

### 3 Analisis Segmentasi Pelanggan Ritel Z Menggunakan *K-Means Clustering*

Roaida Yanti<sup>\*1)</sup>, Prita Nurkhalisa Maradjabessy<sup>2)</sup>

<sup>1,2)</sup>Program Studi Teknik Industri, Fakultas Teknologi Industri, Universitas Islam Indonesia, Sleman, Yogyakarta, 55588, Indonesia

Email: roaida.yanti@students.uui.ac.id, prita.maradjabessy@students.uui.ac.id

#### ABSTRAK

Pesatnya perkembangan industri ritel saat ini membuat persaingan antar ritel menjadi sangat tinggi. Titik persaingan tersebut berupa perebutan konsumen. Maka peritel harus melakukan inovasi dan strategi agar dapat memenangkan hati konsumen. Salah satu cara yang dapat membantu perusahaan dalam merumuskan strategi mereka adalah dengan mensegmentasi pelanggannya. Dalam penelitian ini segmentasi pelanggan perlu dilakukan ritel Z untuk mengetahui karakteristik pelanggan dan mengelompokkannya agar dapat memberikan strategi untuk setiap kelompok segmen tersebut. Metode yang digunakan adalah *clustering* dengan algoritma K-Means terhadap 100 data pelanggan yang diperoleh melalui penyebaran angket kuesioner. Hasil yang didapatkan dari proses data mining menggunakan software SPSS bahwa *cluster* 1 merupakan *cluster* yang kurang potensial dengan persentase 46%, *cluster* 2 dan *cluster* 3 potensial dengan persentase berturut-turut 1% dan 12%, kemudian *cluster* 4 juga *cluster* yang kurang potensial dengan persentase 41%. Dengan demikian berdasarkan hasil tersebut, diberikan rekomendasi strategi yang sesuai untuk masing-masing *cluster*.

**Kata kunci:** algoritma k-means, *cluster*, ritel, segmentasi pelanggan.

#### 1. Pendahuluan

Ritel merupakan segala aktivitas bisnis yang berkaitan dengan penjualan produk atau jasa secara langsung kepada konsumen akhir. Perkembangan bisnis ritel yang semakin pesat menjadikan industri ritel sebagai salah satu jenis usaha yang memiliki tingkat persaingan yang tinggi. Persaingan pada bisnis ritel berpusat pada perebutan pelanggannya. Konsumen atau pelanggan menjadi pilar utama bagi keberlangsungan perusahaan ritel. Sehingga bisnis ritel harus berfokus pada konsumen karena bisnis ini berperan sebagai elemen terakhir dalam saluran distribusi produk. Menurut Surjandari dan Seruni (2010), keberhasilan bisnis ritel bergantung bagaimana kecepatan dan kemampuannya dalam merespon perilaku pelanggan. Maka perusahaan ritel harus melakukan inovasi dan menjalankan strategi khusus agar dapat bersaing dengan perusahaan lain. Strategi tersebut dilakukan untuk meningkatkan pendapatan yang berasal dari jumlah transaksi yang dilakukan oleh pelanggan.

Segmentasi pelanggan adalah salah satu strategi yang dilakukan dengan mengelompokkan pelanggan yang mempunyai kebutuhan, karakteristik, atau perilaku yang berbeda menjadi kelompok-kelompok pelanggan yang memiliki kesamaan. Menurut Kansal dkk. (2018), segmentasi pelanggan merupakan penerapan penambangan data yang membantu mensegmentasi pelanggan dengan pola serupa ke dalam kelompok yang serupa sehingga memudahkan bisnis untuk menangani basis pelanggan yang besar. Segmentasi pelanggan ini menjadi penting karena keterlibatannya sebagai data pendukung untuk merumuskan strategi pemasaran yang tepat (Sulistiyowati, 2018). Hal ini dikarenakan segmentasi pelanggan dapat membuka banyak pintu baru untuk menemukan pengetahuan terkait segmen mana yang potensial untuk suatu produk tertentu, menyesuaikan rencana pemasaran sesuai dengan masing-masing segmen, memberikan diskon untuk segmen tertentu, serta menguraikan karakteristik pelanggan lainnya dan menghubungkan dengan objek yang sebelumnya tidak diketahui oleh perusahaan.

Maka penelitian ini bertujuan untuk mengidentifikasi karakteristik pelanggan dalam setiap segmen pelanggan pada ritel Z. Metode *clustering* digunakan untuk mensegmentasi pelanggan tersebut. Metode *clustering* merupakan proses mengelompokkan sekumpulan obyek data

menjadi beberapa kelompok atau *cluster* yang lebih kecil dengan setiap cluster memuat objek yang memiliki kemiripan yang tinggi, namun sangat berbeda dengan objek dalam cluster lainnya. Terdapat dua algoritma *clustering* yang sangat populer yaitu *Hierarchical Clustering* dan *Non Hierarchical Clustering* atau biasa disebut K-Means (Muhidin, 2017). Namun pada penelitian ini menggunakan algoritma K-Means, karena merupakan algoritma *clustering* yang cepat dan sederhana. K-Means adalah salah satu algoritma terkenal yang pembelajarannya tanpa memerlukan pengawasan sehingga menjadi paling sederhana tapi tetap handal dalam memecahkan masalah pengelompokan (Kodinariya dan Makwana, 2013).

Banyak penelitian yang telah menggunakan *clustering* untuk melakukan segmentasi pelanggan. Penelitian yang dilakukan Widiyanto & Witanti (2021), menggunakan Algoritma K-Means sebagai dasar strategi pemasaran untuk membantu perusahaan dalam mengelompokkan pelanggan dan membantu memprioritaskan tenaga dan sumber daya kepada pelanggan yang potensial. Selain itu, Algoritma K-Means membantu dalam mengklasifikasikan setiap kelompok pelanggan dan menentukan tingkat loyalitasnya (Adiana dkk, 2018). Lubis (2016) menyatakan bahwa hasil dari analisis algoritma K-Means *Clustering* dapat mendorong pertumbuhan jangka panjang dan peningkatan profitabilitas perusahaan dengan mengetahui perilaku pelanggan dan memenuhi kebutuhan mereka. Menurut Murpratiwi dkk (2021) hal yang paling penting dalam proses clustering adalah ketika menentukan algoritma dan nilai k awal. Sehingga pada penelitian ini dilakukan analisis jumlah cluster yang ideal berdasarkan pada percobaan dengan menginputkan jumlah *cluster* pada *software* SPSS dan membandingkan hasil yang terbentuk.

## 2. Metode

Metode penelitian menggambarkan tahapan yang akan dilakukan dalam upaya pemecahan masalah. Pada penelitian terdapat beberapa tahapan untuk menganalisis segmentasi konsumen yaitu pengumpulan data, *selection data*, *pre-processing*, *transformation*, pengolahan data menggunakan k-means, hasil dan pembahasan.

### 2.1 Pengumpulan Data

Pada penelitian ini, data diperoleh dari penyebaran angket kuesioner kepada pelanggan di ritel Z. Dalam kuesioner tersebut, terdapat 10 variabel yang menjadi indikator pertanyaan meliputi usia, jenis kelamin, *family life cycle*, pendapatan, durasi belanja, jarak tempat tinggal ke ritel, pekerjaan, *recency* (waktu terakhir belanja), frekuensi belanja, dan *monetary* (uang yang dihabiskan untuk belanja). Sepuluh variabel tersebut didapatkan dari studi literatur penelitian sebelumnya.

### 2.2 Data Selection

Selanjutnya terdapat proses pemilihan data dari kumpulan data yang didapatkan. Hal ini dilakukan untuk meringankan beban dalam pemrosesan data mining (Fong dkk., 2016). Sehingga bisa mendapatkan hasil yang lebih spesifik dan akurat sesuai dengan tujuan penelitian yang hendak dicapai. Jumlah data yang dikumpulkan dan telah melewati tahap seleksi sebanyak 100 data.

### 2.3 Pre-processing

Tahap ini merupakan proses menyiapkan data sebelum digunakan pada tahap pengolahan data mining. *Pre-processing* dilakukan untuk memastikan bahwa data yang akan digunakan telah lengkap dan sempurna. Terdapat beberapa proses yang dilakukan seperti menghilangkan *noise* dan sebagai penanganan terhadap *missing value* pada data.

### 2.4 Transformation

Data transformasi merupakan proses perubahan skala data aktual menjadi skala data tertentu. Tahap ini bertujuan untuk mengubah bentuk data agar dapat diolah menggunakan suatu algoritma yang membutuhkan data dalam bentuk klasifikasi atau numerik (Triyansyah dan Fitrihanah, 2018). Data yang diperoleh dari penyebaran kuesioner belum sesuai dengan format bentuk data untuk pemrosesan karena memerlukan data dengan tipe numerik, sehingga perlu dilakukan

pengkategorian data. Jenis data nominal yang dilakukan transformasi meliputi pekerjaan dan jenis kelamin. Hasil transformasi ditunjukkan pada tabel 1 dan tabel 2.

**Tabel 1.** Transformasi Data Atribut Jenis Kelamin

Jenis Kelamin	Transformasi
Laki-laki	1
Perempuan	2

**Tabel 2.** Transformasi Data Atribut Pekerjaan

Pekerjaan	Transformasi
Pelajar/Mahasiswa	1
Ibu rumah tangga/Wirausaha	2
Karyawan swasta/PNS	3

### 2.5 Pengolahan data dengan K-means

Proses pengolahan data menggunakan salah satu teknik data mining yaitu *clustering* dengan software SPSS. Algoritma yang dipilih untuk menunjang proses data mining tersebut adalah algoritma k-means yang merupakan salah satu metode *clustering* yang paling populer digunakan. Metode k-means akan menerima masukan berupa data tanpa label kelas lalu akan membagi data tersebut menjadi beberapa kelompok. Dalam penelitian ini, pengolahan data dilakukan dengan memakai *cluster tools* berupa data *training* yang diambil dari data mentah melalui proses seleksi, *pre-processing*, dan transformasi data. Untuk mengkalkulasi iterasi jarak terdekat antar *cluster* digunakanlah rumus *Euclidean Distance*. Nilai centroid awal atau nilai tengah ditentukan lalu kemudian partisi data menggunakan nilai minimum dilakukan. Iterasi akan terus dihitung dan berhenti ketika nilai atau posisi data tidak mengalami perubahan, iterasi terakhir itulah yang menjadi centroid akhir dari data yang sudah di partisi sebelumnya. Tahap ini dimulai dengan memilih sejumlah nilai *cluster* awal sesuai dengan jumlah yang diinginkan dan kemudian objek digabungkan ke dalam *cluster-cluster* tersebut. Sehingga output dari proses ini adalah *cluster-cluster* dengan karakteristik pelanggan.

### 2.6 Hasil Penelitian

Pada tahap ini, hasil proses *data mining* akan divisualisasikan dan dipaparkan analisisnya yang berupa kelompok konsumen dengan cerminan perilaku dan karakteristik yang berbeda-beda di setiap *cluster*-nya. Hasil berupa *knowledge* ini nantinya digunakan sebagai bahan pertimbangan untuk memberi rekomendasi kepada perusahaan dalam menentukan strategi pemasaran yang akan digunakan.

## 3. Hasil dan Pembahasan

Penelitian ini menggunakan salah satu teknik data mining yaitu *cluster* yang memiliki dapat membantu proses pengelompokan data. Dalam hal ini, algoritma K-Means digunakan pada 100 data pelanggan. Sebelum melakukan *clustering*, jumlah *cluster* K perlu ditentukan terlebih dahulu. Pemilihan jumlah *cluster* didasarkan pada percobaan dengan menginputkan beberapa jumlah *cluster* pada *software* SPSS dan membandingkan hasil yang terbentuk. Sehingga diputuskan berdasarkan hasil tersebut, jumlah *cluster* yang diambil yaitu 4 *cluster*. Dalam pembentukan 4 *cluster* tersebut, terdapat beberapa proses pengelompokan yang dilakukan seperti *initial cluster centers* yang ditunjukkan oleh Tabel 1 dan *iteration history* sebagai berikut.

**Tabel 3.** *Initial Cluster Center*

Atribut	1	2	3	4
	<i>Cluster</i>			
Zscore(Usia)	-0,2205	-0,2205	5,2922	-0,4410

Atribut	Cluster			
	1	2	3	4
Zscore(Gender)	-1,1455	-1,1455	0,8642	-1,1455
Zscore(Pekerjaan)	-0,3462	-0,3462	1,5772	-0,3462
Zscore(Pendapatan)	-0,1148	8,6229	0,5323	0,2087
Zscore(Anggota_Keluarga)	-0,3811	-0,3811	2,7090	4,7690
Zscore(Jarak_Tempuh)	-0,6884	-0,6884	-0,8908	-0,0809
Zscore(Waktu_Tempuh)	-0,0098	-0,0098	-0,0098	-0,0098
Zscore(Kunjungan_Terakhir)	-0,8203	-0,8203	-0,7252	2,0487
Zscore(Intensitas_Kunjungan)	0,2349	0,2349	0,2349	-0,6686
Zscore(Durasi_Belanja)	-1,0211	-1,0211	0,7651	-0,1280
Zscore(Jumlah_Biaya)	4,3328	4,3328	1,0406	-0,3311

Tabel *Initial Cluster Centers* menunjukkan hasil proses sementara pengelompokan data yang dilakukan. *Cluster Centers* merupakan titik data dengan kepadatan lebih tinggi, ambang batas radius lebih kecil dan jauh dari satu sama lain (Li dkk., 2019). Penentuan *initial cluster centers* dengan melakukan pengelompokan dinamis dan kemudian melakukan pengelompokan tradisional dengan algoritma K-Means (Zhu dan Ma, 2018). Selanjutnya, melakukan iterasi atau pengulangan untuk mendapatkan karakteristik yang tepat untuk setiap *cluster*. Hasil iterasi disajikan pada Tabel 4.

**Tabel 4.** *Iteration History*

Iteration	Change in Cluster Centers			
	1	2	3	4
1	4.618	0.000	3.479	4.406
2	0.272	0.000	0.859	0.431
3	0.227	0.000	0.000	0.368
4	0.204	0.000	0.000	0.254
5	0.145	0.000	0.000	0.150
6	0.039	0.000	0.000	0.041
7	0.000	0.000	0.000	0.000

Berdasarkan Tabel 4, diketahui bahwa proses *clustering* dilakukan melalui 7 iterasi untuk dapat menentukan *cluster* yang tepat. Dari tabel di atas disebutkan bahwa jarak minimum antar pusat *cluster* yang terjadi dari hasil iterasi adalah 7,537. Sehingga didapatkan 4 *cluster* terbaik yang terbentuk seperti pada Tabel. 5

**Tabel 5.** *Cluster yang Terbentuk*

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Umur	<20 tahun	<20 tahun	>30 tahun	<20 tahun
Jenis Kelamin	perempuan	laki-laki	perempuan	laki-laki
Pekerjaan	pelajar/mahasiswa	pelajar/mahasiswa	karyawan swasta/PNS	pelajar/mahasiswa
Pendapatan	1-2 juta	>5 juta	2-3juta	1-2 juta
Anggota keluarga	1	1	>3	2-3
Jarak Tempuh	<5 Km	<5 Km	<5 Km	>10 km
Waktu Tempuh	<15 menit	< 15 menit	< 15 menit	>25 menit

Intensitas Kunjungan	2 kali	3 kali	2 kali	1 kali
Durasi Belanja	<1 jam	<30 menit	<1,5 jam	<1 jam
Jumlah Biaya	>200.000	>500.000	>200.000	<150.000

Berdasarkan proses *cluster* yang telah dilakukan menggunakan algoritma k-means, diperoleh 4 *cluster* yang terdiri dari beberapa anggota yang memiliki karakteristik hampir sama. Dalam *cluster-1* terdiri dari konsumen dengan karakteristik usia dibawah 20 tahun, jenis kelamin perempuan, pekerjaan sebagai pelajar/mahasiswa, pendapatan 1-2 juta, tinggal sendiri, memiliki jarak tempuh dengan toko kurang dari 5 kilometer dengan waktu tempuh kurang dari 15 menit, intensitas kunjungan selama sebulan sebanyak 2 kali dengan durasi belanja dibawah 1 jam dan dapat menghabiskan jumlah biaya belanja kurang dari 200.000. *Cluster-2* terdiri dari konsumen dengan karakteristik usia dibawah 20 tahun, jenis kelamin laki-laki, pekerjaan sebagai pelajar/mahasiswa, pendapatan diatas 5 juta, tinggal sendiri, memiliki jarak tempuh dengan toko kurang dari 5 kilometer dengan waktu tempuh kurang dari 15 menit, intensitas kunjungan selama sebulan sebanyak 3 kali dengan durasi belanja dibawah 30 menit dan dapat menghabiskan jumlah biaya belanja lebih dari 500.000.

Pada *Cluster-3* terdiri dari konsumen dengan karakteristik usia diatas 30 tahun, jenis kelamin perempuan, pekerjaan sebagai karyawan swasta/PNS, pendapatan 2-3 juta, tinggal bersama lebih dari 3 orang, memiliki jarak tempuh dengan toko kurang dari 5 kilometer dengan waktu tempuh kurang dari 15 menit, intensitas kunjungan selama sebulan sebanyak 2 kali dengan durasi belanja dibawah 30 menit dan dapat menghabiskan jumlah biaya belanja lebih dari 200.000. Sedangkan *Cluster-4* terdiri dari konsumen dengan karakteristik usia dibawah 20 tahun, jenis kelamin laki-laki, pekerjaan sebagai pelajar/mahasiswa, pendapatan 1-2 juta, tinggal bersama lebih dari 2-3 orang, memiliki jarak tempuh dengan toko lebih dari 10 kilometer dengan waktu tempuh lebih dari 25 menit, intensitas kunjungan selama sebulan sebanyak 1 kali dengan durasi belanja dibawah 1 jam dan dapat menghabiskan jumlah biaya belanja kurang dari 150.000. Berikut detail jumlah anggota pada masing-masing *cluster* dari 100 konsumen sebagai berikut.

**Tabel 6.** Jumlah Anggota *Cluster*

<i>Number of Cases in each Cluster</i>		
<i>Cluster</i>	1	46
	2	1
	3	12
	4	41
Valid		100

Tabel 6. menyajikan jumlah konsumen yang masuk ke dalam tiap *cluster*. Maka berdasarkan tabel, *cluster-1* merupakan *cluster* dengan anggota paling banyak yang terdiri dari 46 konsumen, *cluster-2* hanya terdiri dari 1 konsumen, *cluster-3* terdiri dari 12 konsumen, dan *cluster-4* terdiri dari 41 konsumen.

Maka dari hasil *cluster* yang telah didapatkan, diketahui bahwa *cluster* yang paling potensial dilihat dari karakteristi yang dimiliki yakni *cluster-2* karena jumlah biaya yang dihabiskan dalam sekali belanja relatif tinggi dengan intensitas kunjungan yang relatif sering. Sedangkan *cluster* yang paling tidak potensial adalah *cluster-4* karena jumlah biaya yang dihabiskan dalam sekali belanja relatif rendah dengan intensitas kunjungan yang sangat jarang dan memiliki jarak tempuh dengan toko yang relatif jauh. Sedangkan *cluster-1* dan *cluster-4* merupakan *cluster* dengan anggota yang banyak.

Hasil dari jumlah anggota *cluster* tersebut dapat mewakili dominasi konsumen perusahaan. Sehingga perusahaan perlu memberikan strategi untuk konsumen pada *cluster-1* dan *cluster-4* seperti memberikan promosi berupa diskon untuk pembelian produk yang sesuai dengan karakteristik *cluster-1* dan *cluster-4*, dengan minimal belanja tertentu dan menggunakan batasan waktu, agar konsumen pada kelompok tersebut dapat meningkatkan belanjanya. Hal ini dikarenakan *cluster-1*, dan *cluster-4* memiliki jarak tempuh yang lumayan dekat dengan pembelian yang lumayan rendah. Hamali dan Anggriani (2020) dalam penelitiannya menyatakan bahwa faktor promosi penjualan memberikan pengaruh yang signifikan terhadap keputusan pembelian suatu produk. Selain itu, menurut Septika dkk. (2022), strategi promosi berupa diskon atau potongan harga dapat menarik konsumen untuk berbelanja sehingga dapat mempertahankan bisnis.

#### 4. Simpulan

Melalui semua proses pengolahan serta analisis data yang telah dilakukan, hasil dari penelitian ini menunjukkan bahwa terbentuk 4 *cluster* dengan algoritma k-means pada metode *clustering* untuk mengetahui segmentasi pelanggan pada Toko Ritel Z. Pelanggan pada *cluster 1* dan *cluster 4* mendapat kategori kurang potensial dengan persentase berturut-turut 46% dan 41%, *cluster 2* dengan persentase 1% dan *cluster 3* dengan persentase 12% kedua dengan kategori potensial. Rekomendasi yang diberikan adalah untuk meningkatkan strategi penjualan dengan memanfaatkan *knowledge* yang didapat dari karakteristik pelanggan yang berbeda-beda dengan tingkat kategori potensial di setiap *cluster*, terutama pada *cluster 1* dan *cluster 4* yang mendominasi jumlah pelanggan pada Toko Ritel Z.

#### Daftar Pustaka

- Adiana, B. E. Soesanti, I., Permanasari, A. E. (2018). Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering. *JUTEI*, Vol. 2 No. 1.
- Fong, S., Wong, R., & Vasilakos, A. V. (2015). *Accelerated PSO swarm search feature selection for data stream mining big data*. *IEEE transactions on services computing*, 9(1), 33-45.
- Hamali, Yusuf, A., & Anggriani, N. L. (2020). Pengaruh Promosi Penjualan Terhadap Keputusan Pembelian Produk Handuk Pada PT Tata Pusaka Sentosa Textile Mills Bandung. *YUME: Journal of Management*, 3(1).
- Kansal, T., Bahuguna, S., Singh, V., & Choudhury, T. (2018, December). *Customer segmentation using K-means clustering*. *International conference on computational techniques, electronics and mechanical systems (CTEMS)* (pp. 135-139). IEEE.
- Kodinariya, T. M., & Makwana, P. R. (2013). *Review on determining number of Cluster in K-Means Clustering*. *International Journal*, 1(6), 90-95.
- Li, Y., Cai, J., Yang, H., Zhang, J., & Zhao, X. (2019). *A novel algorithm for initial cluster center selection*. *iee Access*, 7, 74683-74693.
- Lubis, A. H. (2016). Model Segmentasi Pelanggan Dengan Kernel K-Means Clustering Berbasis *Customer Relationship Management*, *Jurnal dan Penelitian Teknik Informatika*, Vol. 1 No. 1.
- Muhidin, A. (2017). Analisa Metode Hierarchical Clustering Dan K-Mean Dengan Model Lrfmp Pada Segmentasi Pelanggan. *Jurnal SIGMA*, 8(3), 237-244.
- Murprawati, S. I., Indrawan, I. G. A., Aranta, A. (2021). Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail. *Jurnal Pendidikan Teknologi dan Kejuruan*, Vol. 18 No. 2.
- Septika, B. H., Krisnahadi, T., & Aryani, M. (2022). Strategi Pemasaran dalam Mempertahankan Eksistensi Bisnis Oleh-Oleh Khas Lombok Ditengan Pandemi Covid 19 (Studi Kasus Pada

- Toko Oleh-Oleh Sasak Maiq). *Jurnal Visionary: Penelitian dan Pengembangan dibidang Administrasi Pendidikan*, 10(1), 67-73.
- Sulistyowati, S. (2018). Analisa Segmentasi Konsumen Menggunakan Algoritma K-Means Clustering. *Prosiding Seminar Nasional Sains dan Teknologi Terapan* (pp. 51-58).
- Surjandari, I., & Seruni, A. (2010). *Design of product placement layout in retail shop using market basket analysis*. *Makara: Journal of Technology*, 9(2), 43-47.
- Triyansyah, D., & Fitriannah, D. (2018). Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing. *InComTech: Jurnal Telekomunikasi dan Komputer*, 8(3), 163-182.
- Widiyanto, A. T., & Witanti, A. (2021). Segmentasi Pelanggan Berdasarkan Analisis RFM menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global). *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, Vol. 1 No. 1.
- Zhu, E., & Ma, R. (2018). *An effective partitional clustering algorithm based on new clustering validity index*. *Applied soft computing*, 71, 608-621.