

Penerapan Algoritma K-Medoids untuk Menentukan Segmentasi Pelanggan

Anggi Ayu Dwi Sulistyawati*, Mujiono Sadikin

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana
Jl. Meruya Selatan No. 1 Kembangan Jakarta Barat 11650, Indonesia

*e-mail: 41518120117@student.mercubuana.ac.id

(received: 23 Februari 2021, revised: 12 April 2021, accepted: 7 September 2021)

Abstrak

Strategi pemasaran berorientasi pelanggan memiliki peranan penting dalam mengelola hubungan baik dengan pelanggan. Agar strategi pemasaran tepat sasaran, segmentasi pelanggan dapat digunakan untuk mengelompokkan pelanggan berdasarkan karakteristik yang sama. Dalam penyusunan strategi pemasaran dapat memanfaatkan TI di bidang komputasi, salah satunya adalah *data mining*. Pemanfaatan teknologi komputasi untuk pengolahan data yang belum maksimal mengakibatkan penumpukan data yang miskin informasi. Pada penelitian ini dilakukan penerapan teknik *clustering* dengan menggunakan algoritma K-Medoids pada dataset transaksi penjualan untuk menentukan segmentasi pelanggan. Penyusunan strategi pemasaran ditentukan berdasarkan tipe dan karakteristik pelanggan pada setiap *cluster* atau segmen pelanggan yang terbentuk. Uji validitas *cluster* menggunakan Silhouette Index dan Davies Boulbin Index dilakukan untuk menentukan jumlah *cluster* yang paling optimal. Hasil penelitian ini menunjukkan bahwa jumlah *cluster* optimal adalah 3 (tiga) *cluster* dengan nilai maksimum Silhouette Index adalah 0,375 dan nilai minimum Davies Doulbin Index adalah 1,030. Segmen pelanggan hasil penelitian adalah *lost customer*, *core customer*, dan *new customer*.

Kata kunci: algoritma k-medoids, *clustering*, data mining, segmentasi pelanggan, strategi pemasaran

Abstract

Customer-oriented marketing strategies play an important role in managing good relationships with customers. To keep marketing strategies on target, customer segmentation can be used to group customers based on the same characteristics. In the preparation of marketing strategies can utilize IT in the field of computing, one of which is data mining. The utilization of computing technology for data processing that has not been maximized resulted in a poor accumulation of information data. In this study, the application of clustering techniques using the K-Medoids algorithm on sales transaction dataset to determine customer segmentation. The preparation of a marketing strategy is determined based on the characteristics and types of customers in each cluster or segment of customers formed. cluster validity tests using the Silhouette Index and Davies-Boulbin Index are performed to determine the most optimal number of clusters. The results of this study showed that the optimal number of clusters is 3 (three) clusters with a maximum silhouette index value of 0.375 and the minimum value of the davies-bouldin index is 1.030. The customer segments of the research results are lost customers, core customers, and new customers.

Keywords: *k-medoids algorithm, clustering, data mining, customer segmentation, marketing strategy*

1 Pendahuluan

Strategi pemasaran memiliki peran penting dalam persaingan bisnis antar perusahaan. Selain mengutamakan strategi pemasaran berorientasi produk, perusahaan juga harus mampu mengutamakan strategi pemasaran berorientasi pelanggan. Hal ini berguna untuk mengelola hubungan yang baik dengan pelanggan sehingga kepuasan pelanggan tercapai dan mendapatkan loyalitas [1]. Dalam penyusunan strategi pemasaran berorientasi pelanggan, segmentasi pelanggan dapat dimanfaatkan untuk mengidentifikasi karakteristik pelanggan dan mengelompokkan pelanggan berdasarkan perilaku atau kebiasaan yang sama [2], [3].

Perusahaan juga dapat memanfaatkan TI di bidang komputasi dalam melakukan penyusunan strategi pemasaran. Perum BULOG yang saat ini memiliki kebijakan fokus pada penguatan peran komersial harus mampu bersaing dengan perusahaan lainnya dalam perdagangan komoditi pangan. Perusahaan yang bergerak di bidang logistik ini belum memanfaatkan teknologi komputasi untuk pengolahan data secara maksimal. Data transaksi penjualan hanya diolah sebagai laporan penjualan harian, mingguan dan bulanan. Hal tersebut dapat mengakibatkan penumpukan data yang miskin informasi [4]. Data transaksi tersebut memiliki potensi untuk digunakan dalam menentukan strategi pemasaran, salah satunya segmentasi pelanggan. Segmentasi pelanggan perlu dilakukan untuk mengelompokkan pelanggan yang memiliki kesamaan karakteristik [1]. Selama ini, data transaksi penjualan hanya digunakan untuk melihat *volume* pembelian dan *omzet* dalam setahun untuk mengetahui kriteria pelanggan yang mendapatkan *reward*. Pengolahan data transaksi penjualan dengan komputasi tepat dapat memberikan manfaat bagi perusahaan agar pemasaran lebih efektif dan efisien. Dengan memanfaatkan *data mining*, data dapat diolah untuk mengetahui pola dan dapat diidentifikasi informasi yang tersembunyi dalam data tersebut [5].

Salah satu metode pada *data mining* adalah *clustering* yang mana metode ini berguna untuk mengelompokkan data [6]. Teknik *clustering* dapat digunakan dalam menentukan segmentasi pelanggan dengan menganalisis kelompok data untuk mengetahui karakteristik dari kelompok pelanggan yang terbentuk [7]. Algoritma K-Medoids merupakan salah satu algoritma teknik *clustering* yang dapat digunakan untuk mengelompokkan objek-objek kedalam *cluster - cluster* dengan objek yang *similar* atau sama. Kelebihan dari algoritma ini yaitu tidak sensitif terhadap outlier, dapat mengurangi *noise*, dan jika dibandingkan dengan algoritma K-Means, K-Medoids lebih unggul dalam melakukan klasterisasi dataset heterogen/campuran, pemilihan *cluster*, kompleksitas antar ruang *cluster*, dan waktu eksekusi [8]–[10].

Pada penelitian ini, dilakukan penerapan algoritma K-Medoids pada dataset transaksi penjualan untuk menentukan segmentasi pelanggan. Tujuan dari penelitian ini adalah untuk mengetahui tipe dan karakteristik pelanggan untuk memberikan rekomendasi strategi pemasaran untuk perusahaan. Strategi pemasaran lebih efektif dan tepat sasaran apabila disusun berdasarkan dari tipe dan karakteristik pelanggan dari hasil segmentasi pelanggan yang terbentuk.

2 Tinjauan Literatur

Berbagai penelitian yang telah dilakukan terkait dengan segmentasi pelanggan maupun non segmentasi pelanggan menggunakan teknik *clustering*. Berbagai dataset juga digunakan untuk menambang informasi nontrivial yang terkandung dalam data. Pada bagian ini disajikan beberapa tinjauan literatur yang isi penelitiannya dijadikan sebagai acuan dalam penelitian ini.

R.W.S Berahmana d.k.k [2], menyajikan penelitian yang menggunakan teknik *clustering* untuk menentukan segmentasi pelanggan. Dalam penelitian ini, penulis menggunakan algoritma K-Means, K-Medoids, dan DBSCAN yang diterapkan pada dataset transaksi tahunan dari bulan tahun 2013 sampai dengan 2018. Hasil penelitian menunjukkan DBSCAN dalam segmentasi pelanggan menghasilkan *cluster* lebih baik daripada K-Means dan K-Medoids, sedangkan uji validitas menggunakan Silhouette Index dan Davies Bouldin Index menunjukkan K-Means menghasilkan validitas yang baik daripada K-Medoids dan DBSCAN. Teknik *clustering* lain untuk pendukung keputusan strategi pemasaran juga disajikan oleh Triyansyah dkk [11]. Pada penelitian ini, penulis menggunakan K-Means *clustering* yang diterapkan pada dataset transaksi penjualan online shop Hoyweapstore dari bulan Juli 2017 sampai dengan Februari 2018. Sebagai hasil penelitian, penulis merekomendasikan tiga produk terlaris.

Teknik *clustering* digunakan untuk menganalisis segmentasi pelanggan juga disajikan oleh Christy dkk [3]. Penelitian ini melakukan analisis segmentasi pelanggan dengan mengelompokkan pelanggan berdasarkan perilaku atau kebiasaan yang sama yang diterapkan pada dataset ritel online. Proses segmentasi dilakukan menggunakan analisis *Recency*, *Frequency*, dan *Monetary* (RFM) dan dilanjutkan dengan proses *clustering* menggunakan algoritma K-Means, Fuzzy C-Means dan RM K-Means. Hasil penelitian menunjukkan model RFM dapat digunakan untuk segmentasi pelanggan dan algoritma RM K-Means mempunyai waktu eksekusi lebih kecil dan menurunkan jumlah iterasi.

Sedangkan penelitian terkait non segmentasi pelanggan menggunakan teknik *clustering* khususnya algoritma K-Medoids disajikan oleh Arora dkk [9]. Pada penelitian ini dilakukan

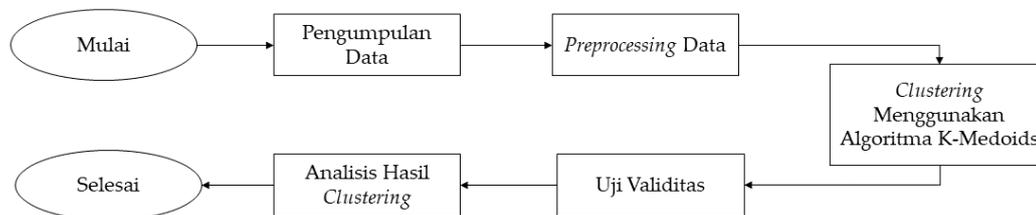
eksplorasi kelebihan dan batasan teknik *clustering* dengan menggunakan algoritma K-Means dan K-Medoids. Algoritma tersebut diujikan pada dataset *transaction10k of KEEL*. Hasil penelitian menunjukkan bahwa K-Medoids jauh lebih baik dibandingkan K-Means seperti tidak sensitif terhadap *outlier*, dapat mengurangi *noise*, waktu eksekusi yang diperlukan saat pemilihan *cluster* dan kompleksitas antar ruang cluster. Harikumar dkk [10] dalam penelitian terkait teknik *clustering* menyajikan studi dengan membandingkan algoritma K-Medoids dan K-Means untuk mengelompokkan dataset heterogen/ campuran. Algoritma tersebut diterapkan pada dataset Bank yang mempunyai tipe data yang bervariasi. Hasil perbandingan dari kedua algoritma tersebut menunjukkan bahwa K-Medoids lebih unggul dari K-Means dalam melakukan klasterisasi dataset heterogen/campuran.

Penelitian lain terkait dengan teknik *clustering* menggunakan algoritma K-Medoids juga dilakukan oleh Andini dkk [12] dan Marlina dkk [13]. Semua penelitian tersebut menggunakan algoritma K-Medoids untuk melakukan proses *clustering*. Dataset yang digunakan pada penelitian [12] adalah dataset penyakit pasien di RSUD Kota Bandung pada tahun 2019, sedangkan pada penelitian [13] adalah dataset sebaran anak cacat di provinsi Riau. Hasil kedua penelitian menunjukkan bahwa Algoritma K-Medoids bekerja dengan baik karena setiap objek pada setiap *cluster* memiliki mutu yang baik, dimana setiap objek telah dikelompokkan sesuai dengan tingkat kemiripan yang tinggi serta K-Medoids lebih baik dalam melakukan pengelompokan data dibandingkan dengan algoritma K-Means berdasarkan nilai validitasnya.

Dari berbagai literatur yang direview di atas, dari beberapa metode *clustering* untuk segmentasi pelanggan K-Medoids memberikan hasil yang lebih baik. Dengan dasar ini, untuk kepentingan segmentasi pelanggan Perum Bulog dipilih K-Medoids. Teknik K-Medoids diterapkan pada dataset transaksi penjualan kantor wilayah DKI Jakarta & Banten yang dikumpulkan dari tanggal 1 Januari 2019 sampai dengan 31 Desember 2019. Dataset transaksi penjualan dilakukan analisis untuk menentukan segmentasi pelanggan yang bertujuan untuk menyusun rekomendasi strategi pemasaran berdasarkan karakteristik dan tipe pelanggan. Untuk mengetahui karakteristik pelanggan dilakukan analisis berdasarkan aspek geografis, demografi dan perilaku.

3 Metode Penelitian

Metode penelitian berupa alur tahapan yang menjadi acuan untuk menentukan segmentasi pelanggan. Skema alur tahapan yang digunakan pada penelitian ini ditunjukkan pada Gambar 1. Penelitian ini menggunakan bahasa pemrograman python dengan *tool Google Colaboratory* untuk pengolahan data.



Gambar 1. Skema Alur Tahapan Penelitian

3.1 Pengumpulan Data

Pada tahap pengumpulan data dilakukan dengan melakukan wawancara dan observasi. Hasil dari wawancara dan observasi, diketahui bahwa data yang dimiliki perusahaan hanya diolah sebagai laporan penjualan dan belum dimanfaatkan untuk kepentingan strategi pemasaran dalam hal ini yaitu segmentasi pelanggan. Data yang digunakan untuk segmentasi pelanggan yaitu data transaksi penjualan Perum BULOG kantor wilayah DKI Jakarta & Banten pada periode tahun 2019 sebanyak 9.945 instan.

3.2 Pre-Processing Data

Pada tahapan preprocessing data dilakukan pembersihan data, pemilihan data, dan transformasi data. Tahap pembersihan pada data awal untuk memastikan tidak ada data yang tidak lengkap atau *null* dan *redundant* data serta mempersiapkan data agar data yang siap diolah sesuai dengan kebutuhan [4],[5],[7]. Tahapan pembersihan data dimaksudkan untuk menghilangkan nilai kosong atau tidak lengkap. Tahapan pemilihan data dilakukan pemilihan atribut-atribut, dari jumlah atribut

awal sebanyak 15 atribut dipilih 7 atribut (dengan simbol \checkmark) yang digunakan untuk keperluan pengolahan data. Pemilihan atribut-atribut yang digunakan disajikan pada Tabel 1. Data yang siap untuk dilakukan pengolahan data memiliki dimensi 9.945 baris x 7 kolom.

Tabel 1. Pemilihan Atribut

| Atribut | Detail Penggunaan | Atribut | Detail Penggunaan |
|----------------|-------------------|-------------------|-------------------|
| Tanggal SPS | x | Merk Komoditi | \checkmark |
| Tanggal Bayar | \checkmark | Kemasan | x |
| No. SPS | \checkmark | Satuan | x |
| Saluran | \checkmark | Kolli Terjual | x |
| ID Pembeli | \checkmark | Kuantum Terjual | x |
| KAB/ Kota | \checkmark | Harga per Kemasan | x |
| Nama Komoditi | x | Nominal Penjualan | \checkmark |
| Jenis Komoditi | x | | |

Tahapan transformasi data yaitu tahap untuk mengubah data agar sesuai dengan format untuk diolah dalam proses data mining [14]. Pengubahan data ini dilakukan dengan proses konversi data yang masih berupa label menjadi numerik dikarenakan pada komputasi *clustering* hanya dapat melakukan pengolahan data yang bersifat numerik [7]. Transformasi data juga dilakukan pada data transaksi penjualan dengan menggunakan agregasi atau meringkas data yang dikelompokkan berdasarkan model *recency*, *frequency*, dan *monetary* (RFM)[15]. Model RFM adalah teknik pemasaran yang digunakan untuk mempelajari nilai dan perilaku pelanggan [16],[17]. Pengelompokan perilaku yang serupa berdasarkan nilai RFM pelanggan dapat membantu dalam menentukan segmentasi pelanggan [3]. Model RFM memiliki 3 (tiga) variabel, yaitu [14]: *recency* (R) adalah rentang waktu (hari) pelanggan melakukan transaksi terakhir sampai saat ini, *frequency* (F) adalah jumlah total transaksi yang dilakukan pelanggan dalam satu periode, dan *monetary* (M) adalah jumlah total nilai pembelian yang dilakukan pelanggan dalam satu periode.

Setelah data ditransformasi sesuai kebutuhan, dilakukan normalisasi data menggunakan normalisasi Min-Max agar data memiliki *range* nilai antara 0 sampai 1 sehingga tidak memiliki nilai yang terlalu jauh [18]. Perhitungan normalisasi Min-Max dapat dilihat pada rumus persamaan (1). Dimana, x' adalah nilai hasil normalisasi, x adalah nilai data aktual yang dinormalisasi, x_{max} adalah nilai maksimum data aktual dan x_{min} adalah nilai minimum data aktual.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

3.3 Clustering Menggunakan Algoritma K-Medoids

Algoritma K-Medoids atau sering dikenal sebagai PAM (*Partitioning Around Medoids*), merupakan algoritma yang menerapkan objek sebagai perwakilan di setiap *cluster* yaitu medoid[2]. Algoritma K-Medoids dapat meminimalkan jumlah perbedaan antara titik data dalam *cluster* dengan titik data terpilih di setiap *cluster* sebagai pusat (medoid) [15], [19]. K-Medoids memiliki karakteristik dimana pusat *cluster* berada di antara titik-titik datanya [20]. Penggunaan Euclidean Distance untuk menghitung jarak antara kedekatan objek dengan pusat (medoid) sehingga kemungkinan besar objek yang dipilih secara acak mirip dengan objek medoidnya [21]. Tahapan penyelesaian K-Medoids adalah sebagai berikut [8]:

- Melakukan inisialisasi pusat *cluster* sebanyak jumlah *cluster* (k).
- Mengalokasikan setiap objek ke *cluster* terdekat menggunakan Euclidean Distance dengan rumus persamaan (2).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

- Memilih objek secara acak pada masing-masing *cluster* sebagai calon medoid yang baru.
- Menghitung jarak setiap objek pada masing-masing *cluster* dengan calon medoid yang baru.

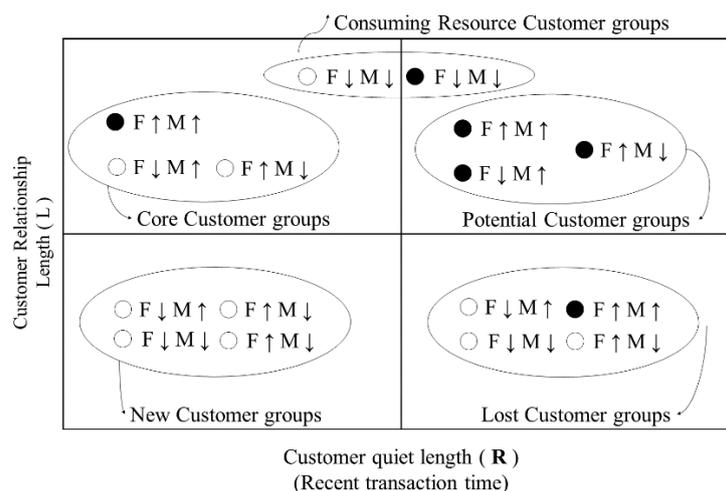
- e. Total simpangan (S) didapatkan dengan menghitung nilai total jarak yang baru dikurangi total jarak yang lama. Jika nilai S kurang dari 0, maka dilakukan penukaran objek dengan data *cluster* untuk membentuk sekumpulan jumlah *cluster* objek yang baru sebagai medoid.
- f. Mengulangi tahap c sampai dengan e apabila masih terjadi perubahan medoid, jika tidak ada perubahan maka diperoleh *cluster* serta anggota dari *cluster* masing-masing.

3.4 Uji Validitas

Uji validitas bertujuan untuk mengevaluasi hasil proses *clustering* untuk menentukan nilai k (jumlah cluster) yang paling optimal [18]. Pada penelitian ini, validasi *cluster* menggunakan Silhouette Index (SI) dan Davies Doulbin Index (DBI). SI adalah teknik validasi untuk mengukur seberapa dekat setiap objek dalam satu *cluster* dan objek di *cluster* lain serta mengetahui seberapa baik objek terletak dalam clusternya [22]. Nilai SI berada di antara -1 hingga +1 dan hasil nilai maksimum menunjukkan jumlah *cluster* yang optimal [22]. Sedangkan, DBI adalah salah satu teknik untuk mengevaluasi hasil dari *clustering* dengan memaksimalkan jarak antar cluster, dimana jumlah *cluster* yang optimal memiliki nilai DBI yang minimum [23].

3.5 Analisis Hasil Clustering

Dilakukan analisis hasil *clustering* terhadap *cluster* atau segmen yang terbentuk untuk mengetahui tipe dan karakteristik pelanggan. Untuk mengetahui tipe pelanggan dapat dilakukan pemetaan pelanggan sesuai dengan penelitian [17] yang berdasarkan matriks loyalitas pelanggan dari penelitian Dai dan Tseng [24], yang ditunjukkan pada Gambar 2. Matrik tersebut dapat mengklasifikasikan perilaku pelanggan berdasarkan segmennya dengan menggunakan simbol (↑) untuk merepresentasikan nilai rata-rata *cluster* yang terbentuk lebih tinggi dari nilai rata-rata keseluruhan dan simbol (↓) untuk merepresentasikan nilai rata-rata *cluster* yang terbentuk lebih rendah dari nilai rata-rata keseluruhan [18]. Matrik tersebut mengelompokkan objek berdasarkan nilai LRFM, sedangkan pada penelitian ini hanya menggunakan nilai RFM. Sedangkan, karakteristik pelanggan dapat dilihat dari aspek geografis, demografi dan perilaku. Selanjutnya, dari setiap tipe dan karakteristik pelanggan per segmen dapat dilakukan penyusunan rekomendasi strategi pemasaran yang tepat.



Gambar 2.. Matrik Loyalitas Pelanggan

4 Hasil dan Pembahasan

Dataset transaksi penjualan sebanyak 9.945 instan yang telah dikumpulkan dilakukan *preprocessing data* dengan membersihkan data *null* atau tidak lengkap. Hasil eksperimen menunjukkan data transaksi penjualan tidak terdapat nilai kosong. Pada tahap pemilihan data dilakukan pemilihan atribut-atribut untuk menghasilkan data yang siap untuk dilakukan pengolahan data. Data transaksi penjualan yang diolah memiliki dimensi 9.945 baris x 7 kolom. Tahap transformasi data dilakukan untuk mengubah data agar sesuai dengan format untuk diolah dalam proses data mining. Transformasi data diawali dengan proses konversi data yang masih berupa kategori label menjadi kode numerik dilakukan pada atribut *cab_kota* dan saluran. Selanjutnya,

dilakukan agregasi atau meringkas data transaksi penjualan menjadi model RFM. Hasil dari tahapan transformasi data transaksi penjualan disajikan pada Tabel 2 dan diketahui sebanyak 906 pelanggan yang melakukan transaksi di tahun 2019. Dari hasil transformasi data dilakukan normalisasi data pada atribut *recency*, *frequency*, *monetary*, *kode_kab_kota* dan *kode_saluran*. Pada penelitian ini menggunakan normalisasi Min-Max scaler dari sklearn. Hasil normalisasi data merupakan hasil akhir dari tahapan *preprocessing* data yang disajikan pada Tabel 3.

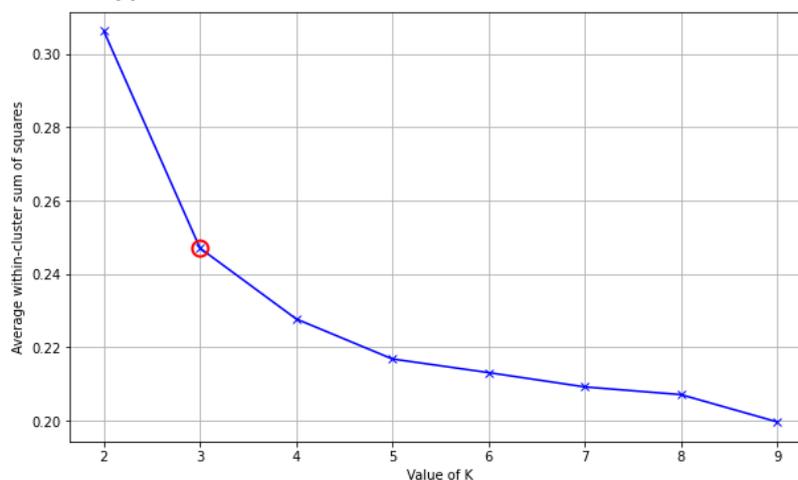
Tabel 2.. Hasil Transformasi Data

| No | id_pelanggan | <i>recency</i> | <i>frequency</i> | <i>monetary</i> | kode_kab_kota | kode_saluran |
|------|--------------|----------------|------------------|-----------------|---------------|--------------|
| 1 | 490 | 1 | 26 | 63.333.750 | 16 | 6 |
| 2 | 1025 | 11 | 9 | 6.721.600 | 15 | 6 |
| 3 | 1027 | 4 | 19 | 228.761.000 | 9 | 6 |
| 4 | 1030 | 176 | 10 | 5.660.500 | 14 | 6 |
| 5 | 1033 | 111 | 7 | 1.421.100 | 14 | 6 |
| | | | | | | |
| 906 | 118562 | 0 | 1 | 5.522.000 | 8 | 6 |

Tabel 3.. Hasil Normalisasi Data

| <i>recency</i> | <i>frequency</i> | <i>monetary</i> | kode_kab_kota | kode_saluran |
|----------------|------------------|-----------------|---------------|--------------|
| 0,003 | 0,030 | 0,002 | 0,615 | 0,667 |
| 0,030 | 0,010 | 0,000 | 0,577 | 0,667 |
| 0,011 | 0,022 | 0,007 | 0,346 | 0,667 |
| 0,485 | 0,011 | 0,000 | 0,538 | 0,667 |
| 0,306 | 0,007 | 0,000 | 0,538 | 0,667 |
| ... | ... | ... | ... | ... |
| 0,000 | 0,000 | 0,000 | 0,308 | 0,667 |

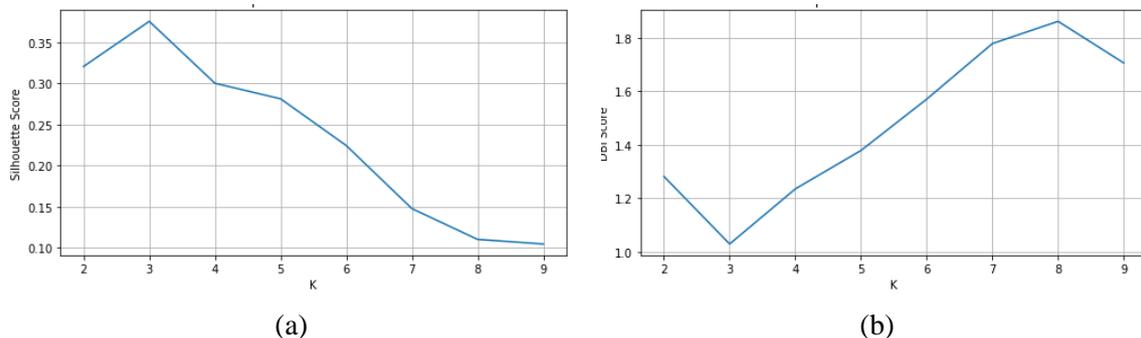
Pada tahapan *clustering* dengan algoritma K-Medoids diterapkan pada data hasil akhir tahapan *preprocessing* data. Untuk melakukan inisialisasi pusat *cluster* sebanyak jumlah *cluster* (k) menggunakan metode Elbow. Metode Elbow digunakan untuk menentukan jumlah *cluster* (k), dimulai dari k=2 dan seterusnya [25]. Hasil grafik nilai k dengan Metode Elbow yang ditunjukkan pada Gambar 3, menunjukkan bahwa jumlah *cluster* (k) yang optimal adalah 3 (tiga). Pada grafik yang terlihat pada gambar tersebut, terlihat jumlah *cluster* 4 dan seterusnya tidak memberikan perubahan nilai yang signifikan. Proses *clustering* dengan algoritma K-Medoids menghasilkan 3 (tiga) *cluster* yaitu *cluster* 0 memiliki anggota *cluster* sebanyak 271, *cluster* 1 sebanyak 307 anggota, dan *cluster* 2 sebanyak 328 anggota.



Gambar 3. Grafik Nilai K Menggunakan Metode Elbow

Uji validitas dilakukan untuk mengevaluasi hasil nilai k dari tahapan *clustering* menggunakan algoritma K-Medoids apakah nilai k merupakan jumlah *cluster* yang paling optimal. Pengujian

dilakukan pada *cluster* dengan nilai $k = 2$ sampai dengan nilai $k = 9$. Grafik hasil dari uji validitas *cluster* dengan menggunakan SI dan DBI ditunjukkan pada Gambar 4. Dari Tabel 4 yang disajikan, didapatkan nilai SI maksimum yaitu 0,375 dan nilai DBI minimum yaitu 1,030 menunjukkan bahwa $k = 3$ merupakan jumlah *cluster* yang optimal.



Gambar 4. (a) Grafik Uji Validitas SI dan (b) Grafik Uji Validitas DBI

Tabel 4. Validasi *cluster* Optimal

| Jumlah <i>cluster</i> (k) | SI | DBI |
|---------------------------|-------|-------|
| 2 | 0,321 | 1,282 |
| 3 | 0,375 | 1,030 |
| 4 | 0,300 | 1,235 |
| 5 | 0,281 | 1,379 |
| 6 | 0,224 | 1,571 |
| 7 | 0,148 | 1,778 |
| 8 | 0,110 | 1,861 |
| 9 | 0,105 | 1,706 |

Pada tahapan analisis hasil *clustering* bertujuan untuk mengetahui karakteristik dan tipe pelanggan per segmen pelanggan berdasarkan jumlah *cluster* yang terbentuk. Hasil analisis dapat dijadikan acuan untuk menyusun rekomendasi strategi pemasaran kepada perusahaan. Hasil *cluster* yang terbentuk secara optimal dari proses *clustering* adalah 3 (tiga) *cluster* atau segmen. Dalam melakukan pemetaan pelanggan untuk mengetahui tipe pelanggan, perlu dilakukan perhitungan nilai rata-rata keseluruhan pada atribut RFM yang disajikan pada Tabel 5. Didapatkan jumlah pelanggan sebanyak 906 dengan rata-rata nilai *recency* adalah 87 hari, nilai *frequency* adalah 11 kali transaksi dan nilai *monetary* adalah Rp129.239.476,00.

Tabel 5. Nilai RFM Rata-Rata Keseluruhan

| Jumlah Pelanggan | Nilai Rata-Rata Keseluruhan | | |
|------------------|-----------------------------|----|-------------|
| | R | F | M |
| 906 100% | 87 | 11 | 129.239.476 |

Sedangkan, perhitungan nilai rata-rata pada atribut RFM pada setiap segmen disajikan di Tabel 6. Didapatkan jumlah pelanggan pada segmen 0 sebanyak 271 dengan prosentase 30% dengan rata-rata nilai *recency* adalah 210 hari, nilai *frequency* adalah 4 kali transaksi dan nilai *monetary* adalah Rp39.300.768,00. Pada segmen 1 jumlah pelanggan sebanyak 307 dengan prosentase 34% dengan rata-rata nilai *recency* adalah 36 hari, nilai *frequency* adalah 21 kali transaksi dan nilai *monetary* adalah Rp275.483.299,00. Pada segmen 2 jumlah pelanggan sebanyak 328 dengan prosentase 36% dengan rata-rata nilai *recency* adalah 32 hari, nilai *frequency* adalah 7 kali transaksi dan nilai *monetary* adalah Rp66.667.939,00. Untuk mengklasifikasikan perilaku pelanggan berdasarkan segmennya dengan menggunakan simbol (↑) untuk nilai rata-rata per segmen lebih tinggi dari nilai rata-rata keseluruhan dan simbol (↓) untuk nilai rata-rata per segmen lebih rendah dari nilai rata-rata keseluruhan. Hasil klasifikasi berdasarkan nilai RFM dengan representasi simbol disajikan pada Tabel 6.

Berdasarkan grafik loyalitas pelanggan yang ditunjukkan pada Gambar 2, klasifikasi berdasarkan nilai RFM dengan representasi simbol menunjukkan bahwa segmen 0 berada di *lost customer group*, segmen 1 berada di *core customer group*, dan segmen 2 berada di *new customer group*. Dalam menentukan karakteristik pelanggan dilihat dari aspek geografis, demografi dan perilaku. Aspek

Tabel 6. Nilai RFM Rata-Rata Per Segmen

| Segmen | Jumlah Pelanggan | Nilai Rata-Rata Per Segmen | | |
|--------|------------------|----------------------------|----|-------------|
| | | R | F | M |
| 0 | 271 | 210 | 4 | 39.300.768 |
| | 30% | R↑ | F↓ | M↓ |
| 1 | 307 | 36 | 21 | 275.483.299 |
| | 34% | R↓ | F↑ | M↑ |
| 2 | 328 | 32 | 7 | 66.667.939 |
| | 36% | R↓ | F↓ | M↓ |

geografis berdasarkan mayoritas pelanggan berasal. Aspek demografi berdasarkan mayoritas saluran pembelian. Sedangkan, aspek perilaku berdasarkan kebiasaan dalam bertransaksi dan merk yang paling diminati. Tabel 7 menyajikan tipe dan karakteristik per segmen yang merupakan hasil dari analisis hasil *clustering*.

Tabel 7. Tipe dan Karakteristik Pelanggan

| Segmen | Tipe | Karakteristik |
|--------|----------------------|---|
| 0 | <i>Lost Customer</i> | <ul style="list-style-type: none"> a. Mayoritas berasal dari kota Jakarta Selatan dan transaksi terbesar melalui saluran pembelian toko/agen/ pengecer. b. Memiliki rentang waktu pembelian paling lama, frekuensi beli rendah, dan rata-rata pembelian kecil. c. <i>Revenue</i> per bulan paling tinggi di bulan Juni, dan pembelian hanya sampai di bulan Oktober. d. Aktif melakukan transaksi di bulan Juni. e. Merk yang diminati yaitu daging kita disusul manis kita. |
| 1 | <i>Core Customer</i> | <ul style="list-style-type: none"> a. Mayoritas berasal dari kota Serang dan transaksi terbesar melalui saluran pembelian perusahaan/ CV/UD. b. Memiliki rentang waktu pembelian agak pendek, frekuensi beli tinggi, dan rata-rata pembelian besar. c. <i>Revenue</i> per bulan naik signifikan di bulan Desember. d. Aktif melakukan transaksi di bulan November. e. Merk yang diminati yaitu daging kita disusul manis kita. |
| 2 | <i>New Customer</i> | <ul style="list-style-type: none"> a. Mayoritas berasal dari kabupaten Lebak dan pembelian terbesar melalui saluran pembelian toko/agen/ pengecer. b. Pelanggan memiliki rentang waktu pembelian pendek, frekuensi beli sedang, dan rata-rata pembelian sedang. c. <i>Revenue</i> per bulan paling tinggi di bulan Desember. d. Aktif melakukan transaksi di bulan Desember. e. Merk yang diminati yaitu daging kita disusul beras polos. |

Dari hasil analisis hasil *clustering*, didapatkan karakteristik dan tipe pelanggan per segmen, maka selanjutnya dilakukan penyusunan rekomendasi strategi pemasaran. Adapun strategi pemasaran yang direkomendasikan sebagai berikut :

- a. Segmen 0 dengan tipe pelanggan "*Lost Customer*" adalah pelanggan yang pernah melakukan transaksi namun tidak melakukan transaksi pada 2 bulan terakhir dan sebagai pelanggan musiman karena pembelian tertinggi di bulan Juni 2019 pada musim lebaran. Strategi pemasaran untuk pelanggan ini yaitu "*Don't Let Go Strategy*" yang mana perusahaan mempertahankan pelanggan meskipun sulit untuk membuatnya bertransaksi kembali. Mengidentifikasi pelanggan yang masih bisa dipertahankan. Penting untuk melakukan komunikasi kembali dan menawarkan produk-produk sesuai preferensi merk yang diminati. Memperbaiki pelayanan apabila pelanggan

- pergi dikarenakan pelayanan kurang memuaskan. Selanjutnya, memeriksa kembali apakah mereka masih minat untuk melakukan transaksi, apabila tidak hapus dari daftar pelanggan yang harus dipertahankan.
- b. Segmen 1 dengan tipe pelanggan “*Core Customer*” adalah pelanggan yang sering melakukan transaksi dengan jumlah nominal pembelian paling besar. Strategi pemasaran untuk pelanggan ini yaitu “*Defensive-Forward Strategy*” yang mana perusahaan harus mempertahankan pelanggan ini demi kelancaran bisnis. Menjaga komunikasi yang baik dengan segmen pelanggan ini, memberikan reward seperti diskon atau harga khusus dengan minimal jumlah atau kuantum pembelian yang telah ditentukan, memberikan informasi harga dan produk secara berkala, memberikan pelayanan terbaik, dan meningkatkan aktivitas promosi agar pelanggan melakukan transaksi pembelian kembali. Dikarenakan pelanggan terbesar berasal dari saluran pembelian perusahaan/CV/UD yang merupakan perusahaan yang mempunyai *downline* pasar dan keterjangkauan pasar yang luas, perlu menerapkan strategi *forward* atau meneruskan. Melakukan promosi produk baru kepada segmen pelanggan ini dan dorong mereka untuk mereferensikan produk-produk perusahaan ke *downline* pasar mereka.
 - c. Segmen 2 dengan tipe pelanggan “*New Customer*” adalah pelanggan yang baru melakukan transaksi. Strategi pemasaran untuk pelanggan ini yaitu “*Building Relationship Strategy*” yang mana perusahaan harus membangun hubungan dengan segmen pelanggan ini agar bisa menjadi *core customer*. Penting untuk membuat mereka mengetahui promosi penjualan yang sedang berjalan di perusahaan. Aktif melakukan promosi di sosial media untuk menjangkau pelanggan. Menjaga kualitas produk dan pelayanan agar memberikan kesan pertama yang baik. Melakukan monitoring secara berkala agar segmen pelanggan ini tidak menjadi *lost customer*.

5 Kesimpulan

Dalam penelitian ini dilakukan penambahan data terhadap data transaksi penjualan pada periode tahun 2019 untuk tujuan segmentasi pelanggan. Dari hasil penelitian yang telah dilakukan, segmentasi pelanggan dengan menerapkan algoritma K-Medoids menghasilkan jumlah *cluster* optimal adalah 3 (tiga). Secara umum dapat disimpulkan bahwa segmentasi paling baik untuk kasus dataset Perum BULOG ini adalah 3 (tiga) segmen yaitu *Lost Customer*, *Core Customer*, dan *New Customer*. Segmentasi pelanggan dapat mengidentifikasi tipe dan karakteristik pelanggan berdasarkan perilaku atau kebiasaan yang sama. Hasil segmentasi pelanggan dapat menjadi acuan dalam menyusun rekomendasi strategi pemasaran. Potensi riset masa depan yang bisa dikembangkan adalah memprediksi *Customer Lifetime Value (CLV)* dan memprediksi *customer churn*. Hal tersebut bertujuan untuk penargetan pelanggan agar strategi pemasaran yang disusun lebih tepat sasaran.

Referensi

- [1] B. E. Adiana, I. Soesanti, and A. E. Permanasari, “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.
- [2] R. W. Sembiring Brahmana, F. A. Mohammed, and K. Chairuang, “Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, pp. 32–43, 2020, doi: 10.24843/lkjiti.2020.v11.i01.p04.
- [3] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, “RFM ranking – An effective approach to customer segmentation,” *J. King Saud Univ. - Comput. Inf. Sci.*, 2018, doi: 10.1016/j.jksuci.2018.09.004.
- [4] A. R. Riszky and M. Sadikin, “Data Mining Menggunakan Algoritma Apriori untuk Rekomendasi Produk bagi Pelanggan,” *J. Teknol. dan Sist. Komput.*, vol. 7, no. 3, pp. 103–108, 2019, doi: 10.14710/jtsiskom.7.3.2019.103-108.
- [5] W. A. Triyanto, “Algoritma K-Medoids Untuk Penentuan Strategi Pemasaran Produk,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 6, no. 1, pp. 183–188, 2015, doi: 10.24176/simet.v6i1.254.
- [6] I. Kamila, U. Khairunnisa, and M. Mustakim, “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau,” *J. Ilm.*

- Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, pp. 119–125, 2019, doi: 10.24014/rmsi.v5i1.7381.
- [7] R. Dinar Dana, C. Lukman Rohmat, and A. Rizki Rinaldi, “Strategi Marketing Penerimaan Mahasiswa Baru Menggunakan Machine Learning dengan Teknik Clustering,” *J. Inform. J. Pengemb. IT*, vol. 04, no. 2, pp. 201–204, 2019, doi: 10.30591/jpit.v4i2-2.1879.
- [8] C. Astria, A. P. Windarto, and D. Hartama, “Penerapan K-Medoid Pada Rumah Tangga Yang Memiliki Sumber Penerangan Listrik Pln Berdasarkan Provinsi,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 604–609, 2019, doi: 10.30865/komik.v3i1.1667.
- [9] P. Arora, Deepali, and S. Varshney, “Analysis of K-Means and K-Medoids Algorithm for Big Data,” *Phys. Procedia*, vol. 78, no. December 2015, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.
- [10] S. Harikumar and P. V. Surya, “K-Medoid Clustering for Heterogeneous DataSets,” *Procedia Comput. Sci.*, vol. 70, pp. 226–237, 2015, doi: 10.1016/j.procs.2015.10.077.
- [11] D. Triyansyah and D. Fitriana, “Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing,” *J. Telekomun. dan Komput.*, vol. 8, no. 3, pp. 163–182, 2018, doi: 10.22441/incomtech.v8i3.4174.
- [12] A. D. Andini and T. Arifin, “Implementasi Algoritma K-Medoids Untuk Klasterisasi Data Penyakit Pasien Di Rsud Kota Bandung,” *J. RESPONSIF Ris. Sains ...*, vol. 2, no. 2, pp. 128–138, 2020, [Online]. Available: <http://ejurnal.ars.ac.id/index.php/jti/article/view/247>.
- [13] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, “Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak,” *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, pp. 64–71, 2018, doi: 10.24014/coreit.v4i2.4498.
- [14] H. Astuti, “Penerapan Data Mining Menggunakan Metode K-Means Clustering Untuk Pengelompokan Data Pelanggan (Studi Kasus: PT. Pinus Merah Abadi),” *J. Web Inform. Teknol.*, vol. 4, no. 1, pp. 9–16, 2019.
- [15] M. Aryuni, E. Didik Madyatmadja, and E. Miranda, “Customer Segmentation in XYZ Bank Using K-Means and K-Medoids Clustering,” *Proc. 2018 Int. Conf. Inf. Manag. Technol. ICIMTech 2018*, no. September, pp. 412–416, 2018, doi: 10.1109/ICIMTech.2018.8528086.
- [16] M. Kadir and A. Achyar, “Customer Segmentation on Online Retail using RFM Analysis: Big Data Case of Bukku.id,” *ICEASD*, 2019, doi: 10.4108/eai.1-4-2019.2287279.
- [17] F. Hadi, M. Mustakim, D. O. Rahmadia, F. H. Nugraha, N. P. Bulan, 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),” *J. Sains dan Teknol. Ind.*, vol. 15, no. 1, pp. 69–76, 2017.
- [18] S. Monalisa, “Segmentasi Perilaku Pembelian Pelanggan Berdasarkan Model RFM dengan Metode K-Means,” *QUERY J. Sist. Inf.*, vol. 02, no. April, pp. 9–15, 2018.
- [19] W. S. Manjoro, M. Dhakar, and B. K. Chaurasia, “Operational analysis of k-medoids and k-means algorithms on noisy data,” *Int. Conf. Commun. Signal Process. ICCSP 2016*, pp. 1500–1505, 2016, doi: 10.1109/ICCSP.2016.7754408.
- [20] N. Arbin, N. S. Suhaimi, N. Z. Mokhtar, and Z. Othman, “Comparative analysis between k-means and k-medoids for statistical clustering,” *Proc. - AIMS 2015, 3rd Int. Conf. Artif. Intell. Model. Simul.*, pp. 117–121, 2016, doi: 10.1109/AIMS.2015.82.
- [21] S. Defiyanti, M. Jajuli, and N. Rohmawati, “K-Medoid Algorithm in Clustering Student Scholarship Applicants,” *Sci. J. Informatics*, vol. 4, no. 1, pp. 27–33, 2017, doi: 10.15294/sji.v4i1.8212.
- [22] G. Ogbuabor and U. F. N, “Clustering Algorithm for a Healthcare Dataset Using Silhouette Score Value,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 10, no. 2, pp. 27–37, 2018, doi: 10.5121/ijcsit.2018.10203.
- [23] A. K. Singh, S. Mittal, P. Malhotra, and Y. V. Srivastava, “Clustering Evaluation by Davies-Bouldin Index(DBI) in Cereal data using K-Means,” *Proc. 4th Int. Conf. Comput. Methodol. Commun. ICCMC 2020*, no. Iccmc, pp. 306–310, 2020, doi: 10.1109/ICCMC48092.2020.ICCMC-00057.
- [24] D. C. Li, W. L. Dai, and W. T. Tseng, “A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing <http://sistemasi.ftik.unisi.ac.id>

- business,” *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7186–7191, 2011, doi: 10.1016/j.eswa.2010.12.041.
- [25] C. D. Rumiarti and I. Budi, “Customer Segmentation for Customer Relationship Management on Retail Company: Case Study PT Gramedia Asri Media,” *J. Sist. Inf.*, vol. 13, no. 1, pp. 1–10, 2017, doi: 10.21609/jsi.v13i1.525.