

IMPLEMENTASI ALGORITMA K-MEANS UNTUK SEGMENTASI PELANGGAN PADA PT. BINTANG MULTI SARANA CABANG TUGUMULYO

Herlinda Mawarni^{1*}, Gusmelia Testiana², Muhammad Leandry Dalafranka³

^{1,2,3}Program Studi Sistem Informasi, Universitas Islam Negeri Raden Fatah,
Jl. Pangeran Ratu 5 Ulu Palembang, Indonesia

^{1*}Email: herlindamawarni@gmail.com

²Email: gusmeliatestiana_uin@radenfatah.ac.id

³Email: leandry_uin@radenfatah.ac.id

ABSTRAK

Salah satu aset terpenting perusahaan adalah pelanggannya. Berbagai cara dilakukan oleh perusahaan untuk mempertahankan pelanggan dengan menerapkan berbagai strategi seperti pada PT Bintang Multi Sarana yang melakukan berbagai macam cara untuk menjalin hubungan baik dengan pelanggan seperti mengadakan promo dan diskon. Perusahaan menawarkan barang promo atau diskon dengan cara menghubungi pelanggan melalui telepon satu per satu atau menawarkan secara langsung ketika pelanggan selesai membeli barang sebelumnya. Penawaran diskon dan promo keseluruhan pelanggan ini kurang efisien karena memakan banyak biaya untuk operasional dan menyita banyak waktu. Oleh sebab itu, pelanggan yang ada di PT Bintang Multi Sarana perlu dikelompokkan terlebih dahulu agar dapat mempermudah perusahaan dalam menentukan layanan yang sesuai. Selain itu data transaksi penjualan pada PT Bintang Multi Sarana hanya diolah menjadi laporan bulanan saja. Padahal jika diolah dengan tepat, data transaksi penjualan dapat digunakan untuk membantu menentukan strategi dalam mempertahankan pelanggan. Tujuan penelitian ini yaitu melakukan segmentasi pelanggan untuk mengetahui karakteristik pelanggan. Salah satu solusi untuk permasalahan tersebut yaitu dengan mengolah data transaksi penjualan dengan memanfaatkan peran dari data mining dengan teknik *clustering* menggunakan algoritma *k-means*. Algoritma *k-means* merupakan algoritma berbasis jarak yang membagi data kedalam *cluster* terpisah. Berdasarkan perhitungan menggunakan *k-means*, pelanggan tercluster kedalam 3 kelompok yaitu *cluster* 1 sebanyak 742 pelanggan, *cluster* 2 sebanyak 795 pelanggan dan *cluster* 3 sebanyak 223 pelanggan. Dari hasil analisis segmentasi pelanggan diperoleh hasil bahwa *cluster* 1 merupakan pelanggan segmen *silver*, *cluster* 2 pelanggan segmen *gold*, dan *cluster* 3 pelanggan segmen *platinum*. Adapun strategi pemasaran untuk segmen *silver* dengan pemberian diskon harga, promosi dan iklan di media sosial, untuk segmen *gold* diberikan layanan berupa *point reward* atau diskon khusus, sedangkan untuk segmen *platinum* yaitu pemberian hadiah menarik, atau akses eksklusif terhadap produk tertentu.

Kata Kunci: Segmentasi Pelanggan, Data Mining, *Clustering*, *K-means*

ABSTRACT

One of a company's most important assets is its customers. Companies use various methods to retain customers by implementing various strategies, such as PT Bintang Multi Sarana which uses various methods to establish good relationships with customers, such as holding promos and discounts. Companies offer promotional items or discounts by contacting customers by telephone one by one or offering them directly when the customer has finished purchasing the previous item. Offering discounts and promos to all customers is less efficient because it costs a lot of operational costs and time. Therefore, customers at PT Bintang Multi Sarana need to be grouped first to make it easier for the company to determine appropriate services. Apart from that, sales transaction data at PT Bintang Multi Sarana is only processed into monthly reports. In fact, if processed properly, sales transaction data can be used to help determine strategies for retaining customers. The aim of this research is to segment customers to determine customer characteristics. One solution to this problem is to process sales transaction data by utilizing the role of data mining with clustering techniques using the *k-means* algorithm. The *k-means* algorithm is a distance-based algorithm that divides data into separate clusters. Based on calculations using *k-means*, customers are clustered into 3 groups, namely cluster 1 with 742 customers, cluster 2 with 795 customers and cluster 3 with 223 customers. From the results of the customer segmentation analysis, the results show that cluster 1 is a silver segment customer, cluster 2 is a gold segment customer, and cluster 3 is a platinum segment customer. The marketing strategy for the silver segment is by providing price discounts, promotions and advertising on social media, for the gold segment services are provided in the form of reward points or special discounts,

while for the platinum segment, this is the provision of attractive prizes or exclusive access to certain products.

Keywords: Customer Segmentation, Data Mining, Clustering, K-means

1. PENDAHULUAN

Pelanggan merupakan aset yang sangat berharga bagi perusahaan. Tanpa adanya pelanggan akan mengurangi pendapatan perusahaan. Hal ini mengakibatkan terhambatnya perkembangan bisnis dan mengganggu operasional dari perusahaan tersebut. Selain itu, pelanggan yang loyal pada perusahaan dapat memberikan informasi tentang barang apa yang mereka butuhkan sehingga perusahaan dapat mengembangkan produknya sesuai kebutuhan pelanggan.

Di tengah persaingan yang ketat saat ini, perusahaan mulai mengubah fokus mereka dari yang semula hanya pada pengembangan produk dan layanan yang unggul menjadi fokus pada hubungan dengan pelanggan untuk mempertahankan pelanggan. Hal ini dilakukan karena kesadaran bahwa hubungan yang baik antara perusahaan dan pelanggannya dapat membantu pertumbuhan perusahaan [1].

PT Bintang Multi Sarana merupakan salah satu perusahaan yang menjual alat-alat elektronik dan *furniture* seperti mesin cuci, televisi, kulkas, kipas angin, lemari, sofa, kasur dan lainnya. PT Bintang Multi Sarana berpusat di Palembang dengan salah satu cabangnya di desa Tugumulyo, kabupaten Ogan Komering Ilir. PT Bintang Multi Sarana melakukan berbagai macam cara untuk menjalin hubungan baik dengan pelanggan seperti mengadakan promo dan diskon. Namun diskon dan promo yang diberikan perusahaan belum tepat sasaran karena diberikan kepada seluruh pelanggan. Perusahaan menawarkan barang promo atau diskon dengan cara menghubungi pelanggan melalui telepon satu per satu atau menawarkan secara langsung ketika pelanggan selesai membeli barang sebelumnya. Hal ini kurang efisien karena memakan banyak biaya dan waktu. Oleh sebab itu, pelanggan yang ada di PT Bintang Multi Sarana perlu dikelompokkan terlebih dahulu agar dapat mempermudah perusahaan dalam menentukan layanan yang sesuai.

Dalam kegiatan transaksi, PT Bintang Multi Sarana mencatat transaksi penjualan ke dalam bentuk file Microsoft Excel. Data yang dicatat dalam file tersebut yaitu tanggal pembelian, nama pelanggan, nomor faktur, pendapatan, sales, alamat, nama barang, dan harga barang. Data transaksi penjualan pada PT Bintang Multi Sarana belum diolah secara maksimal yang mana data ini hanya diolah menjadi laporan penjualan bulanan saja. Padahal jika diolah dengan tepat, data transaksi penjualan dapat digunakan untuk membantu menentukan strategi untuk mempertahankan pelanggan. Salah satu bentuk pemanfaatan data transaksi penjualan dengan membagi pelanggan menjadi beberapa kelompok untuk mengetahui pelanggan-pelanggan yang memiliki potensi bagi perusahaan atau bisa juga disebut segmentasi pelanggan [2]. Pada PT Bintang Multi Sarana juga belum pernah dilakukan segmentasi pelanggan sehingga karakteristik dari pelanggan yang ada belum diketahui oleh perusahaan.

Segmentasi pelanggan adalah proses mengumpulkan kelompok pelanggan berdasarkan ciri-ciri yang sama. Segmentasi ini diperlukan untuk mempelajari perilaku konsumen dan membantu menerapkan rencana pemasaran terbaik guna meningkatkan pendapatan bisnis [3]. Solusi yang dapat digunakan dalam kegiatan segmentasi pelanggan yaitu dengan menggunakan *data mining*.

Data mining dapat diterapkan dalam bidang bisnis untuk membantu dalam proses pengambilan keputusan. *Data mining* digunakan untuk mengekstraksi data menjadi sebuah informasi yang berguna bagi perusahaan. Salah satu metode yang dapat digunakan untuk melakukan kegiatan segmentasi pelanggan yaitu metode *clustering* yang digunakan untuk mengelompokkan data. Metode ini berguna untuk menentukan segmentasi pelanggan dengan mengidentifikasi data-data yang mempunyai ciri atau karakteristik tertentu kemudian mengelompokkan data ke dalam sebuah *cluster* yang homogen. Dalam metode *clustering*, banyak algoritma yang bisa digunakan salah satunya algoritma *k-means*.

Algoritma *k-means* merupakan metode *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster*. Penelitian ini menggunakan algoritma *k-means* karena algoritma ini merupakan algoritma yang sederhana dan bersifat umum. Selain itu, *K-means* mampu mengelompokkan data yang berjumlah besar dengan waktu komputasi yang relatif cepat dan efisien [4].

Telah banyak penelitian yang membahas tentang segmentasi pelanggan seperti penelitian yang dilakukan Adiana [3] tentang analisis segmentasi pelanggan menggunakan kombinasi model *Recency, Frequency, and Monetary* (RFM) dan metode *clustering*. Segmentasi pelanggan menggunakan algoritma *clustering* ini menghasilkan 3 *cluster* dengan *cluster* pertama sebanyak 30 pelanggan yang termasuk ke *typical customer*, *cluster* kedua sebanyak 8 pelanggan masuk ke kelompok *superstar* dan *cluster* ketiga masuk ke kelompok *dormant customer*.

Pada penelitian ini menggunakan model RFM tetapi juga menggunakan variabel pendapatan yang merupakan salah satu aspek dalam segi demografis yang bisa digunakan untuk segmentasi pelanggan.

Penggunaan variabel pendapatan ini didasarkan bahwa pendapatan merupakan salah satu faktor yang mempengaruhi tingkat permintaan konsumen terhadap suatu produk [5]. Dalam penelitian lain tentang pengaruh pendapatan terhadap pilihan konsumen untuk membeli produk kosmetik impor, ditemukan bahwa tingkat pendapatan seseorang memiliki pengaruh terhadap pilihan mereka untuk membeli produk yang lebih mahal [6]. Selain itu, terdapat perbedaan tempat penelitian dan jumlah data yang digunakan.

Berdasarkan latar belakang yang telah diuraikan diatas, maka tujuan dari penelitian ini yaitu mengetahui karakteristik dari setiap segmen pelanggan yang terbentuk dari proses *clustering* menggunakan algoritma *k-means* dan membantu perusahaan mengidentifikasi strategi pemasaran yang tetap melalui segmentasi pelanggan.

2. MATERI DAN METODE

Data Set

Beragam-macam data yang disebut dataset sering kali mencakup objek dan properti. Fitur atau kualitas yang membentuk atribut objek. Sebagai ilustrasi, pertimbangkan suhu atau warna mata seseorang. Variabel, bidang, sifat, dan fitur adalah istilah yang digunakan untuk mendeskripsikan atribut. Catatan, titik, kasus, sampel, atau entitas adalah istilah untuk sekelompok atribut yang mencirikan suatu item. Kumpulan data adalah kumpulan data atau data statistik tunggal yang setiap elemen datanya dapat mewakili suatu variabel [7].

Data Mining

Data mining adalah proses menemukan pola dan tren yang bermanfaat secara besar-besaran [8]. Data mining merupakan teknik untuk memperoleh informasi berharga yang tersembunyi atau terpendam dalam database dengan jumlah yang besar sehingga ditemukan sebuah pola menarik yang belum diketahui sebelumnya [9].

Tahapan Data Mining

Berikut ini merupakan proses KDD secara garis besar [8]:

1. *Data Selection* merupakan tahapan pemilihan (seleksi) data. Tahapan ini perlu dilakukan sebelum penggalian informasi dilakukan. Pada tahap ini memilih kumpulan data yang akan digunakan untuk menemukan pengetahuan.
2. *Preprocessing/cleaning*, dilakukan pembersihan data yang mencakup pembuangan data yang duplikat, *noise*, kesalahan cetak/*tipografi* serta menangani data yang tidak konsisten dan relevan.
3. *Transformation*, data disesuaikan dengan format ekstensi yang sesuai untuk proses data mining. Ini dilakukan karena beberapa metode data mining memerlukan format tertentu sebelum diproses..
4. *Data mining*, menemukan pola atau informasi menarik dalam data melalui penggunaan teknik tertentu. *Data mining* menggunakan banyak teknik atau algoritma bergantung pada tujuan dan proses KDD secara keseluruhan, pendekatan atau algoritma yang sesuai akan dipilih.
5. *Interpretation/evaluation*, tahap ini termasuk dalam proses KDD yang dikenal sebagai interpretasi. Pola yang dihasilkan dari proses *data mining* harus ditampilkan sehingga pihak yang berkepentingan dapat memahaminya [10]. Pada titik ini juga dilakukan pemeriksaan ulang untuk memastikan apakah pola atau informasi yang dikumpulkan bertentangan dengan fakta atau hipotesis sebelumnya.

Clustering

Clustering merupakan kegiatan memisahkan dataset ke dalam beberapa sub-set sehingga elemen-elemen dalam kelompok tertentu memiliki tingkat persamaan yang tinggi dalam satu kelompok dan memiliki tingkat persamaan yang rendah dengan kelompok lain [11].

Saat data dikelompokkan, kelas objek yang sebanding dikelompokkan bersama. Berbeda dengan data pada kluster lain, kluster merupakan kumpulan data yang dapat diperbandingkan satu sama lain. Karena tidak ada variabel tujuan untuk pengelompokan ini, ini bervariasi dari klasifikasi. Tidak ada klasifikasi, estimasi, atau prediksi nilai variabel target yang dibuat selama pekerjaan pengelompokan. Namun, algoritma *clustering* bertujuan untuk membagi seluruh kumpulan data menjadi subkelompok atau cluster yang relatif homogen, di mana kesamaan karakter di dalam cluster dimaksimalkan dan kesamaan karakter di luar cluster ini diminimalkan [8].

K-Means

K-means adalah teknik clustering berbasis jarak yang menciptakan banyak cluster dari data. Metode yang membagi atau mempartisi data menjadi bagian yang berbeda adalah algoritma *k-means*. Karena kesederhanaan dan kemampuannya untuk mengelompokkan volume data yang sangat besar dengan cepat, teknik ini cukup terkenal. Setiap potongan data dalam algoritma *k-means* harus masuk ke dalam satu cluster tertentu. Selain itu, data dapat ditransfer ke cluster lain pada tahap selanjutnya setelah memasuki cluster tertentu selama langkah pemrosesan. Berikut ini merupakan langkah-langkah dari algoritma *k-means* [8]:

1. Tentukan berapa banyak *cluster* k .
2. Tentukan secara acak k *record* sebagai *centroid* awal.
3. Untuk setiap *record*, temukan *centroid* terdekat. Untuk menentukan *record* dengan *centroid* terdekat, dapat menggunakan jarak *Euclidean* (Persamaan 1).

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (1)$$

keterangan:

$d(p, q)$ = jarak antara p dan q .

4. Untuk setiap k *cluster*, temukan *centroid cluster*, dan perbarui lokasi setiap pusat *cluster* ke nilai *centroid* yang baru. *Centroid cluster* dapat ditemukan dengan cara misal n titik data $(a_1, b_1, c_1) \dots (a_2, b_2, c_2) \dots (a_n, b_n, c_n)$, *centroid* dari titik-titik ini adalah pusat *cluster* yang terletak di tempat dan terletak dititik dengan Persamaan 2.

$$\text{Cluster Center} = \left(\sum \frac{a_i}{n}, \sum \frac{b_i}{n}, \sum \frac{c_i}{n} \right) \quad (2)$$

Misalnya, titik $(1,1,1)$, $(1,2,1)$, $(1,3,1)$, dan $(2,1,1)$ akan menjadi pusat *centroid*.
 $\left(\frac{1+1+1+1}{4}, \frac{1+2+3+1}{4}, \frac{1+1+1+1}{4} \right) = 1.25, 1.75, 1.00$.

5. Ulangi langkah 3 dan 4 hingga konvergen. Algoritma ini berakhir ketika semua data yang dimiliki oleh setiap *centroid* berada di *cluster* yang tetap. Dengan kata lain, algoritma berhenti ketika untuk semua *cluster* C_1, C_2, \dots, C_k , semua *record* yang "dimiliki" oleh setiap pusat *cluster* tetap berada di *cluster* tersebut.

Euclidean Distance

Rumus Pythagoras dapat digunakan untuk menghitung jarak *Euclidean*, yaitu jarak antara dua titik. Perhitungan menggunakan geometri *Euclidean* memberikan hasil yang positif. Dengan memilih nilai terendah, kesamaan yang diukur menggunakan jarak *Euclidean* ditentukan. Jika hasil perbandingan dan perhitungan dengan jarak *Euclid* adalah nilai terendah atau bahkan sangat mendekati 0, maka dua item dianggap sebanding [12].

Normalisasi Min-Max

Teknik normalisasi yang mengubah data asli disebut normalisasi min-max. Untuk mengkonversi data linier, teknik ini menggunakan nilai terendah dan nilai maksimum dengan Persamaan 3 [13].

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

Keterangan: x' = nilai hasil normalisasi, x = nilai pada data yang akan dinormalisasi, x_{\max} = nilai maksimum dari data aktual, x_{\min} = nilai minimum dari data aktual.

Segmentasi Pelanggan

Pelanggan adalah pelanggan yang melakukan pembelian berulang atau menggunakan jasa karena kepuasannya terhadap penjual atau penyedia jasa [14]. Pelanggan sendiri memiliki berbagai jenis yang terdiri dari:

1. Individu yang bekerja untuk organisasi dan berdampak pada pekerjaan atau kinerja bisnis dikenal sebagai pelanggan internal.
2. Sekelompok orang yang tidak menggunakan produk perusahaan secara maksimal disebut pelanggan perantara. Ungkapan "pelanggan perantara" mengacu pada klien yang menjadi jembatan antara produsen dan konsumen. Pengecer, dealer sepeda motor, biro jasa, dan bisnis serupa lainnya adalah contoh klien perantara.
3. Seseorang yang memanfaatkan objek atau komoditas yang diperoleh untuk tujuannya sendiri dikenal sebagai klien eksternal. Klien eksternal juga berperan sebagai konsumen, namun berbeda dengan klien internal karena mereka secara konsisten membeli barang dari vendor atau lokasi yang sama [15].

Dalam konteks bisnis, pelanggan biasanya dianggap sebagai individu atau organisasi yang melakukan pembelian atau penggunaan jasa secara teratur dari suatu perusahaan. Dalam beberapa kasus, perusahaan biasanya menetapkan batasan minimal pembelian atau penggunaan jasa untuk mendapatkan status pelanggan atau untuk mendapatkan keuntungan tertentu, seperti diskon atau *reward*.

Karakteristik pelanggan dapat direpresentasikan oleh beberapa kategori variabel yang terkait dengan pengelompokan yaitu sebagai berikut:

1. Demografis: dalam segmentasi pasar dikelompokkan atas dasar variabel demografis meliputi umur, jenis kelamin, besarnya keluarga, besarnya kediamaan, siklus hidup keluarga, pemasukan, pekerjaan atau profesi, pendidikan, kepemilikan rumah, status sosial ekonomi, agama, kewarganegaraan.
2. Psikografis: dalam segmentasi ini, pembeli dibedakan menjadi kelompok berdasarkan variabel meliputi kepribadian, gaya hidup, kelas sosial.

3. Perilaku/*behavior*: meliputi manfaat yang dicari, status pembelian, tingkat penggunaan produk, frekuensi pembelian. Variabel perilaku banyak digunakan sebagai variabel terbaik untuk membentuk segmentasi pasar.
4. Geografis: meliputi Negara, provinsi, kota, kode pos, iklim [16].

Segmentasi pelanggan menjadi kelompok yang berbeda berdasarkan karakteristik yang sama memungkinkan perusahaan untuk mengembangkan strategi pemasaran yang disesuaikan dengan karakteristik pelanggan. Mengembangkan hubungan dengan pelanggan melalui pengembangan hubungan pelanggan akan membantu perusahaan mempertahankan dan menciptakan nilai tambahan bagi pelanggan..

Dalam penyusunan strategi pemasaran terdapat 3 komponen yaitu *segmentation*, *targeting*, dan *positioning*. *Segmentation* adalah upaya perusahaan memecah pelanggan menjadi kelompok yang lebih kecil di mana orang-orang dalam satu kelompok tersebut memiliki kemiripan sehingga memudahkan perusahaan mengidentifikasi dan menjangkau pelanggan berdasarkan perilaku mereka. *Targeting*, pada tahap ini perusahaan menentukan kelompok pelanggan mana yang akan menjadi target dari strategi pemasaran mereka. Sedangkan *positioning* adalah tahap di mana perusahaan menempatkan produk atau layanan di benak pelanggan. Pada tahap ini perusahaan perlu membuat program yang unik yang membedakan produk perusahaan dengan produk lain melalui iklan, *branding*, dan kegiatan pemasaran lainnya [17].

Pada penelitian ini pelanggan akan dibagi menjadi 3 yaitu *platinum*, *gold* dan *silver*, berikut merupakan karakteristik dari setiap segmennya pada Tabel 1.

Tabel 1. Karakteristik Segmen Pelanggan

Nama Segmen	Karakteristik Segmen
<i>Platinum</i>	<ul style="list-style-type: none">- Nilai kebaruan tertinggi- Nilai pendapatan pelanggan tinggi- Frekuensi pembelian tinggi- Total pembelian tinggi
<i>Gold</i>	<ul style="list-style-type: none">- Nilai kebaruan tinggi ke-2- Nilai pendapatan pelanggan tinggi ke-2- Frekuensi pembelian tinggi urutan ke2- Total pembelian tinggi di urutan ke-2
<i>Silver</i>	<ul style="list-style-type: none">- Nilai kebaruan cukup tinggi diurutan ke-3- Nilai pendapatan pelanggan tinggi ke-3- Frekuensi pembelian tinggi di urutan ke-3- Total pembelian tinggi diurutan ke-3

R Studio

R adalah bahasa *open-source* yang dapat digunakan untuk menjelajahi dan menganalisis kumpulan data. Analisis yang menggunakan R dapat memanfaatkan banyak paket, rutinitas, dan antarmuka pengguna grafis yang tersedia secara bebas, untuk mengatasi sebagian besar masalah analisis data [18].

Metode Penelitian

Metode yang digunakan dalam penelitian ini yaitu metode kuantitatif dengan menggunakan pendekatan deskriptif. Pendekatan penelitian kuantitatif menggunakan data numerik untuk mendeskripsikan fenomena sebelum menggunakan statistik untuk menilai hasilnya [19] [20]. Mengidentifikasi masalah yang dimaksudkan sebagai tujuan penelitian adalah tujuan dari metode penelitian kuantitatif, membuat prediksi untuk menentukan masalah, memperoleh data yang akan diprediksi, dan menafsirkan atau menganalisis data untuk melihat apakah mendukung untuk diprediksi [21].

Teknik Pengumpulan Data

Pada penelitian ini menggunakan teknik dalam proses pengambilan data, diantaranya sebagai berikut:

1. Wawancara merupakan kegiatan untuk mendapatkan keterangan untuk tujuan penelitian yang dilakukan dengan cara tanya jawab. Pada kegiatan ini, peneliti melakukan kegiatan wawancara dengan pimpinan unit cabang Tugumulyo untuk mengetahui informasi terkait data transaksi penjualan yang berhubungan dengan penelitian.
2. Observasi merupakan kegiatan pengamatan secara langsung yang dilakukan oleh peneliti menggunakan indera manusia [22]. Dalam kegiatan observasi ini, dilakukan dengan mengamati dan mencatat hal-hal yang dianggap penting yang berhubungan dengan penelitian.
3. Studi Pustaka untuk melakukan studi literatur, seseorang harus mencari bahan bacaan, seperti buku, jurnal, dan karya ilmiah, yang relevan dengan penelitian yang dipublikasikan di media.

3. HASIL DAN PEMBAHASAN

Hasil

Penelitian ini dilanjutkan dengan melakukan proses data mining yang digunakan untuk mengetahui segmen pelanggan yang terbentuk dari proses *clustering* menggunakan algoritma *k-means*. Dalam proses pembentukan segmen pelanggan menggunakan data transaksi penjualan dari tahun 2021-2022.

Pengumpulan Data

Pengumpulan data pada dilakukan pada PT Bintang Multi Sarana cabang Tugumulyo pada tanggal 03 November 2022 setelah melalui diskusi dengan pimpinan cabang dari PT Bintang Multi Sarana, serta kepala bagian Gudang. Data yang digunakan yaitu data transaksi penjualan sejak tahun 2021 sampai Desember 2022. Data diperoleh dari proses pembelian produk oleh pelanggan yang di simpan dalam format excel. Data transaksi penjualan PT Bintang Multi Sarana memiliki 2407 entri data.

Pengolahan Data

Tahapan pengolahan data terdiri dari beberapa aktivitas, yaitu *data selection*, *preprocessing*, *transformation*, *data mining*, dan *interpretation*. Penjelasan lebih rinci terakait tahapan pengolahan data dijelaskan pada bagian selanjutnya.

Data Selection

Data selection dilakukan dengan memilih data mana saja yang akan digunakan dalam proses data mining. Data yang digunakan pada penelitian ini yaitu data transaksi penjualan pada PT Bintang Multi Sarana cabang Tugumulyo dari tahun 2021 sampai tahun 2022 yang telah dilakukan oleh pelanggan terdapat 2407 transaksi.

Data Preprocessing

Pada tahap ini dilakukan proses pembersihan data. Proses pembersihan data dilakukan dengan mengidentifikasi variabel yang diperlukan dalam penelitian. Variabel yang tidak diperlukan dapat dihilangkan atau dihapuskan. Selain itu data transaksi penjualan yang digunakan pada penelitian ini ditemukan data kosong sebanyak 1 data sehingga terdapat 2406 data (Tabel 2).

Tabel 2. Data Siap Proses

No	Tanggal	Nama konsumen	Penghasilan	Total
1	3-mar-21	Imam badowi/wariyanti	3.000.000	950.000
2	3-mar-21	Nurmiati/dahlian	3.000.000	2.922.000
3	3-mar-21	Sopiah/sodik	3.000.000	3.400.000
4	3-mar-21	Dwi fitriasari/yahman	3.000.000	400.000
...
2406	31-dec-22	Rohani sugianto/katiem	2.500.000	3.300.000

Data Transformation

Data transaksi penjualan pada penelitian ini tidak dapat langsung diolah untuk meng*cluster* pelanggan. Variabel selisih/kebaruan, frekuensi dan total pembelian atau lebih dikenal dengan Variabel RFM pada penelitian sebelumnya dijadikan sebagai variabel untuk menghitung *cluster* pelanggan, pada penelitian ini juga dilakukan hal serupa dengan menambahkan variabel pendapatan pelanggan.

1. Transformasi Variabel Selisih

Variabel Tanggal pada dataset menunjukkan waktu terjadinya transaksi. Variabel tanggal perlu ditransformasikan kedalam bentuk selisih untuk menemukan interval waktu pembelian produk. Variabel selisih merupakan hasil yang menunjukkan kebaruan dari transaksi pelanggan dimana semakin kecil nilai selisih pelanggan berarti pelanggan tersebut baru saja melakukan transaksi. Variabel selisih diperoleh dengan cara menghitung interval waktu antara tanggal transaksi yang dilakukan pelanggan dengan tanggal yang ditetapkan penelitian, yaitu tanggal 01 Januari 2023. Variabel yang digunakan sebagai identitas pelanggan yaitu Nama Konsumen. Berikut ini merupakan cuplikan nilai hasil dari variabel selisih pada Tabel 3.

2. Transformasi Variabel pendapatan

Variabel Pendapatan merupakan variabel yang digunakan untuk mengetahui pendapatan perbulan dari masing-masing pelanggan.

Tabel 3. Transformasi Variabel Selisih

No	Tanggal	Selisih	Nama konsumen
1	3-marc-21	669	Imam badowi/wariyanti
2	6-may-21	605	Suwati/anton
3	6-may-21	605	Jungkir/dewi
4	6-may-21	605	Muhamad zen/sahilin
...
2406	31-dec-22	1	Rohani sugianto/katiem

3. Transformasi Variabel Frekuensi

Frekuensi adalah variabel untuk melihat berapa kali pelanggan melakukan transaksi (Tabel 4).

Tabel 4. Transformasi Variabel Frekuensi

No	Selisih	Nama Konsumen	Penghasilan	Frekuensi
1	613	Muryati/ lani dwi	3.000.000	2
2	605	Suwati / anton	3.400.000	1
3	605	Jungkir / dewi	3.000.000	1
4	605	Muhamad zen / sahilin	3.000.000	1
....
1815	1	Rohani sugianto/katiem	2.500.000	3

4. Transformasi Variabel Total Pembelian

Variabel Total Pembelian merupakan total keseluruhan transaksi yang dilakukan oleh setiap pelanggan (Tabel 5).

Tabel 5. Transformasi Variabel Total Pembelian

No	Selisih	Nama konsumen	Penghasilan	Frekuensi	Total pembelian
1	613	Muryati/ lani dwi	3.000.000	2	950.000
2	605	Suwati / anton	3.400.000	1	3.400.000
3	605	Jungkir / dewi	3.000.000	1	552.000
....
1815	1	Rohani sugianto/katiem	2.500.000	3	6.780.000

Pada tahap ini data transaksi pelanggan yang nilai total pembeliannya < Rp.500.000 (lima ratus ribu rupiah) dihapus karena tidak sesuai dengan kriteria pelanggan pada PT Bintang Multi Sarana sehingga data transaksi pelanggan yang siap diolah ke tahap selanjutnya yaitu 1760 data.

5. Normalisasi Variabel

Setelah data transaksi pelanggan dilakukan transformasi pada setiap variabel yang digunakan, proses selanjutnya yaitu normalisasi. Variabel yang didapat dari proses transformasi data memiliki satuan yaitu Tanggal dalam satuan hari, frekuensi menggambarkan jumlah kunjungan pelanggan, total pembelian dan pendapatan menggunakan satuan rupiah. Data yang akan digunakan perlu dilakukan normalisasi terdahulu agar selisih nilai antar pelanggan tidak terlalu jauh, selain itu normalisasi dilakukan untuk mengantisipasi outlier (Tabel 6).

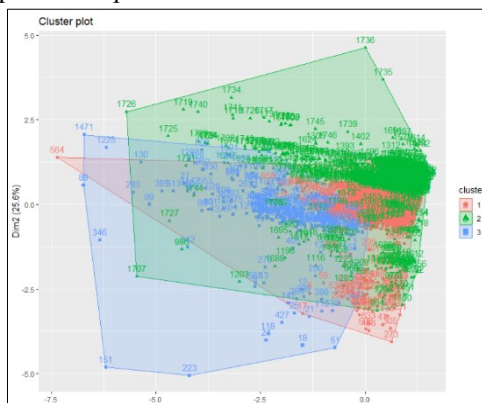
Tabel 6. Normalisasi Min-Max

No	Selisih	Penghasilan	Frekuensi	Total pembelian
1	1	0.4	0.25	0.013238916
2	0.986928105	0.48	0	0.088669951
3	0.986928105	0.4	0	0.000985222
...
1760	0	0.3	0.5	0.19273399

Data Mining

Pada tahap *data mining* dilakukan untuk mencari pola atau informasi yang menarik yang terdapat dalam data transaksi pelanggan dengan metode *clustering* dan algoritma *k-means* yang akan dilakukan dengan menggunakan perhitungan manual menggunakan Excel dan implementasi menggunakan *tools R studio*.

Setelah diketahui hasil dari pengelompokan data ke dalam 3 cluster menggunakan algoritma *k-means*, selanjutnya yaitu menganalisis *cluster-cluster* tersebut ke dalam segmen pelanggan. Hasil pengolahan dataset transaksi pelanggan pada PT Bintang Multi Sarana terbagi ke dalam 3 *cluster* dimana *cluster* 1 sebanyak 742 pelanggan, *cluster* 2 sebanyak 795 pelanggan dan *cluster* 3 sebanyak 223 pelanggan dengan visualisasi persebaran data dapat dilihat pada Gambar 1 berikut ini.



Gambar 1. Grafik Persebaran Data

Dari Gambar 1, dapat dilihat bahwa persebaran data dari setiap *cluster* saling berdekatan sehingga diketahui tingkat perbedaan *inter cluster*-nya rendah. Setelah diketahui jumlah anggota dari masing-masing *cluster*, langkah selanjutnya yaitu mengidentifikasi tipe atau karakteristik *cluster* berdasarkan nilai rata-rata dari variabel selisih, penghasilan, frekuensi dan total pembelian pelanggan. Berikut ini merupakan rata-rata nilai dari setiap variabel pada Tabel 7.

Tabel 7. Cluster Rata-rata

	Jumlah	Selisih	Penghasilan	Frekuensi	Total pembelian
Cluster 1	742	0,5673943	0,4241240	0,002021563	0,1005415
Cluster 2	795	0,1681260	0,4273208	0,068238994	0,1011852
Cluster 3	223	0,7002770	0,4286996	0,412556054	0,1644210

Setelah karakteristik dari masing-masing *cluster* berhasil dianalisis berdasarkan nilai rata-rata variabel, langkah selanjutnya adalah memetakan setiap *cluster* kedalam segmen pelanggan berdasarkan karakteristik dari setiap segmen, dari Tabel 7 didapatkan hasil yaitu *Cluster* 1 dengan total sebanyak 742 pelanggan merupakan pelanggan bertipe *silver*. Kelompok pelanggan *silver* merupakan pelanggan dengan nilai rata-rata kebaruan 348 hari, rata-rata penghasilan perbulan Rp. 3.120.620, rata-rata frekuensi transaksi sebanyak 1 kali, dan total pembelian sebesar Rp. 3.785.588. *Cluster* 2 dengan total sebanyak 795 pelanggan merupakan tipe pelanggan *gold*. Kelompok pelanggan *gold* merupakan pelanggan dengan nilai rata-rata kebaruan 104 hari, rata-rata penghasilan perbulan sebesar Rp. 3.136.604, rata-rata frekuensi transaksi sebanyak 1-2 kali, dan total pembelian sebesar Rp. 3.806.497. *Cluster* 3 dengan total sebanyak 223 merupakan tipe pelanggan *platinum* dengan nilai rata-rata kebaruan sebesar 430, rata-rata penghasilan perbulan sebesar Rp. 3.143.498, rata-rata frekuensi transaksi 3 kali, dan rata-rata total pembelian sebesar Rp. 5.860.395.

Pada penelitian sebelumnya yang dilakukan oleh Adiana [3], karakteristik *cluster* yang dihasilkan berupa *cluster* 1 dengan ciri berupa nilai rata-rata kebaruan, rata-rata frekuensi tertinggi setelah *cluster* 2, tetapi nilai rata-rata total pembelian paling rendah dari kedua *cluster* lainnya. Pada *cluster* 2 menghasilkan ciri berupa nilai rata-rata kebaruan, rata-rata frekuensi, rata-rata total pembelian tertinggi. Pada *cluster* 2 ini merupakan *cluster* terbaik. Sedangkan pada *cluster* 3 menghasilkan ciri nilai rata-rata kebaruan, rata-rata frekuensi paling rendah, akan tetapi rata-rata total pembelian berada diantara *cluster* 1 dan *cluster* 2. Sehingga jika dibandingkan hasil *cluster* pada penelitian ini dengan penelitian sebelumnya yaitu pada *cluster* terbaik terdapat kesamaan karakteristik berupa nilai rata-rata setiap variabel paling tinggi dari

cluster lain. Sedangkan pada dua *cluster* lainnya terdapat perbedaan karakteristik pada variabel total pembelian.

Hasil olah data transaksi pelanggan pada PT Bintang Multi Sarana berupa *cluster 1* atau segmen silver, *cluster 2* atau segmen gold dan *cluster 3* segmen platinum dari penelitian ini diketahui karakteristik setiap segmen pelanggan yang terbentuk dari proses *clustering* sehingga dapat digunakan perusahaan untuk membantu mengidentifikasi segmen pelanggan mana yang akan menjadi target pemasaran perusahaan. Berikut ini merupakan layanan yang direkomendasikan sesuai dengan segmen pelanggan berdasarkan karakteristik yang dihasilkan yaitu Segmen *silver* merupakan kelompok pelanggan dengan nilai rata-rata pendapatan, frekuensi, dan total pembelian rendah namun nilai rata-rata kebaruan pelanggan melakukan transaksinya menengah. Pelanggan pada segmen ini perlu diberikan layanan berupa pemberian diskon harga dan promosi produk. Selain itu kelompok pelanggan ini perlu dijangkau dengan menggunakan media elektronik seperti telepon, dan media sosial. Segmen *gold* merupakan kelompok dengan nilai pendapatan, frekuensi dan total pembelannya menengah akan tetapi kebaruan transaksinya rendah. Pelanggan pada segmen ini perlu diberikan layanan berupa *point reward* atau diskon khusus untuk pelanggan dimana program ini membantu meningkatkan minat transaksi mereka. Segmen *platinum* merupakan kelompok pelanggan dengan nilai pendapatan, frekuensi, total pembelian dan kebaruan transaksi paling tinggi. Pelanggan pada segmen ini perlu diberikan layanan berupa pemberian hadiah yang dapat menarik perhatian pelanggan platinum. Memberikan akses secara eksklusif terhadap produk-produk yang tidak tersedia guna menjaga hubungan baik antara perusahaan dengan dengan pelanggan.

4. KESIMPULAN DAN SARAN

Hasil *cluster* terbagi kedalam 3 *cluster* di mana *cluster 1* merupakan segmen pelanggan silver dengan jumlah 742 pelanggan dengan nilai rata-rata kebaruan 348 hari, penghasilan rata-rata Rp.3.120.620, frekuensi 1 kali, dan rata-rata total pembelian Rp. 3.785.588. Pada *cluster 2* merupakan segmen pelanggan *gold* dengan jumlah 795 pelanggan dengan nilai rata-rata kebaruan 104 hari, rata-rata penghasilan perbulan sebesar Rp. 3.136.604, rata-rata frekuensi transaksi sebanyak 1-2 kali, dan total pembelian sebesar Rp. 3.806.497. Pada *cluster 3* merupakan segmen pelanggan platinum dengan jumlah 223 pelanggan nilai rata-rata kebaruan sebesar 430, rata-rata penghasilan perbulan sebesar Rp. 3.143.498, rata-rata frekuensi transaksi 3 kali, dan rata-rata total pembelian sebesar Rp. 5.860.395. Oleh karena itu pada penelitian ini *cluster* terbaik merupakan *cluster 3*. Strategi pemasaran berdasarkan segmentasi pelanggan untuk segmen *silver* perlu diberikan layanan berupa pemberian diskon harga, promosi produk dan menjangkau dengan menggunakan media elektronik seperti telepon, dan media sosial. Segmen *gold* perlu diberikan layanan berupa *point reward* atau diskon khusus untuk pelanggan Segmen *platinum* perlu diberikan layanan berupa pemberian hadiah yang dapat menarik, memberikan akses secara eksklusif terhadap produk-produk yang jarang tersedia. Adapun saran untuk penelitian berikutnya yaitu menggunakan data transaksi penjualan yang lebih banyak agar pelanggan terbaru dapat ter-*cluster* kedalam segmen pelanggan. Menambahkan variabel lain dari aspek demografis, psikografis, ataupun geografis untuk mengidentifikasi karakteristik segmen pelanggan yang lebih spesifik.

DAFTAR PUSTAKA

- [1] N. R. Syarif and Windarto, "Implementasi Algoritma Fuzzy C-means dan Metode Recency Frequency Monetary (RFM) pada Aplikasi Data Mining untuk Pengelompokan Pelanggan," *Sebatik*, vol. Vol.22, no, pp. 88–94, 2018. [Online]. Available: <https://jurnal.wicida.ac.id/index.php/sebatik/article/view/313>.
- [2] A. A. D. Sulistyawati and M. Sadikin, "Penerapan Algoritma K-Medoids untuk Menentukan Segmentasi Pelanggan," *Sistemasi*, vol. 10, no. 3, p. 516, 2021, doi: [10.32520/stmsi.v10i3.1332](https://doi.org/10.32520/stmsi.v10i3.1332).
- [3] B. E. I. S. dan A. E. P. Adiana, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," *JUTEI (Jurnal Terap. Teknol. Informasi)*, vol. Volume.2 N, pp. 23–32, 2018, doi: [10.21460/jutei.2018.21.76](https://doi.org/10.21460/jutei.2018.21.76).
- [4] M. H. Siregar, "Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan)," *J. Teknol. Dan Open Source*, vol. 1, no. 2, pp. 83–91, 2018, doi: [10.36378/jtos.v1i2.24](https://doi.org/10.36378/jtos.v1i2.24).
- [5] Arwin, *Buku Ajar Pengantar Ekonomi Mikro*. Cendikia Publisher, 2020.
- [6] Rizky and Rosalina Noor, "Pengaruh Harga, Pendapatan, dan Halal Awareness Terhadap Keputusan Pembelian Produk Kosmetik Impor Pada Konsumen Pengguna E-commerce Sociolla (Studi Pada Mahasiswa di Universitas Brawijaya)," *J. Ilm. Fak. Ekon. dan Bisnis*, pp. 1–14, 2020. [Online]. Available: <https://jimfeb.ub.ac.id/index.php/jimfeb/article/view/7036>.
- [7] E. Listiana and M. A. Muslim, "Penerapan Adaboost untuk Klasifikasi Support Vector Machine guna

- meningkatkan Akurasi pada Diagnosa Chronic Kidney Disease,” *Pros. SNATIF*, no. 2015, pp. 875–881, 2017. [Online]. Available: <https://jurnal.umk.ac.id/index.php/SNA/article/view/1463>.
- [8] D. T. Larose and C. D. Larose, *Data Mining and Predictive Analytics*, Second Edi. Canada: John Wiley & Sons, Inc, 2015.
- [9] A. M. dan A. P. Siregar, *Data Mining Pengolahan Data Menjadi Informasi dengan RapidMiner*. Surakarta: CV Kekata Group, 2017.
- [10] W. Lestari, “Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus : STMIK Bina Bangsa Kendari),” *Simkom*, vol. 4, no. 2, pp. 35–48, 2019, doi: [10.51717/simkom.v4i2.37](https://doi.org/10.51717/simkom.v4i2.37).
- [11] F. Astuti Hermawati, *Data mining*. Yogyakarta: Andi, 2013.
- [12] A. Wanto dkk, *Biometrika: Teknologi Identifikasi*. Yayasan Kita Menulis, 2020.
- [13] Suyanto, *Data Mining untuk Klasifikasi dan Klasterisasi*. Bandung: Informatika Bandung, 2017.
- [14] D. Fatihudin dan A. Firmansyah, *Pemasaran Jasa (Strategi, Mengukur Kepuasan Dan Loyalitas Pelanggan)*. Yogyakarta: Deepublish, 2019.
- [15] Candrianto, *Kepuasan Pelanggan Suatu Pengantar*, I. Malang: Literasi Nusantara, 2021.
- [16] J. P. Nugraha dkk, *Teori Perilaku Konsumen*, 1st ed. Pekalongan: Nasya Expanding Management, 2021.
- [17] D. Z. Basil, G. Diaz-Meneses, and M. D. Basil, *Correction to: Social Marketing in Action*. 2019. doi: [10.1007/978-3-030-13020-6_30](https://doi.org/10.1007/978-3-030-13020-6_30).
- [18] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data An Introduction to Data Mining*, Second. United States of America, 2014.
- [19] U. Suharsaputra, *Metode Penelitian Kuantitatif, kualitatif, dan Tindakan*, Ketiga. Bandung: PT Refika Aditama, 2018.
- [20] I. dan H. M. Santoso, *Metodologi Penelitian Kuantitatif*. Indigo Media, 2021.
- [21] S. dan M. A. S. Siyoto, *Dasar Metodologi Penelitian*. Yogyakarta: Literasi Media Publishing, 2015.