_r^n) \right\}, i \quad (3)$$

$i = 1, 2, \dots, m_n$

dimana:

$d(x_i^j, x_r^j)$ = Jarak data ke-i dengan data ke r dalam satu klaster j.

m_j = Jumlah data dalam klaster ke-j.

Persamaan (2) merupakan persamaan untuk pencarian nilai a_i dimana nilai a_i adalah rata-rata jarak data ke-i terhadap semua data lainnya dalam satu *cluster*.

Persamaan (3) adalah persamaan untuk mencari nilai b_i adalah hasil yang didapatkan melalui perhitungan rata-rata jarak a_i terhadap seluruh data dari *cluster* lain, kemudian dari hasil rata-rata dari b_i diambil perolehan nilai minimum yang akan digunakan untuk pencarian nilai S_i .

2. Setelah seluruh rata-rata jarak dari seluruh data selesai dicari, kemudian mencari nilai terkecil dari *cluster* yang telah dicari sebelumnya.
3. Nilai S_i dapat ditemukan dengan cara seperti pada persamaan (4).

$$S_i^j = \frac{b_i^j - a_i^j}{\max\{a_i^j, b_i^j\}} \quad (4)$$

dimana:

S_i^j = Rumus *Silhouette Coefficient*.

b_i^j = Rata-rata jarak data ke-i terhadap semua data dari *cluster* lain, diambil data yang paling kecil.

a_i^j = Rata-rata jarak data ke-i terhadap semua data lainnya dalam satu *cluster*.

Pada persamaan (4) nilai yang digunakan dalam perhitungan $b_i - a_i$ adalah nilai minimum dari kelompok yang diselisihkan dengan jarak rata-rata a_i . Perhitungan tersebut dilakukan agar mendapatkan nilai maksimum dari kedua kelompok. Apabila hasil dari seluruh *silhouette coefficient* tiap *cluster* telah ditemukan, perhitungan S_i global dapat dilakukan dengan cara menghitung rata-rata dengan membagi seluruh nilai S_i dengan sejumlah data yang ada.

Berikut ini contoh perhitungan dari Silhouette Coefficient:

Cluster 1

$$a_1 = \sqrt{(1-1-4)^2} = 4$$

$$b_1 =$$

Cluster 3

$$(1,1,4) \square (5,4,5) = \sqrt{(1-5)^2 + (1-4)^2 + (4-5)^2} = 5,0990$$

$$(1,1,4) \square (5,2,3) = \sqrt{(1-5)^2 + (1-2)^2 + (4-3)^2} = 4,2426$$

$$(1,1,4) \square (4,1,3) = \sqrt{(1-4)^2 + (1-1)^2 + (4-3)^2} = 3,1622$$

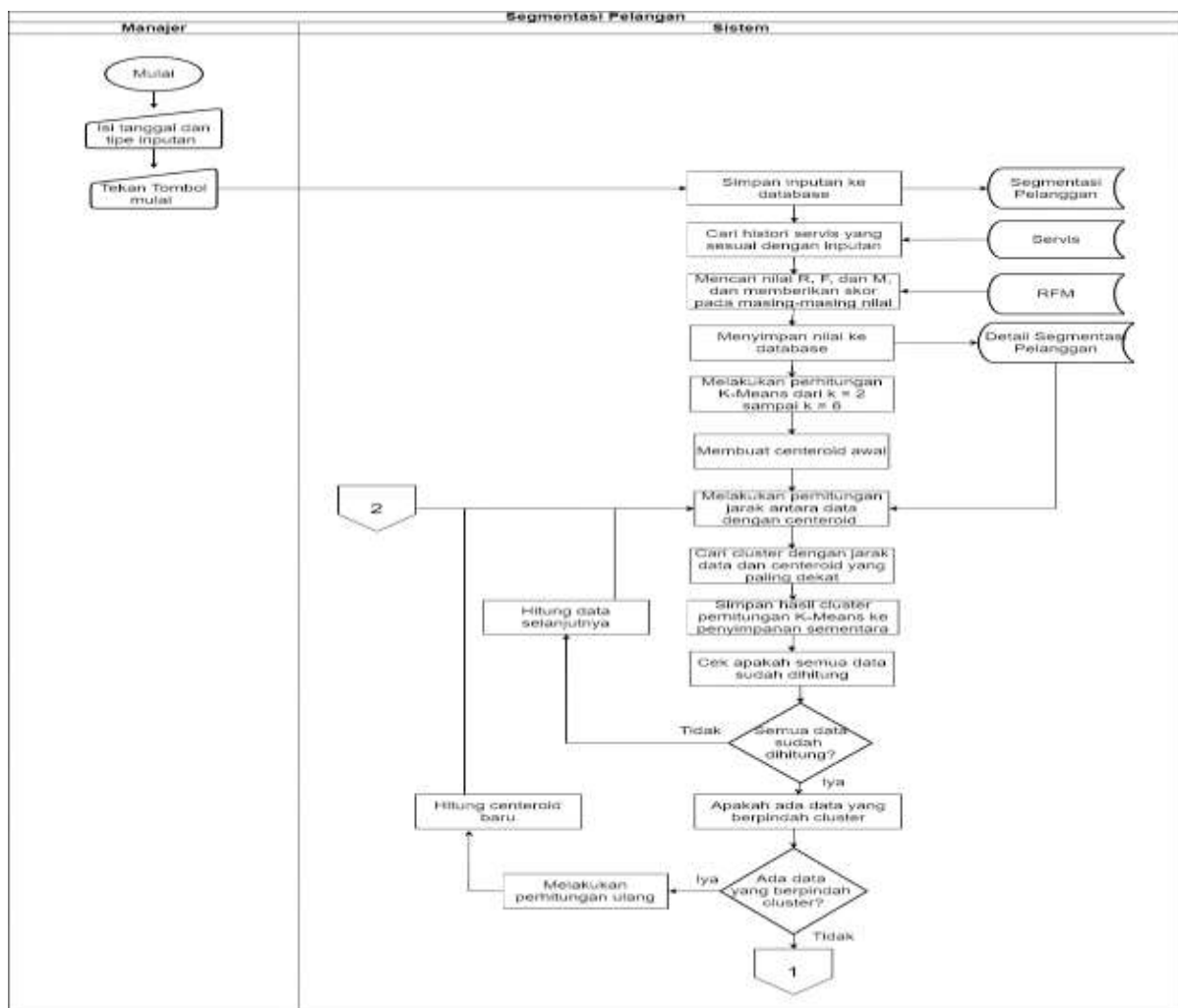
$$(1,1,4) \square (2,5,5) = \sqrt{(1-2)^2 + (1-5)^2 + (4-5)^2} = 4,2426$$

$$b_1 = 5,0990 + 4,2426 + 3,1622 + 4,2426 = 16,7464/4 = 4,1866$$

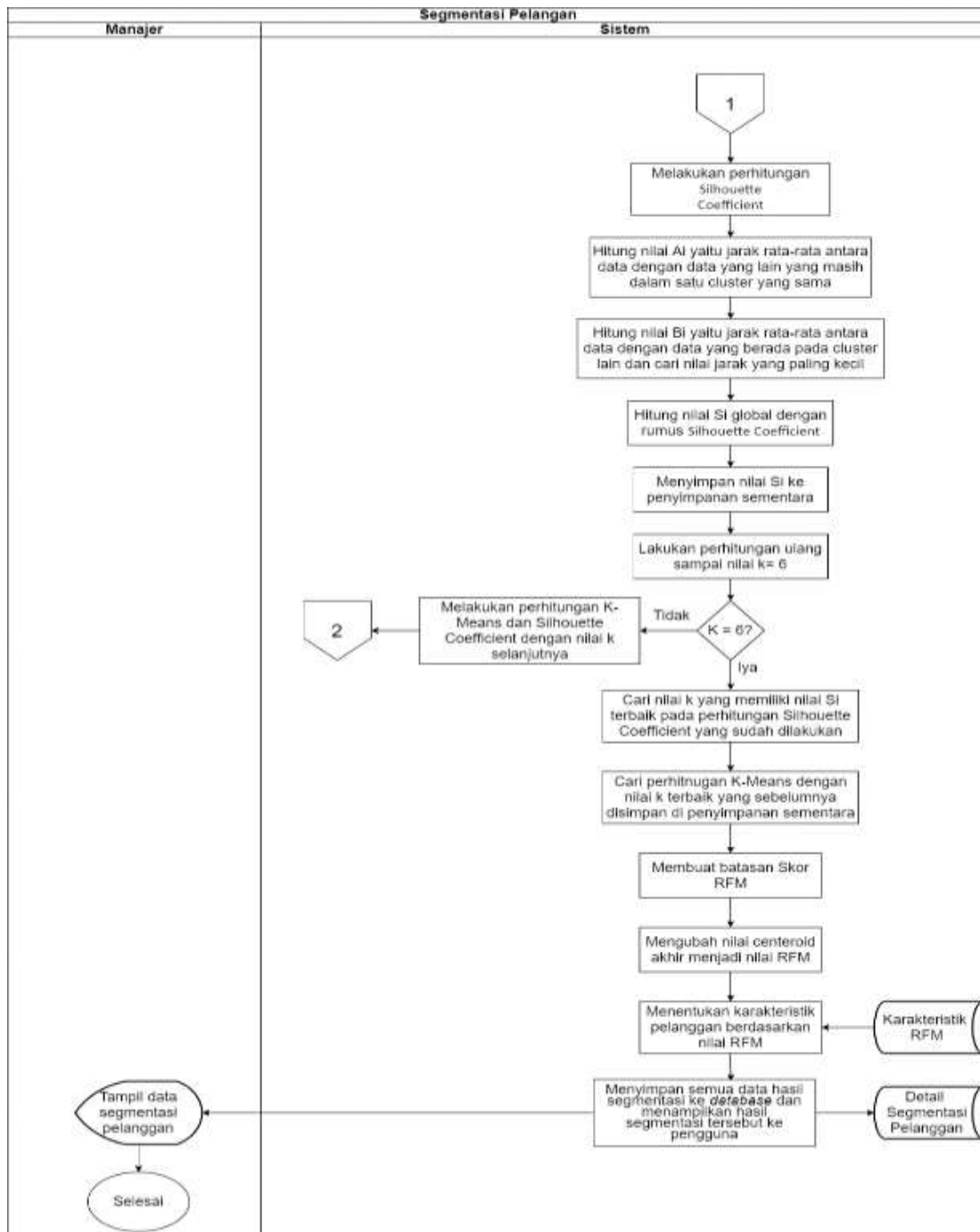
$$S_1 = 4,1866 - 44,1866 = 0,0445$$

3.4 Proses Segmentasi Pelanggan

pada tahap *system flowchart* dilakukan perancangan bagan alur sistem yang akan dibuat, terdiri dari pengelolaan data, segmentasi pelanggan, dan laporan segmentasi pelanggan. Berikut Gambar 1 dan Gambar 2 merupakan *system flowchart* dari proses segmentasi pelanggan:



Gambar 1. System Flowchart Segmentasi Pelanggan-1



Gambar 2. System Flowchart Segmentasi Pelanggan-2

Pada Gambar 1 dan Gambar 2 merupakan *System Flowchart* yang menggambarkan proses untuk melakukan segmentasi pelanggan. Proses segmentasi pelanggan pada aplikasi terbagi menjadi 3 proses yaitu RFM, K-Means, dan *Silhouette Coefficient*. RFM digunakan untuk mengidentifikasi karakteristik dari pelanggan dan membantu K-Means untuk melakukan transformasi data, data yang tidak berbentuk numerik akan dijadikan bentuk numerik agar dapat diolah, sedangkan K-Means digunakan untuk melakukan segmentasi pelanggan dan *Silhouette Coefficient* digunakan untuk menentukan nilai k terbaik pada perhitungan K-Means. Berikut merupakan langkah-langkah segmentasi pelanggan yang berjalan pada aplikasi segmentasi pelanggan:

1. Untuk memulai proses segmentasi pelanggan pengguna harus memilih tanggal segmentasi dan tipe inputan yang akan digunakan, untuk tipe inputan terdiri dari 3 pilihan yaitu: 1) Auto2000 Kenjeran: yaitu hanya data servis pelanggan yang melakukan servis di Auto2000 kenjeran Surabaya saja yang akan diproses. 2) Surabaya, yaitu hanya data servis pelanggan yang melakukan servis di Auto2000 wilayah Surabaya saja yang akan diproses. 3) Bebas, yaitu semua data akan diproses sesuai tanggal yang ditentukan.
2. Setelah pengguna menekan tombol mulai, sistem akan menyimpan data yang dimasukan pelanggan ke dalam *database*, selanjutnya sistem akan mencari histori servis pelanggan yang sesuai dengan tanggal dan tipe inputan yang sudah dipilih.
3. Selanjutnya akan dilakukan tahap pencarian nilai dari R, F, dan M. Nilai R didapatkan dari jarak tanggal segmentasi dilakukan dengan tanggal terakhir servis yang dilakukan pelanggan, nilai F didapatkan dari jumlah servis yang sudah dilakukan, dan nilai M didapatkan dari total jumlah *revenue*.
4. Setelah didapatkan nilai dari masing-masing atribut selanjutnya memberikan skor untuk masing-masing nilai. Pemberian skor dengan cara mencocokkan nilai RFM yang didapat dengan nilai RFM pada *database* RFM. Setelah skor didapat, nilai dan skor RFM akan disimpan ke *database*.
5. Selanjutnya sistem akan melakukan perhitungan K-Means dari k=2 sampai k=6. Nilai K pada K-Means dibatasi hanya 2 – 6 karena 2 adalah batasan minimum K dalam K-Means, sedangkan 6 karena jika nilai K terlalu banyak akan membuat manajer atau pembuat keputusan sulit untuk menentukan strategi perusahaan untuk pelanggan serta semakin tinggi nilai K maka nilai *Silhouette Coefficient* cenderung semakin turun [17].
6. Langkah pertama perhitungan K-Means yaitu sistem akan membuat *centeroid* awal secara acak.
7. Kemudian sistem akan melakukan perhitungan jarak antara data skor RFM dan *centeroid* awal menggunakan teori *Euclidean Distance*
8. Lalu sistem akan mencari *cluster* terdekat dan hasil *cluster* tersebut akan disimpan ke penyimpanan sementara.
9. Setelah semua data sudah dihitung sistem akan mengecek apakah ada perubahan *cluster* atau tidak, jika ada data yang berpindah *cluster* maka sistem akan melakukan perhitungan ulang dengan membuat *centeroid* yang baru dengan cara menjumlahkan nilai RFM tiap-tiap *cluster* lalu dibagi dengan jumlah data tiap *cluster* tersebut. Selanjutnya mulai melakukan perhitungan jarak lagi menggunakan teori *Euclidean Distance*, tapi jika data tidak ada yang berpindah maka sistem akan melanjutkan ke proses selanjutnya yaitu *Silhouette Coefficient*.
10. Untuk melakukan proses *Silhouette Coefficient* pertama sistem akan mencari nilai A_i dengan cara menghitung rata-rata jarak antara data dengan data yang lain yang masih dalam satu *cluster*
11. Selanjutnya sistem akan mencari nilai B_i yaitu dengan cara menghitung jarak rata-rata antara data dengan data yang berada pada *cluster* lain lalu dicari nilai jarak yang paling kecil
12. Setelah itu sistem akan mencari nilai S_i tiap data menggunakan rumus *Silhouette Coefficient*
13. Selanjutnya sistem akan mencari nilai SI Global dengan menggunakan rumus sebagai berikut:
$$SI = \frac{1}{k} \sum_{j=1}^k SI_j \quad (5)$$
dimana:
 SI_j = SI global.
 k = Banyaknya cluster.
lalu nilai S_i global disimpan ke penyimpanan sementara.
14. Jika sistem sudah melakukan perhitungan sampai nilai k=6 maka sistem akan melanjutkan ke proses selanjutnya, tapi jika tidak maka sistem akan melakukan perhitungan ulang dengan nilai k selanjutnya.
15. Proses selanjutnya yaitu mencari perhitungan K-Means dengan nilai k yang memiliki nilai S_i terbaik yang sebelumnya sudah disimpan pada penyimpanan sementara.

16. *Centeroid* akhir pada perhitungan K-Means dengan hasil k terbaik akan diambil untuk dibuat menjadi batasan skor RFM, dengan cara pembagian sama rata sebanyak 5 bagian (20%) dari nilai minimum dan maksimum dari tiap atribut RFM. Batasan skor RFM ini digunakan untuk mengubah nilai *centeroid* akhir menjadi skala RFM pada proses selanjutnya.
17. Kemudian nilai *centeroid* akhir tiap *cluster* akan diubah menjadi nilai RFM lagi menggunakan batasan skor RFM yang sudah dibuat pada proses sebelumnya.
18. Kemudian sistem akan menentukan karakteristik pelanggan dengan cara mencocokkan nilai RFM yang sudah didapat pada proses sebelumnya dengan nilai skala pada *database* karakteristik RFM. Terakhir, sistem akan menyimpan semua data hasil segmentasi ke *database* dan menampilkan hasil segmentasi tersebut ke pengguna.

4. Hasil dan Pembahasan

Dalam melakukan proses segmentasi pelanggan menggunakan model RFM dan metode algoritma K-Means, berikut adalah data yang digunakan. Pada Tabel 2 menunjukkan data sampel yang diambil secara *random* dari data pelanggan pada cabang auto2000 kenjeran Surabaya. Data pada Tabel 2 telah dilakukan pemilihan terhadap parameter yang diperlukan untuk melakukan *clustering*.

Tabel 2. Data Pelanggan

| No | Nama | Jarak tanggal service terakhir | Jumlah service | Jumlah uang yang dikeluarkan |
|----|---------------|--------------------------------|----------------|------------------------------|
| 1 | Arif Satria | 22 Hari | 8 | 5.500.000 |
| 2 | Hansamu Yama | 24 Hari | 4 | 2.550.000 |
| 3 | Irfan Jaya | 32 Hari | 1 | 2.900.000 |
| 4 | David Dasilva | 142 Hari | 15 | 8.000.000 |
| 5 | Koko Ari Arya | 261 Hari | 2 | 3.500.000 |

4.1 Transformasi Data

Agar data di atas dapat diolah dengan menggunakan metode K-Means *clustering*, maka data harus diinisialisasikan terlebih dahulu dalam bentuk angka. Pada Tabel 3, 4 dan 5 menunjukkan hasil inialisasi dari parameter jumlah *service*, jumlah uang yang dikeluarkan dan jarak *service* terakhir sesuai dengan model RFM yaitu *Recency*, *Frequency*, dan *Monetary*.

Tabel 3. Inialisasi Data Jumlah Service

| No | Jumlah Service | Inisial |
|----|----------------|---------|
| 1 | 1 – 2 | 1 |
| 2 | 3 – 4 | 2 |
| 3 | 5 – 6 | 3 |
| 4 | 7 – 8 | 4 |
| 5 | >8 | 5 |

Tabel 4. Inialisasi Data Jumlah Uang Yang Dikeluarkan

| No | Jumlah Uang Yang Dikeluarkan | Inisial |
|----|--------------------------------|---------|
| 1 | > Rp4.000.000 | 5 |
| 2 | Rp 3.000.000 – 3.999.999 | 4 |
| 3 | Rp 2.000.000 – Rp Rp 2.999.999 | 3 |
| 4 | Rp 1.000.000 – Rp 1.999.999 | 2 |
| 5 | Rp 0 – Rp 999.999 | 1 |

Tabel 5. Inisialisasi Data Jarak Tanggal Service Terakhir

| No | Jarak Tanggal Service Terakhir | Inisial |
|----|--------------------------------|---------|
| 1 | 30 hari | 5 |
| 2 | 31 – 60 hari | 4 |
| 3 | 61 – 90 hari | 3 |
| 4 | 91 – 180 hari | 2 |
| 5 | >180 | 1 |

4.2 Pengolahan Data

Setelah semua data pelanggan ditransformasi ke dalam bentuk angka, maka data tersebut dapat dikelompokkan dengan menggunakan algoritma *K-Means Clustering*. Untuk dapat melakukan pengelompokan data-data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah, yaitu:

1. Tentukan jumlah *cluster* yang diinginkan. Dalam penelitian ini jumlah *cluster* untuk pengelompokan pelanggan akan ditentukan melalui metode *Silhouette Coefficient*, dimana range nilai *K (cluster)* yang dibuat untuk uji coba adalah 2- 6.
2. Tentukan titik pusat awal dari setiap *cluster*. Dalam penelitian ini titik pusat awal ditentukan secara random dan didapat titik pusat setiap *cluster* dapat dilihat pada Tabel 6.

Tabel 6. Titik Pusat Awal Setiap Cluster

| <i>Titik Pusat Awal</i> | <i>Tanggal service terakhir</i> | <i>Jumlah service</i> | <i>Jumlah uang yang dikeluarkan</i> |
|-------------------------|---------------------------------|-----------------------|-------------------------------------|
| <i>Cluster 1</i> | 1 | 1 | 3 |
| <i>Cluster 2</i> | 2 | 2 | 1 |
| <i>Cluster 3</i> | 3 | 3 | 3 |
| <i>Cluster 4</i> | 1 | 1 | 5 |

3. Tempatkan setiap data pada *cluster*. Dalam penelitian ini digunakan metode *Hard K-Means* untuk mengalokasikan setiap data ke dalam suatu *cluster*, sehingga data akan dimasukkan dalam suatu *cluster* yang memiliki jarak paling dekat dengan titik pusat dari setiap *cluster*. Untuk mengetahui *cluster* mana yang paling dekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap *cluster*. Sebagai contoh, akan dihitung jarak dari data pelanggan pertama ke pusat *cluster* pertama:

$$D(1,1) = \sqrt{(5-1)^2} + \sqrt{(4-1)^2} + \sqrt{(5-3)^2} = 5,33$$

$$D(1,2) = \sqrt{(5-2)^2} + \sqrt{(4-2)^2} + \sqrt{(5-1)^2} = 5,38$$

$$D(1,3) = \sqrt{(5-3)^2} + \sqrt{(4-3)^2} + \sqrt{(5-3)^2} = 3$$

$$D(1,4) = \sqrt{(5-1)^2} + \sqrt{(4-1)^2} + \sqrt{(5-5)^2} = 5$$

Berdasarkan hasil perhitungan di atas dapat disimpulkan bahwa jarak data pelanggan pertama yang paling dekat adalah *cluster* 3, sehingga data pelanggan pertama dimasukkan ke dalam *cluster* 3. Hasil perhitungan selengkapnya untuk 5 data pelanggan pertama dapat di lihat pada Tabel 7.

Tabel 7. Contoh Hasil Perhitungan Setiap Data ke Setiap Cluster

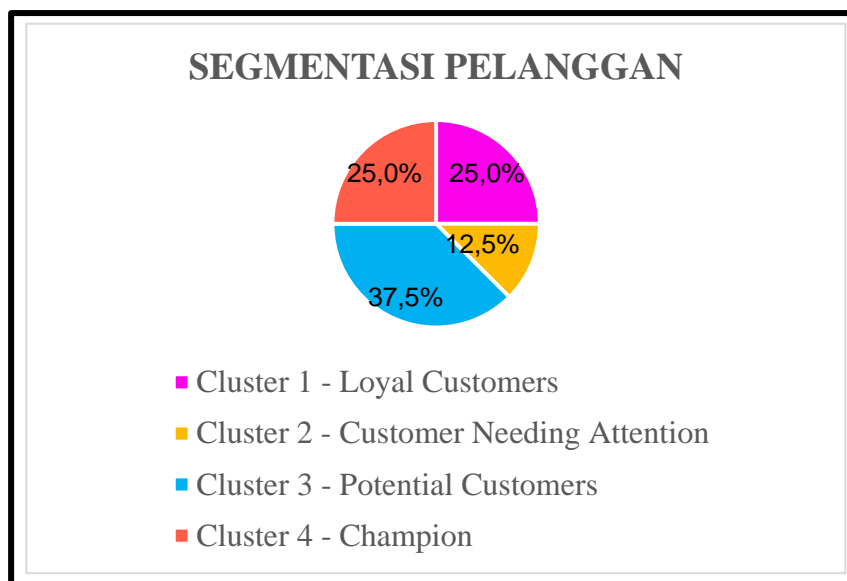
| No | Nama | Jarak tanggal service terakhir | Jumlah service | Jumlah uang yang dikeluarkan | Jarak ke | | | | Jarak terdekat ke Cluster |
|----|---------------|--------------------------------|----------------|------------------------------|----------|------|------|------|---------------------------|
| | | | | | C1 | C2 | C3 | C4 | |
| 1 | Arif Satria | 5 | 4 | 5 | 5,38 | 5,38 | 3 | 5 | C3 |
| 2 | Hansamu Yama | 5 | 2 | 3 | 4,12 | 3,60 | 2,23 | 4,58 | C3 |
| 3 | Irfan Jaya | 4 | 1 | 3 | 3 | 3 | 2,23 | 3,60 | C3 |
| 4 | David Dasilva | 2 | 5 | 5 | 4,58 | 5 | 3 | 4,12 | C3 |
| 5 | Koko Ari Arya | 1 | 1 | 4 | 1 | 3,31 | 3 | 1 | C1 |

- Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut.
 - Setelah didapatkan titik pusat yang baru dari setiap *cluster*, lakukan kembali tahapan langkah ketiga hingga titik pusat dari setiap *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain.
- hasil perhitungan proses segmentasi pelanggan pada aplikasi sudah tepat.

| No | Nama Pelanggan | Total Skor | Karakteristik | Jumlah Kunjungan |
|----|-----------------|------------|-----------------|------------------|
| 1 | Richmad Irianto | 14 | Champion | 8 |
| 2 | Davis Dasilva | 12 | Champion | 15 |
| 3 | Rizky Ridho | 11 | Loyal Customers | 1 |
| 4 | Hansamu Yama | 10 | Loyal Customers | 4 |
| 5 | Irfan Jaya | 8 | Loyal Customers | 1 |

Gambar 3. Tampilan 5 Pelanggan Terbaik

Gambar 3 menunjukkan hasil urutan 5 pelanggan terbaik pada cabang auto2000 cabang kenjeran. hal ini diperoleh setelah melakukan pemodelan karakteristik pelanggan berdasarkan model RFM dan *clustering* menggunakan metode K-Means. 5 pelanggan terbaik memiliki karakteristik *champion* dan *loyal customer* dimana berdasarkan model RFM karakteristik *champion* menandakan bahwa pelanggan melakukan servis baru-baru ini, sering melakukan servis dan sudah mengeluarkan banyak uang untuk melakukan servis. Sedangkan untuk *loyal customer* yaitu sering melakukan servis dan mengeluarkan banyak uang serta Responsif terhadap promosi.



Gambar 4. Tampilan Hasil Clustering

Gambar 4 menunjukkan hasil prosentasi karakteristik pelanggan yang diperoleh dari data beberapa pelanggan yang telah diinputkan pada sistem. Pada Gambar 5 dapat dilihat bahwa dari data pelanggan yang ada dihasilkan beberapa karakteristik yaitu 25% berkarakteristik *champion*, 25% *loyal customer*, 37.5% *potential loyalist* dan 12.5% *customer needing attention*. Deskripsi dari masing-masing karakteristik dapat dilihat pada Tabel 1

| Nama Pelanggan | Skor | Monetary | Recency | Frequency | Karakteristik |
|-----------------|------|---------------|---------|-----------|---------------|
| Davis Dasilva | 12 | Rp. 8.000.000 | 142 | 15 | Champion |
| Rachmad Irianto | 14 | Rp. 5.500.000 | 22 | 8 | Champion |

Gambar 5. Data Detail Pelanggan Dengan Karakteristik Champions

Gambar 5 menunjukkan tampilan detail untuk pelanggan yang berkarakter *champion* untuk melihat masing-masing nilai parameter pada model RFM yaitu *recency*, *monetary* dan *frequency* untuk masing-masing pelanggan

5. Kesimpulan

Berdasarkan semua tahapan yang sudah dilakukan yaitu melakukan pemetaan kelompok pelanggan dengan model RFM berdasarkan variabel *Recency*, *Frequency*, dan *Monetary*, clustering pelanggan berdasarkan karakteristik menggunakan *K-Means* dan *Silhouette Coefficient*. Maka aplikasi segmentasi pelanggan dapat melakukan proses segmentasi pelanggan yang dapat membagi pelanggan kedalam beberapa kelompok dan menemukan kelompok pelanggan potensial sehingga dapat memudahkan pihak Auto2000 untuk memberikan strategi pemasaran seperti diskon dan promosi sesuai dengan karakteristik pelanggan yang sudah dikelompokkan. Untuk penelitian selanjutnya, Performa dan waktu pemrosesan proses segmentasi pelanggan pada aplikasi bisa lebih ditingkatkan dengan menggunakan metode yang berbeda untuk menghasilkan proses yang lebih optimal.

Referensi

- [1] K. Rainer, B. Prince, and C. Cegielski, *Introduction To Information System Supporting and Transforming Business*. 2014.
- [2] T. Hardiani, S. Sulisty, and R. Hartanto, "Segmentasi Nasabah Tabungan Menggunakan Model Rfm (Recency , Frequency , Monetary) dan K-Means pada Lembaga Keuangan Mikro," *Semin. Nas. Teknol. Inf. Dan Komun. Terap.*, 2015.
- [3] Eko Prasetyo, *Data Mining : Konsep dan Aplikasi Menggunakan Matlab*. 2013.
- [4] J. O. Ong, "Implementasi Algoritma K-Means Clustering untuk Menentukan Strategi Marketing President University," *J. Ilm. Tek. Ind.*, 2013.
- [5] X. Wu *et al.*, "Top 10 Algorithms in Data Mining," *Knowl. Inf. Syst.*, 2008, doi: 10.1007/S10115-007-0114-2.
- [6] W. A. Taqwim, N. Y. Setiawan, and F. A. Bachtiar, "Analisis Segmentasi Pelanggan Dengan Rfm Model Pada Pt. Arthamas Citra Mandiri Menggunakan Metode Fuzzy C-Means Clustering," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, Vol. 3, No. 2, pp. 2548–964, 2019.
- [7] R. A. Hendrawan and A. Utamima, "Customer Segmentation Via Rfm Model and Rough Set Theory to Understand Customer's Characteristic (Case Study: PT. Abbott Indonesia,Tbk Cabang Malang)," Institut Teknologi Sepuluh Nopember (ITS), 2015.
- [8] A. Mario, S. Herry, and H. Nasution, "Pemilihan Distance Measure pada K-Means Clustering untuk Pengelompokan Member Di Alvaro Fitness," *J. Sist. Dan Teknol. Inf.*, 2016.
- [9] V. Aggelis and Christodoulakis, "Customer Clustering Using RFM Analysis," In *Proceedings of The 9th International Conference On Computers (ICCOMP)*, 2005.
- [10] C. H. Cheng and Y. S. Chen, "Classifying The Segmentation of Customer Value Via RFM Model and RS Theory," *Expert Syst. Appl.*, 2009, doi: 10.1016/J.Eswa.2008.04.003.
- [11] S. A. Sutresno, A. Iriani, and E. Sedyono, "Metode K-Means Clustering dengan Atribut RFM untuk Mempertahankan Pelanggan," *J. Tek. Inform. Dan Sist. Inf.*, Vol. 4, P. 433, 2018, doi: 10.28932/Jutisi.V4i3.878.
- [12] Y. Agusta, "K-Means-Penerapan, Permasalahan dan Metode Terkait," *J. Sist. Dan Inform.*, Vol. 3, pp. 47–60, 2007.
- [13] N. Ningsih *et al.*, "Pengembangan Simulator ESM Untuk Identifikasi Sinyal Radar dengan Metode Clustering Menggunakan K-Means," *Semin. Teknol. Dan Rekayasa*, 2015.
- [14] Suyanto, "Data Mining untuk Klasifikasi dan Klasterisasi Data," *Springerreference*, 2017.
- [15] Agus Nur Khormarudin, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Ilmu Komputer*, 2016.
- [16] L. Hidayat and W. Firdaus Mahmudy, "Pengelompokan Data Hasil Tes Kepribadian 16pf Sopir Bus Menggunakan Algoritma Genetika" *J. Teknol. Inf. Dan Ilmu Komput.*, Vol. 3, No. 3, pp. 163–168, 2016.
- [17] S. Monalisa *et al.*, "Segmentasi Perilaku Pembelian Pelanggan Berdasarkan Model RFM dengan Metode K-Means," *J. Ilm. Mhs. Univ. Surabaya*, 2013.