

Implementasi Algoritma K-Medoids Clustering Untuk Menentukan Segmentasi Pelanggan

Nita Mirantika^{*1}, Tri Septiar Syamfithriani², Ragel Trisudarmo³

^{1,2,3}Sistem Informasi, Fakultas Ilmu Komputer, Universitas Kuningan

Email: ^{*1}nita.mirantika@uniku.ac.id, ²tri@uniku.ac.id, ³ragel.trisudarmo@uniku.ac.id

Abstract

In managing customer relationships, companies must recognize the characteristics of each customer by creating an appropriate management strategy according to these characteristics. Customer characteristics can be identified by segmenting customers into groups based on their characteristic similarities. This study aims to segment customers at XYZ Store using data mining clustering techniques, namely by using the K-Medoids algorithm which uses data representatives as cluster centers. This study uses RFM analysis, which consists of three attributes: recency, frequency, and monetary. Determining the optimal number of clusters uses the Silhouette coefficient method and evaluation of cluster results uses the Davies-Bouldin Index (DBI) method. By using the RStudio application, the results of customer segmentation obtained three customer segments, namely loyal customers who have high (new) recency values, the most frequency and the most monetary are 21 customers; typical customers who have low (old) recency, medium frequency and moderate monetary are 31 customers; and newcomers who have the highest (most recent) recency values, lowest frequency and monetary are 61 customers. By obtaining three customer segments, the company can make policies and business strategies for managing customer relationships so that the company's profits are expected to increase.

Keywords – K-Medoids Algorithm, Clustering, Customer Segmentation

Abstrak

Dalam melakukan pengelolaan hubungan dengan pelanggan, perusahaan harus mengenali karakteristik setiap pelanggan sehingga perusahaan dapat membuat strategi pengelolaan yang tepat sesuai dengan karakteristik tersebut. Identifikasi karakteristik pelanggan dapat diperoleh dengan melakukan segmentasi pelanggan yaitu dengan mengelompokkan pelanggan yang memiliki kesamaan karakteristik. Tujuan penelitian ini adalah untuk melakukan segmentasi pelanggan pada XYZ Store dengan menggunakan teknik data mining clustering yaitu dengan menggunakan algoritma K-Medoids yang menggunakan perwakilan data sebagai pusat cluster. Atribut yang digunakan dalam penelitian ini menggunakan analisis RFM yang terdiri dari tiga atribut yaitu: recency, frequency dan monetary. Penentuan jumlah cluster optimal menggunakan metode Silhouette coefficient dan evaluasi hasil cluster menggunakan metode Davies-Bouldin Index (DBI). Dengan menggunakan bantuan aplikasi RStudio hasil segmentasi pelanggan diperoleh tiga segmen pelanggan yaitu loyal customer yang mempunyai nilai recency tinggi (baru), frequency paling banyak dan monetary paling banyak sebanyak 21 pelanggan; typical customer yang memiliki recency rendah (lama), frequency sedang dan monetary sedang sebanyak 31 pelanggan; dan newcomer yang mempunyai nilai recency paling tinggi (paling baru), frequency dan monetary paling rendah sebanyak 61 pelanggan. Dengan diperolehnya tiga segmen pelanggan, maka perusahaan dapat membuat kebijakan dan strategi bisnis dalam melakukan pengelolaan hubungan dengan pelanggan sehingga diharapkan keuntungan perusahaan akan meningkat.

Kata Kunci – Algoritma K-Medoids, Clustering, Segmentasi Pelanggan

1. PENDAHULUAN

Pengelolaan hubungan dengan pelanggan merupakan faktor penting untuk kesuksesan perusahaan di era globalisasi dengan persaingan bisnis yang semakin ketat [1]. Dengan mengelola hubungan yang baik dengan pelanggan maka kepuasan pelanggan dapat tercapai dan mendapatkan loyalitas. Pencapaian kepuasan dan loyalitas pelanggan yang tinggi akan berkaitan erat dengan perolehan keuntungan perusahaan dan peningkatan pangsa pasar [2]. Dalam melakukan pengelolaan hubungan dengan pelanggan, perusahaan harus mengenali karakteristik setiap pelanggan, sehingga perusahaan dapat membuat strategi pengelolaan yang tepat sesuai dengan karakteristik tersebut [3, 4].

XYZ Store merupakan perusahaan ritel yang bergerak dalam penjualan produk *fashion*, mempunyai pelanggan yang tersebar diberbagai daerah dengan karakteristik yang beragam. Dalam mengidentifikasi karakteristik tersebut, diperlukan segmentasi pelanggan untuk mengelompokkan pelanggan yang memiliki kesamaan karakteristik [4]. Dengan melakukan segmentasi pelanggan maka dapat mengetahui perilaku (*behavior*) pelanggan yang berulang dan membentuk pola yang dapat digunakan untuk membuat strategi bisnis [5, 6]. Analisis RFM dapat digunakan untuk mengamati perilaku pelanggan dengan menganalisis nilai pelanggan [4, 5, 7]. Didalamnya terkandung tiga aspek. Aspek pertama adalah *recency* (R) atau kebaruan yaitu melihat seberapa baru pelanggan melakukan pembelian. Aspek kedua adalah *frequency* (F) yaitu seberapa sering pelanggan melakukan pembelian. Dan aspek ketiga adalah *monetary* (M) atau keuangan yaitu jumlah total uang yang dihabiskan pelanggan dalam melakukan pembelian [5, 6, 7]. Segmentasi pelanggan membagi pelanggan menjadi beberapa kelompok atau *cluster*, dimana setiap *cluster* dapat memberikan *insight* dalam membuat kebijakan bagi perusahaan.

Segmentasi pelanggan dapat dilakukan dengan menggunakan teknik *data mining clustering* [8, 9]. *Data mining* adalah aktivitas yang menggambarkan sebuah

proses analisis yang terjadi secara iteratif pada *database* yang besar, dengan tujuan mengekstrak informasi dan *knowledge* yang akurat dan berpotensi berguna untuk *knowledge workers* yang berhubungan dengan pengambilan keputusan [10]. Melalui *data mining*, perusahaan dapat menganalisis data pelanggan dengan menggunakan data yang dikumpulkan untuk memandu, mengoptimalkan, dan mengotomatisasi transaksi yang berpusat pada pelanggan [11]. Metode analisis *data mining* yang dapat digunakan untuk melakukan segmentasi pelanggan adalah metode *clustering*, yaitu metode analisis yang bertujuan untuk melakukan segmentasi populasi yang heterogen menjadi sejumlah kelompok yang beranggotakan observasi dengan karakteristik yang homogen [10]. Terdapat beberapa algoritma pada metode *clustering*, yaitu diantaranya adalah algoritma K-Medoids.

Algoritma K-Medoids atau sering disebut juga dengan algoritma PAM (*Partitioning Around Medoid*) merupakan salah satu metode *clustering* yang diusulkan untuk mengatasi kelemahan algoritma K-Means yang sensitif terhadap *outlier* karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data [11, 12]. Algoritma K-Medoids menggunakan *medoid* (perwakilan) data sebagai ganti dari *mean* untuk pusat *cluster*. *Medoid* adalah objek data paling sentral di antara titik-titik *cluster*. Oleh karena itu, K objek dipilih secara acak sebagai medoid untuk mewakili *cluster*, dan semua objek data ditugaskan ke *cluster* dengan medoid terdekat [11].

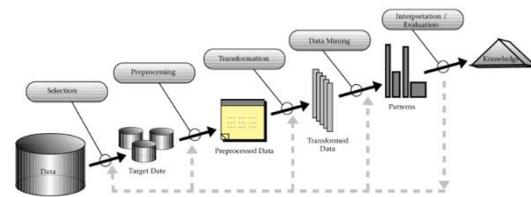
Terdapat beberapa penelitian sebelumnya yang melakukan segmentasi pelanggan dengan menggunakan algoritma K-Medoids. Penelitian yang dilakukan oleh Sulistyawati dan Sadikin [13] mengenai segmentasi pelanggan pada Perum BULOG kantor wilayah DKI Jakarta & Banten dengan menggunakan algoritma K-Medoids, diperoleh hasil jumlah *cluster* optimal adalah 3 *cluster* dengan nilai maksimum *Silhouette Index* adalah 0,375 dan nilai minimum *Davies Doublin Index* adalah 1,030. Segmen pelanggan hasil penelitian adalah *lost customer*, *core customer*, dan

new customer. Penelitian lain yang dilakukan oleh Madani, dkk [14] mengenai segmentasi pelanggan pada BC HNI 2 Pekanbaru dengan menerapkan algoritma K-Medoids, diperoleh hasil 2 *cluster* terbaik berdasarkan DBI (*Davies Bouldin Index*). Dengan nilai DBI 0.492, serta segmen pelanggan hasil penelitian yang didapatkan adalah 452 *core customer* dan 351 *lost customer*. Penelitian lain yang dilakukan oleh Purnama, G., dkk [15] mengenai segmentasi pelanggan pada perusahaan *fashion* di kota Bandung menggunakan algoritma K-Medoids, diperoleh hasil terdapat tiga *cluster* yaitu *cluster* 1 termasuk kelompok *uncertain new customers*, *cluster* 2 termasuk kelompok *lost customer* dan *cluster* 3 termasuk kelompok *core customer*. penyusunan strategi bisnis ditentukan berdasarkan tipe dan karakteristik pelanggan pada setiap *cluster* atau segmen pelanggan yang terbentuk.

Penelitian ini dilakukan untuk menghasilkan segmentasi pelanggan pada XYZ Store dengan menerapkan metode *clustering* menggunakan algoritma K-Medoids. Segmentasi pelanggan yang dihasilkan diharapkan dapat menjadi masukan bagi XYZ Store untuk membuat strategi dalam melakukan pengelolaan hubungan dengan pelanggan.

2. METODE PENELITIAN

Tahapan penelitian dilakukan dengan mengacu pada proses *Knowledge Discovery Data* (KDD). KDD sebagai proses dari penggunaan metode *data mining* untuk mencari informasi-informasi yang berharga, pola yang ada di dalam data, yang melibatkan algoritma untuk mengidentifikasi pola pada data [16]. Terdapat lima tahap dalam proses KDD yaitu *selection*, *preprocessing*, *transformation*, *data mining*, *interpretation/evaluation* yang dapat dilihat pada Gambar 1 [16, 17].



Gambar 1. Proses KDD
(Sumber: Fayyad et al, 1996)

2.1 Selection

Selection adalah pemilihan atau seleksi data dari sekumpulan data yang ada dan penentuan jenis serta kategori data yang akan digunakan dan dibutuhkan pada penelitian. Karena tidak seluruh data akan dipakai pada proses perhitungan [17, 18]. Atribut yang digunakan dalam penelitian ini menggunakan analisis RFM yang terdiri dari tiga atribut yaitu: *recency*, *frequency* dan *monetary* [5, 6, 7].

- 1) *Recency* (R) adalah atribut yang digunakan untuk menilai pelanggan berdasarkan waktu kebaruan pelanggan melakukan pembelian.
- 2) *Frequency* (F) adalah atribut yang digunakan untuk menilai pelanggan berdasarkan jumlah frekuensi total pelanggan melakukan pembelian.
- 3) *Monetary* (M) adalah atribut yang digunakan untuk menilai pelanggan berdasarkan jumlah total uang yang dihabiskan pelanggan dalam melakukan pembelian.

2.2 Preprocessing

Preprocessing adalah tahapan untuk melakukan pembersihan data (*data cleaning*). Proses pembersihan data meliputi antara lain membuang duplikasi data, *missing value* atau yang bernilai (*null*), memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak [17]. Pembersihan data memastikan data yang akan diolah benar-benar data relevan dan siap dilakukan proses perhitungan *clustering* [18].

2.3 Transformation

Transformation adalah proses perubahan dataset menjadi satu format

yang sama untuk mempermudah proses *data mining* [18]. Transformasi dilakukan untuk meningkatkan akurasi model, salah satunya yaitu dengan melakukan normalisasi [4]. Normalisasi merupakan proses transformasi suatu atribut numerik dengan diskalakan pada jarak yang lebih kecil agar skala data tidak terlalu jauh [4, 19]. Pada penelitian ini proses normalisasi menggunakan Metode Indeks Z. Normalisasi Indeks Z (*zero-Mean normalization*) adalah salah satu metode yang diterapkan untuk normalisasi data, dimana nilai dari suatu atribut dinormalisasi berdasarkan rata-rata dan standar deviasi atribut tersebut [4, 20], dapat dilihat pada persamaan (1).

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

Keterangan:

Z : Nilai Z score

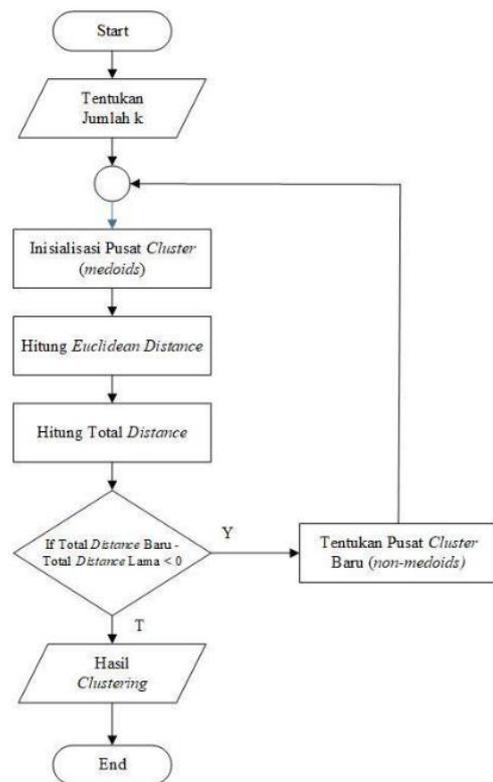
x : Data observasi

μ : Nilai rata-rata atribut

σ : Standar deviasi atribut

2.4 Data mining

Data mining adalah teknik atau proses menemukan pola atau penggalian informasi unik dan menarik dari dataset yang terpilih dengan menggunakan metode atau algoritma tertentu sesuai dengan maksud atau tujuan dari proses KDD secara keseluruhan [17, 18]. Metode yang digunakan pada penelitian ini adalah metode *clustering*, yaitu metode analisis yang bertujuan untuk melakukan segmentasi populasi yang heterogen menjadi sejumlah kelompok yang beranggotakan observasi dengan karakteristik yang homogen [10]. Metode *clustering* dilakukan dengan menerapkan algoritma K-Medoids yaitu algoritma yang menggunakan *medoid* (perwakilan) data sebagai pusat *cluster*. *Medoid* adalah objek data paling sentral di antara titik-titik *cluster*. Oleh karena itu, K objek dipilih secara acak sebagai *medoid* untuk mewakili *cluster*, dan semua objek data ditugaskan ke *cluster* dengan *medoid* terdekat [11]. Proses *Clustering* menggunakan algoritma K-Medoids ditunjukkan oleh flowchart [21] pada Gambar 2.



Gambar 2. Flowchart Algoritma K-Medoids
(Sumber: Andini, A. D., dkk., 2020)

Tahapan Algoritma K-Medoids:

- 1) Menentukan jumlah *cluster* yang akan dibentuk (nilai k).
- 2) Menginisialisasi pusat *cluster* awal (*medoids*) dari masing-masing *cluster* secara acak.
- 3) Menghitung setiap data observasi (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance*.
- 4) Menghitung secara acak objek pada masing-masing *cluster* sebagai kandidat *medoid* baru.
- 5) Menghitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoid* baru.
- 6) Menghitung total simpangan (S) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai *medoid*.
- 7) Mengulangi langkah 4 sampai 6 hingga tidak terjadi perubahan *medoid*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

Penentuan jumlah *cluster* yang akan dibentuk (nilai *k*) pada penelitian ini menggunakan metode *silhouette coefficient* yaitu metode untuk menemukan jumlah *cluster* yang optimal dengan mengkombinasikan nilai *cohesion* dan *separation*. *Cohesion* mengukur seberapa dekat relasi antar satu objek dengan objek lainnya pada sebuah *cluster*, sedangkan *separation* menghitung seberapa jauh sebuah *cluster* berpisah dari *cluster* yang lain [20]. *Silhouette coefficient* terletak pada rentang nilai antara -1 sampai 1. Jika nilai *silhouette coefficient* mendekati nilai 1, maka *cluster* tersebut semakin baik. Sebaliknya jika nilai rata-rata *silhouette coefficient* mendekati -1, maka *cluster* tersebut semakin tidak baik [20, 22].

2.5 Interpretation/Evaluation

Interpretation/Evaluation merupakan tahap terakhir dari proses KDD, yaitu melakukan interpretasi dan evaluasi penerjemahan dan analisa pola-pola atau pengetahuan (*knowledge*) tertentu yang dihasilkan dari perhitungan dalam *data mining*. Proses evaluasi ini dilakukan dengan cara pengujian apakah pola, informasi atau *knowledge* yang diperoleh sesuai dengan fakta atau hipotesa sebelumnya [17]. Pada penelitian ini, pengujian dilakukan dengan menggunakan metode *Davies Bouldin Index* (DBI). DBI adalah salah satu teknik untuk mengevaluasi hasil dari *clustering* dengan memaksimalkan jarak antara objek *cluster* yang satu dengan yang objek data pada *cluster* lainnya dan juga secara bersamaan meminimalkan jarak antara objek dalam sebuah *cluster* [18]. Suatu *cluster* dikatakan memiliki hasil yang optimal atau baik jika memiliki nilai DBI yang minimal [13, 18].

3. HASIL DAN PEMBAHASAN

Tahap pertama penelitian yang dilakukan berdasarkan proses KDD yaitu *selection* atau pemilihan data. Pada penelitian ini, data yang digunakan adalah data transaksi penjualan produk *fashion* XYZ Store untuk triwulan kedua tahun

2021 yang terdiri dari 113 pelanggan. Atribut yang dipilih dari data transaksi penjualan mengacu pada analisis RFM yang terdiri dari tiga atribut yaitu *recency*, *frequency* dan *monetary*. Berikut ini dataset penelitian dapat dilihat pada Tabel 1.

Tabel 1. Dataset Penelitian

Pelanggan	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>
Pelanggan 1	70	2	376.640
Pelanggan 2	60	5	1.259.280
.....			
Pelanggan 112	59	2	479.840
Pelanggan 113	81	1	269.910

Tahap kedua adalah *preprocessing*, yaitu tahapan untuk melakukan pembersihan data atau *data cleaning*. Proses pembersihan data dilakukan dengan menggunakan aplikasi Rstudio. Salah satunya yaitu proses pengecekan *missing value* dilakukan dengan *script* dan hasil yang dapat dilihat pada Gambar 3.

```
> #mengecek missing value  
> anyNA(XYZStore1)  
[1] FALSE  
> |
```

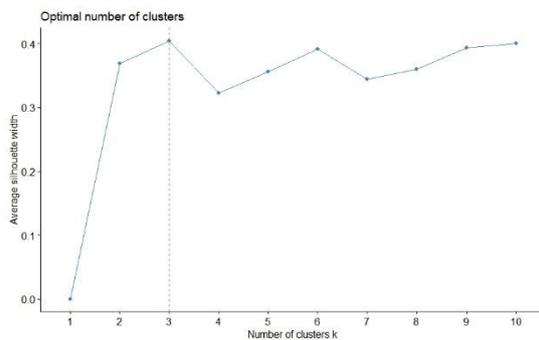
Gambar 3. *Preprocessing* Untuk *Missing Value*

Hasil “FALSE” menunjukkan bahwa tidak ada data yang *missing value*, oleh karena itu data dapat digunakan untuk tahap berikutnya yaitu *transformation*. Pada tahap *transformation* dilakukan proses normalisasi yaitu proses untuk membuat atribut numerik diskalakan pada jarak yang lebih kecil agar skala data tidak terlalu jauh. Normalisasi dilakukan dengan aplikasi Rstudio dengan *script* dan hasil yang dapat dilihat pada Gambar 4.

```
-----  
> #melakukan normalisasi dengan skala Z score  
> XYZStore2=scale(XYZStore1)  
> #melihat data yang sudah di normalisasi  
> XYZStore2  
  
      Recency  Frequency  Monetary  
[1,]  0.5629648352 -0.1310916 -0.32697441  
[2,] -0.0619371921  1.3024580  1.60991424  
[3,]  1.3128472679 -0.6089414 -0.69943443  
[4,] -0.1244273948 -0.1310916 -0.13676152  
[5,]  1.8752590925  0.3467583  0.14368620
```

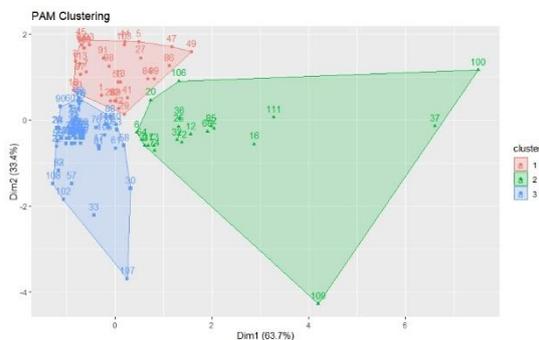
Gambar 4. Hasil Normalisasi

Tahap ketiga adalah proses *data mining*, yaitu teknik untuk menemukan pola atau penggalan informasi dari data dengan menggunakan algoritma. Pada penelitian ini menggunakan algoritma K-Medoids. Langkah pertama pada algoritma K-Medoids adalah menentukan jumlah *cluster* yang akan dibentuk yaitu dengan menggunakan metode *silhouette coefficient*. Hasil visualisasi *silhouette coefficient* dengan menggunakan aplikasi RStudio dapat dilihat pada Gambar 5.

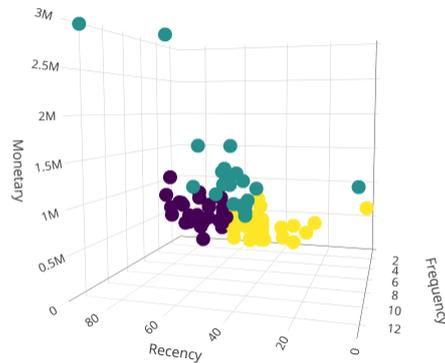


Gambar 5. Hasil Metode *Silhouette*

Berdasarkan Gambar 5 diatas, dapat diperoleh hasil bahwa jumlah *cluster* optimal pada dataset penelitian adalah sebanyak 3 buah *cluster*, nilai $k=3$. Hasil jumlah *cluster* optimal ini digunakan untuk melakukan proses *clustering* pada algoritma K-Medoids. Dengan menggunakan aplikasi RStudio, hasil visualisasi *clustering* tiga buah *cluster* dapat dilihat pada Gambar 6 dan 7.



Gambar 6. Hasil *Cluster 2D*



Gambar 7. Hasil *Cluster 3D*

Berdasarkan hasil *cluster* diatas, dengan menggunakan algoritma K-Medoids pada tiga buah *cluster* optimal diperoleh hasil anggota masing-masing *cluster* pada Tabel 2.

Tabel 2. Hasil Anggota *Cluster*

<i>Cluster</i>	Anggota <i>Cluster</i>
1	31
2	21
3	61

Hasil pada Tabel 2 menunjukkan bahwa segmentasi pelanggan pada XYZ Store terdiri dari tiga jenis pelanggan yaitu segmen 1 terdiri dari 31 pelanggan, segmen 2 terdiri dari 21 pelanggan dan segmen 3 terdiri dari 61 pelanggan. Karakteristik masing-masing segmen pelanggan tersebut dapat diketahui dengan menghitung nilai pada tiga atribut RFM, yaitu nilai *recency* (R), *frequency* (F) dan *monetary* (M). Berikut ini nilai RFM pada masing-masing *cluster*/segmen pelanggan dapat dilihat pada Tabel 3.

Tabel 3. Hasil Segmentasi Pelanggan

<i>Cluster</i>	R	F	M	Jenis Pelanggan
1	79	2	484.652	Typical Customer
2	58	5	1.191.072	Loyal Customer
3	53	1	317.391	Newcomer

Berdasarkan Tabel 3 diatas, dapat diketahui bahwa *cluster* 1 dengan nilai *recency* rendah (lama), *frequency* sedang dan *monetary* sedang dapat dikategorikan sebagai segmen pelanggan *typical customer*, yaitu pelanggan biasa yang melakukan transaksi pembelian sedang. *Cluster* 2 mempunyai nilai *recency* tinggi (baru), *frequency* paling banyak dan *monetary* paling banyak adalah segmen pelanggan *loyal customer*, yaitu pelanggan yang sudah sering dan banyak melakukan transaksi pembelian. *Cluster* 3 mempunyai nilai *recency* paling tinggi (paling baru), *frequency* dan *monetary* paling rendah adalah segmen pelanggan *newcomer* atau pendatang baru yang melakukan transaksi pembelian.

Evaluasi hasil *cluster* sebagai tahapan terakhir proses KDD adalah dengan menggunakan metode *Davies-Bouldin Index* (DBI). DBI adalah teknik untuk mengevaluasi hasil dari *clustering* dengan memaksimalkan jarak antara objek *cluster* yang satu dengan objek data pada *cluster* lainnya dan meminimalkan jarak antara objek dalam sebuah *cluster*. Metode DBI dilakukan dengan aplikasi Rstudio dengan *script* dan hasil yang dapat dilihat pada Gambar 8.

```
> #evaluasi klaster dengan DBI
> library(clustersim)
> (index.DB (XYZStore2,
+          pam.res$clustering,
+          centrotypes='medoids',
+          d=dist(XYZStore2)))
$DB
[1] 1.536369
```

Gambar 8. Hasil DBI

Berdasarkan Gambar 8, diperoleh nilai DBI sebesar 1,536. Hasil *clustering* yang optimal diperoleh jika memiliki nilai DBI minimal, yaitu nilai yang mendekati 0. Nilai DBI yang dihasilkan menunjukkan angka yang mendekati 0. Sehingga dapat disimpulkan bahwa hasil *clustering* dengan tiga buah *cluster* memiliki kemiripan nilai data yang dekat untuk anggota setiap *cluster*

dan berbeda jauh dengan anggota *cluster* lain.

Dengan diperolehnya tiga *cluster* atau segmen pelanggan, maka perusahaan dapat membuat kebijakan dan strategi bisnis dalam melakukan pengelolaan hubungan dengan pelanggan. Segmen *loyal customer* adalah segmen yang paling banyak memberikan keuntungan pada perusahaan maka perusahaan harus menjalin hubungan yang baik dan memberikan pelayanan yang lebih. Perusahaan dapat memberikan *reward*, mendapat *privilege* sebagai *customer* prioritas dalam hal program penjualan maupun termin pembayaran dan dapat dilibatkan dalam berbagai kegiatan. Segmen *typical customer* adalah pelanggan biasa yang perlu mendapat berbagai promosi sehingga dapat meningkatkan transaksi pembeliannya. Segmen *newcomer* adalah pendatang baru atau pelanggan yang baru bertransaksi yang mempunyai potensi untuk menjadi *loyal customer*. Segmen ini perlu mendapat perhatian terutama dalam hal pelayanan dan kemudahan transaksi terutama pada *first impression*.

4. KESIMPULAN

Penelitian ini bertujuan untuk menentukan segmentasi pelanggan pada XYZ Store dengan menggunakan algoritma K-Medoids dan analisis RFM (*Recency*, *Frequency*, *Monetary*). Penentuan jumlah *cluster* optimal menggunakan metode *silhouette coefficient* dan diperoleh hasil tiga buah *cluster*. Evaluasi hasil *cluster* menggunakan metode *Davies-Bouldin Index* (DBI) dan diperoleh hasil nilai DBI sebesar 1,536 yang menunjukkan bahwa tiga buah *cluster* memiliki kemiripan nilai data yang dekat untuk anggota setiap *cluster* dan berbeda jauh dengan anggota *cluster* lain. Hasil segmentasi pelanggan diperoleh tiga jenis pelanggan yaitu *loyal customer* sebanyak 21 pelanggan, *typical customer* sebanyak 31 pelanggan dan *newcomer* sebanyak 61 pelanggan. *Loyal customer*

adalah pelanggan yang mempunyai nilai *recency* tinggi (baru), *frequency* paling banyak dan *monetary* paling banyak. *Typical customer* adalah pelanggan yang memiliki *recency* rendah (lama), *frequency* sedang dan *monetary* sedang. *Newcomer* adalah pelanggan yang mempunyai nilai *recency* paling tinggi (paling baru), *frequency* dan *monetary* paling rendah. Dengan diperolehnya tiga segmen pelanggan, maka perusahaan dapat membuat kebijakan dan strategi bisnis dalam melakukan pengelolaan hubungan dengan pelanggan sehingga diharapkan keuntungan perusahaan akan meningkat.

5. SARAN

Penelitian ini menerapkan metode *clustering* dengan menggunakan algoritma K-Medoids, untuk penelitian selanjutnya dapat dilakukan dengan algoritma lain misalnya algoritma K-Means, Agglomerative, Divisive dan lainnya, atau dapat pula dilakukan dengan membandingkan dua algoritma. Dalam menentukan *cluster* optimal, penelitian ini menggunakan metode *silhouette coefficient*, untuk penelitian selanjutnya dapat dilakukan dengan metode lain seperti metode *elbow* atau membandingkan kedua metode tersebut. Selain itu, penelitian ini menggunakan aplikasi RStudio untuk melakukan proses *clustering* dan visualisasinya. Untuk penelitian selanjutnya dapat menggunakan aplikasi *data mining* lain seperti Rapid Miner, WEKA, Orange, dan lain sebagainya.

REFERENCE

- [1] Fadly, M., & Wantoro, A. (2019, August). Model Sistem Informasi Manajemen Hubungan Pelanggan Dengan Kombinasi Pengelolaan Digital Asset Untuk Meningkatkan Jumlah Pelanggan. In *Prosiding Seminar Nasional Darmajaya* (Vol. 1, pp. 46-55).
- [2] Putri, F. I. W., & Suriyanto, M. A. (2022). Strategi Mengelola Hubungan Pelanggan Pada Kajian Business To Business PT Wakabe Indonesia. *Jurnal Indonesia Sosial Sains*, 3(01), 44-56.
- [3] Muningsih, E., & Kiswati, S. (2018). Sistem aplikasi berbasis optimasi metode elbow untuk penentuan *clustering* pelanggan. *Joutica: Journal of Informatic Unisla*, 3(1), 117-124.
- [4] Adiana, B. E., Soesanti, I., & Permanasari, A. E. (2018). Analisis segmentasi pelanggan menggunakan kombinasi RFM model dan teknik *clustering*. *Jurnal Terapan Teknologi Informasi*, 2(1), 23-32
- [5] Shihab, S. H., Afroge, S., & Mishu, S. Z. (2019, February). RFM based market segmentation approach using advanced k-means and Agglomerative *clustering*: a comparative study. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-4). IEEE.
- [6] Shirole, R., Salokhe, L., & Jadhav, S. (2021). Customer Segmentation using RFM Model and K-Means *Clustering*.
- [7] Parikh, Y., & Abdelfattah, E. (2020, October). *Clustering* algorithms and RFM analysis performed on retail transactions. In 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (pp. 0506-0511). IEEE.
- [8] Hung, P. D., Lien, N. T. T., & Ngoc, N. D. (2019, March). Customer segmentation using hierarchical agglomerative *clustering*. In Proceedings of the 2019 2nd International Conference on Information Science and Systems (pp. 33-37).
- [9] Sheshasaayee, A., & Logeshwari, L. (2018, May). Implementation of *clustering* technique based RFM analysis for customer behaviour in online transactions. In 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1166-1170). IEEE.

- [10] Vercellis, C. (2009). Business intelligence: *data mining* and optimization for decision making. John Wiley & Sons.
- [11] Mousavi, S. H. A. H. L. A., Boroujeni, F. Z., & Aryanmehr, S. A. E. E. D. (2020). Improving customer clustering by optimal selection of cluster centroids in k-means and k-medoids algorithms. *Journal of Theoretical and Applied Information Technology*, 98(18).
- [12] Atmaja, E. H. S. (2019). Implementation of k-Medoids Clustering Algorithm to Cluster Crime Patterns in Yogyakarta. *International Journal of Applied Sciences and Smart Technologies*, 1(1), 33-44.
- [13] Sulistyawati, A. A. D., & Sadikin, M. (2021). Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan. *Sistemasi: Jurnal Sistem Informasi*, 10(3), 516-526.
- [14] Madani, A., Rahmah, A. R. A., Nurunnisa, F. N. F., & Elia, A. E. A. (2022, September). Segmentasi Pelanggan pada BC HNI 2 Pekanbaru dengan Menerapkan Algoritma K-Medoids dan Model RFM (Recency, Frequency, Monetary): Customer Segmentation at BC HNI 2 Pekanbaru by Applying the K-Medoids Algorithm and Recency, Frequency, Monetary (RFM) Model. In *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat* (Vol. 1, No. 1, pp. 179-186).
- [15] Purnama, G., Pudjiantoro, T. H., & Sabrina, P. N. (2021, October). *Data mining* SEGMENTASI PELANGGAN MENGGUNAKAN K-MEDOIDS BERDASARKAN MODEL LENGTH, RECENCY, FREQUENCY, MONETARY (LRFM). In *SNIA (Seminar Nasional Informatika dan Aplikasinya)* (Vol. 5, pp. B29-34).
- [16] Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). Advances in Knowledge Discovery and *Data mining*. Menlo Park, CA: AAAI Press.
- [17] Marlina, D., Putri, N. F., Fernando, A., & Ramadhan, A. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak. *J. CoreIT*, 4(2), 64.
- [18] Supriyadi, A., Triayudi, A., & Sholihati, I. D. (2021). Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 6(2), 229-240.
- [19] Sutresno, S. A., Iriani, A., & Sedyono, E. (2018). Metode K-Means Clustering dengan Atribut RFM untuk Mempertahankan Pelanggan. *Jurnal Teknik Informatika Dan Sistem Informasi*, 4(3), 433-440
- [20] Rahmah, M., Candra, A., & Sembiring, R. W. (2022). Identifikasi Predikat Hasil Pengelompokan Data Kualitas Udara dengan Menggunakan Affinity Propagation dan Silhouette Coefficient. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, 6(2), 177-180.
- [21] Andini, A. D., & Arifin, T. (2020). Implementasi algoritma k-medoids untuk klasterisasi data penyakit pasien di rsud kota bandung. *Jurnal Responsif: Riset Sains dan Informatika*, 2(2), 128-138.
- [22] Farissa, R. A., Mayasari, R., & Umaidah, Y. (2021). Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung. *Journal of Applied Informatics and Computing (JAIC)*, 5(2), 109-116