

Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Untuk Optimalisasi Penjualan Sembako

Adila Rahmawati¹, Tukino², Agustia Hananto³, Fitria Nurapriani⁴

Fakultas Ilmu Komputer, Program Studi Sistem Informasi

Universitas Buana Perjuangan Karawang

si22.adilarahmawati@mhs.ubpkarawang.ac.id

Abstrak

Sembako sebagai kebutuhan pokok memiliki peranan penting dalam kehidupan masyarakat. Dengan meningkatnya persaingan di sektor ini, pelaku usaha dituntut untuk memahami perilaku konsumen secara lebih mendalam agar strategi pemasaran yang dijalankan menjadi lebih efektif. Salah satu metode yang bisa dimanfaatkan adalah segmentasi pelanggan berbasis data transaksi. Penelitian ini bertujuan menerapkan algoritma K-Means Clustering untuk mengelompokkan pelanggan dalam rangka mendukung peningkatan penjualan sembako. Data yang digunakan meliputi empat variabel utama: jumlah pesanan, uang muka, total transaksi, dan pelunasan. Proses awal dilakukan melalui pembersihan dan normalisasi data menggunakan StandardScaler. Penentuan jumlah cluster terbaik dilakukan dengan metode Elbow dan dikonfirmasi menggunakan nilai Silhouette Score. Hasil penelitian menunjukkan bahwa empat cluster merupakan segmentasi paling optimal dengan nilai Silhouette Score sebesar 3,99. Cluster 2 memberikan kontribusi tertinggi terhadap total penjualan dan jumlah pelanggan sebesar 35,62%, disusul oleh cluster 1 sebesar 30,14%, cluster 3 sebesar 21,92%, sedangkan cluster 0 memiliki kontribusi terendah sebesar 12,33%. Visualisasi scatter plot menunjukkan distribusi antar-cluster yang cukup jelas. Segmentasi ini berguna untuk merancang strategi pemasaran yang lebih tepat sasaran.

Kata kunci: Segmentasi Pelanggan, K-Means Clustering, Elbow Method, Silhouette Score, Penjualan Sembako

Abstract

Basic necessities such as staple foods play an important role in people's lives. With increasing competition in this sector, business actors are required to understand consumer behavior in more depth so that the marketing strategies implemented become more effective. One method that can be utilized is customer segmentation based on transaction data. This study aims to apply the K-Means Clustering algorithm to group customers to support increased sales of necessities. The data used includes four main variables: number of orders, down payment, total transactions, and settlement. The initial process is carried out through data cleaning and normalization using StandardScaler. Determination of the best number of clusters is carried out using the Elbow method and confirmed using the Silhouette Score value. The results of the study showed that four clusters were the most optimal segmentation with a Silhouette Score value of 3.99. Cluster 2 contributed the highest to total sales and number of customers 35.62%, followed by Cluster 1 by 30.14%, cluster 3 by 21.92%, and Cluster 0 had the lowest contribution by 12.33%. Scatter plot visualization shows a fairly clear distribution between clusters. This segmentation is useful for designing more targeted marketing strategies.

Keywords: Customer Segmentation, K-Means Clustering, Elbow Method, Silhouette Score, Grocery Sales

PENDAHULUAN

Sembako (sembilan bahan pokok) merupakan kebutuhan dasar yang dibutuhkan oleh seluruh lapisan masyarakat. Karena sifatnya yang esensial, bisnis yang bergerak di bidang penjualan sembako memiliki peluang pasar yang luas namun juga dihadapkan pada tingkat persaingan yang tinggi. Untuk tetap kompetitif, pelaku usaha perlu memiliki pemahaman yang mendalam mengenai karakteristik dan perilaku pelanggan guna menyusun strategi pemasaran yang tepat sasaran (Fauzi & Priati, 2017).

Salah satu pendekatan yang berkembang dalam dunia bisnis dan teknologi informasi adalah penerapan analisis data untuk mendukung pengambilan keputusan (Lia Hananto et al., 2021). Dalam konteks pemasaran, segmentasi pelanggan berbasis data transaksi merupakan salah satu metode yang efektif untuk mengelompokkan pelanggan berdasarkan kesamaan perilaku (Awalina & Rahayu, 2023). Teknik segmentasi ini memungkinkan pelaku usaha untuk merancang promosi atau layanan yang sesuai dengan kebutuhan dan kebiasaan pelanggan di masing-masing segmen (Kotler & Keller, 2021).

Algoritma K-Means Clustering, sebagai salah satu metode *unsupervised learning*, telah banyak digunakan dalam proses segmentasi pelanggan karena kemampuannya dalam mengelompokkan data numerik secara efisien dan efektif (Ramadhan, 2023). Metode ini bekerja dengan membagi data ke dalam sejumlah cluster berdasarkan kemiripan data (Tukino & Huda, 2019), dengan tujuan untuk memaksimalkan homogenitas dalam cluster dan heterogenitas antar cluster (Bangoria et al., 2013).

Penggunaan K-Means dalam bisnis sembako dapat membantu pemilik toko untuk mengidentifikasi kelompok pelanggan berdasarkan transaksi mereka, seperti jumlah pesanan, pembayaran DP, total pembelian, dan pelunasan. Dengan segmentasi yang tepat, toko dapat merancang strategi promosi yang lebih spesifik dan meningkatkan retensi pelanggan.

Rumusan Masalah

1. Bagaimana menerapkan metode K-Means Clustering melakukan segmentasi pelanggan dalam rangka optimalisasi penjualan sembako?
2. Bagaimana strategi pemasaran yang sesuai untuk masing-masing segmen pelanggan?

Tujuan Penelitian

1. Menerapkan metode K-Means Clustering melakukan segmentasi pelanggan dalam rangka optimalisasi penjualan sembako?
2. Menentukan strategi pemasaran yang sesuai untuk masing-masing segmen pelanggan?

Manfaat Penelitian

1. Penelitian ini diharapkan dapat mempermudah pemilik toko sembako untuk menyusun strategi penjualan sesuai dengan segmen pelanggan toko.
2. Memberikan kontribusi keilmuan dalam penerapan data mining.

Tinjauan Pustaka

Penelitian oleh Pramudiansyah dan Munte menyoroti pemanfaatan algoritma K-Means yang dikombinasikan dengan pendekatan *Recency, Frequency, Monetary* (RFM) dalam tahap persiapan data untuk menilai nilai pelanggan secara lebih akurat.

Penelitian tersebut menggunakan empat atribut utama, yakni kode pelanggan, waktu transaksi terakhir, frekuensi pembelian, dan total pembelian. Berdasarkan metode Elbow, jumlah cluster optimal yang diperoleh adalah empat. Setelah penerapan algoritma K-Means, ditemukan bahwa 38,4% pelanggan tergolong dalam kluster Platinum (557 pelanggan), 12,6% dalam kluster Gold (184 pelanggan), 22,7% dalam kluster Silver (330 pelanggan), dan 26% dalam kluster Bronze (378 pelanggan) (Pramudiansyah & Munte, 2021).

Perdana dkk. melakukan studi segmentasi pelanggan pada platform Alfagift dengan mempertimbangkan lima faktor, yaitu usia, jenis kelamin, intensitas pembelian, metode pembayaran, dan kota domisili. Proses analisis menggunakan pendekatan CRISP-DM, dan algoritma K-Means digunakan untuk membentuk kelompok pelanggan. Evaluasi jumlah cluster optimal dilakukan dengan metode Elbow, menghasilkan tiga kluster dengan nilai Sum of Squared Error (SSE) tertinggi sebesar 1119,28644. Hasil akhir segmentasi menghasilkan tiga kelompok, masing-masing berisi 7.219 pelanggan, 6.902 pelanggan, dan 5.371 pelanggan. Kluster ini kemudian dimanfaatkan untuk mendukung strategi pemasaran yang lebih terarah (Perdana et al., 2022).

Rumapea dan tim meneliti segmentasi pelanggan menggunakan kombinasi model RFM dan algoritma K-Means Clustering. Model RFM digunakan untuk menilai pelanggan berdasarkan waktu terakhir transaksi (recency), jumlah transaksi (frequency), dan nilai transaksi (monetary). Setelah proses normalisasi data, dilakukan pengelompokan menggunakan K-Means. Hasilnya menunjukkan bahwa pelanggan terbagi ke dalam tiga segmen, yakni kluster

0 (40,72%), kluster 1 (41,23%), dan kluster 2 (18,05%). Segmentasi ini memungkinkan perusahaan menyusun strategi pemasaran yang lebih personal dan efektif sesuai karakteristik tiap kluster (Rumapea et al., 2024).

Dalam studi lain oleh Awalina dan Rahayu di lingkungan toko ritel, segmentasi pelanggan dilakukan dengan menggunakan algoritma K-Means dan metode Elbow. Data yang digunakan terdiri dari tiga atribut utama: kuantitas barang, harga satuan, dan ID pelanggan. Hasil klusterisasi menghasilkan empat kelompok pelanggan yang berbeda. Kluster 0 mencakup pelanggan dengan kuantitas rendah namun harga satuan tinggi dan ID besar. Kluster 1 terdiri atas pelanggan dengan kuantitas sedang, harga satuan rendah, dan ID tinggi. Kluster 2 mengelompokkan pelanggan dengan volume pembelian tinggi, harga satuan rendah, dan ID sedang. Terakhir, Kluster 3 mencakup pelanggan dengan pembelian sedikit, harga tinggi, dan ID sedang. Studi ini menunjukkan bahwa kuantitas dan harga satuan sangat berpengaruh dalam perilaku pembelian pelanggan (Awalina & Rahayu, 2023).

LANDASAN TEORI

1. Segmentasi Pelanggan

Segmentasi pelanggan merupakan proses untuk mengelompokkan konsumen ke dalam beberapa kelompok berdasarkan karakteristik, kebutuhan, atau perilaku yang serupa. Tujuan utamanya adalah agar perusahaan dapat menyediakan produk, layanan, dan strategi pemasaran yang sesuai dengan masing-masing segmen. Dengan segmentasi yang tepat, perusahaan dapat meningkatkan efektivitas komunikasi, efisiensi biaya pemasaran, dan

loyalitas pelanggan. Pendekatan segmentasi biasanya didasarkan pada variabel demografis, geografis, psikografis, dan perilaku (Kotler & Keller, 2021)

2. K-Means Clustering

K-Means Clustering adalah algoritma unsupervised learning yang digunakan untuk mengelompokkan data ke dalam beberapa klaster berdasarkan tingkat kemiripan (Kastiawan et al., 2024). Prosesnya dimulai dengan menentukan jumlah cluster (k) terlebih dahulu, lalu algoritma akan mengatur titik pusat awal (centroid), dan secara iteratif menyesuaikan posisi centroid hingga data dalam satu cluster saling mendekati pusat cluster-nya. Algoritma ini banyak digunakan dalam data mining karena kemudahannya dalam pengelompokan data skala besar (Han et al., 2012). Meskipun algoritma K-Means populer digunakan dalam pengelompokan data, algoritma ini memiliki sejumlah keterbatasan. Di antaranya adalah kepekaan terhadap pemilihan titik awal (initial centroid), ketergantungan pada jumlah klaster yang telah ditentukan sebelumnya, serta performa yang kurang optimal ketika diterapkan pada data dengan pola distribusi kompleks. Untuk mengatasi kekurangan tersebut, berbagai penelitian telah mengusulkan perbaikan, seperti penerapan teknik inisialisasi centroid yang lebih akurat dan integrasi dengan metode klasterisasi lain melalui pendekatan hibrida (Qtaish et al., 2024).

3. Metode Elbow

Metode Elbow digunakan untuk menentukan jumlah cluster optimal dalam algoritma seperti K-Means. Teknik ini bekerja dengan menghitung nilai *Within-Cluster Sum of Squares* (WCSS) untuk berbagai jumlah cluster, lalu memvisualisasikannya dalam grafik. Titik

optimal ditentukan pada posisi “siku” atau elbow dari kurva, yaitu ketika penurunan nilai WCSS tidak lagi signifikan seiring bertambahnya jumlah cluster. Titik ini merepresentasikan jumlah cluster yang cukup efisien untuk menggambarkan struktur data (Syakur et al., 2018)

4. Silhouette Score

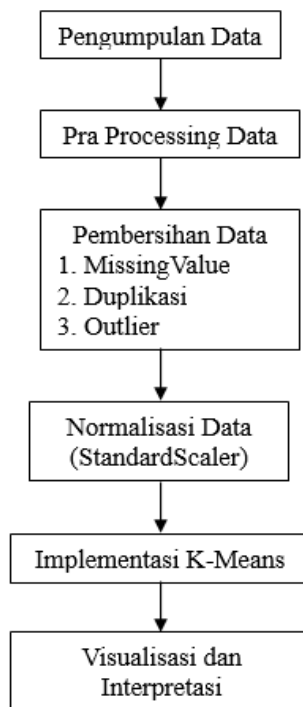
Silhouette Score adalah metrik evaluasi yang digunakan untuk mengukur kualitas hasil klasterisasi. Nilai ini dihitung berdasarkan seberapa dekat satu data dengan data lain dalam klaster yang sama dibandingkan dengan data di klaster lain. Skor Silhouette berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan pemisahan klaster yang baik, sedangkan nilai mendekati 0 menunjukkan tumpang tindih antar klaster. Metrik ini sangat berguna untuk mengevaluasi efektivitas algoritma seperti K-Means dalam membentuk kelompok yang terpisah dan kohesif (Shahapure & Nicholas, 2020).

METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif deskriptif dengan tujuan mengidentifikasi pola segmentasi pelanggan berdasarkan data transaksi historis (Sugiyono, 2019). Pendekatan kuantitatif dipilih karena fokus utama penelitian adalah analisis numerik menggunakan algoritma K-Means Clustering untuk pengelompokan data.

Data yang digunakan dalam penelitian ini adalah data penjualan sembako. Setelah data diperoleh akan dilakukan pembersihan data kembar dan data kosong serta akan dilakukan normalisasi data agar data lebih mudah diolah menggunakan metode StandardScaler. Untuk penentuan jumlah cluster yang digunakan menggunakan metode Elbow dan *Silhouette Score*.

Alur Penelitian



Gambar 1. Alur Penelitian

ANALISIS DAN PERANCANGAN

1. Pengumpulan data

Data yang digunakan berasal dari catatan transaksi pelanggan dalam rentang waktu tertentu. Variabel yang dianalisis mencakup jumlah pesanan, uang muka, total nilai transaksi, dan pelunasan. Penggunaan data historis memberikan landasan analisis berbasis fakta nyata yang menggambarkan kebiasaan pembelian pelanggan.

2. Pra-pemrosesan Data

Tahapan ini meliputi pembersihan data dari nilai kosong, duplikasi, serta outlier. Setelah itu, data dinormalisasi menggunakan metode StandardScaler, yang mengubah skala data numerik menjadi rata-rata nol dan deviasi standar satu (Rachmawati et al., 2024). Normalisasi diperlukan agar variabel dengan nilai besar tidak mendominasi proses klusterisasi.

3. Penentuan Jumlah Cluster

Jumlah kluster yang ideal ditentukan dengan dua pendekatan:

- Metode Elbow, dengan memplot *Within-Cluster Sum of Squares* (WCSS) terhadap jumlah kluster, lalu menentukan titik "siku" sebagai kluster optimal.
 - Silhouette Score*, yang menilai seberapa baik suatu data cocok dengan klusternya dibandingkan kluster lain. Nilai mendekati 1 menunjukkan segmentasi yang baik
4. Penerapan K-Means Clustering

Setelah diperoleh jumlah kluster optimal, algoritma K-Means digunakan untuk membagi pelanggan ke dalam segmen-segmen berdasarkan kesamaan data. Algoritma ini dimulai dengan inisialisasi centroid secara acak, kemudian menghitung jarak Euclidean dan mengelompokkan data berdasarkan centroid terdekat secara iteratif hingga mencapai konvergensi

5. Visualisasi dan interpretasi

Setelah proses klusterisasi selesai, hasilnya divisualisasikan dalam scatter plot untuk menunjukkan sebaran data tiap kluster. Visualisasi ini menampilkan dua dari empat variabel (contohnya: total pembayaran dan DP), dengan pewarnaan berbeda untuk setiap kluster. Analisis lanjutan seperti rata-rata atau median pada masing-masing kluster digunakan untuk mengidentifikasi karakteristik pelanggan dalam setiap segmen.

IMPLEMENTASI DAN PEMBAHASAN

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data transaksi penjualan toko sembako pada bulan April 2025 dengan jumlah transaksi sebanyak 555 transaksi dengan total 73 pelanggan.

Tabel 1. Statistik Data Penjualan

Statistik	Jumlah
Jumlah Transaksi	555
Jumlah Pelanggan	73
Jumlah Unit Terjual	16761
Jumlah Penjualan	RP 1.368.746.300
Jumlah DP	Rp 593.500.000
Jumlah Pelunasan	Rp 775.246.300

2. Pra-pemrosesan Data

a. Cek data kosong

```
checking for missing values:
kode_transaksi      0
kd_pelanggan        0
tanggal_transaksi    0
total_unit           0
total                0
DP                   0
Pelunasan            0
dtype: int64
```

Gambar 1. Hasil Cek Data Kosong

Tidak ada data kosong pada data transaksi penjualan toko sembako.

b. Cek data kembar

```
checking for duplicate rows:
0

No duplicate rows found.
```

Gambar 2. Hasil Cek Data Kembar

Tidak ada data kembar pada data transaksi penjualan toko sembako.

c. Pemilihan Features

Features yang digunakan dalam penelitian ini adalah jum_trans, total penjualan, DP, dan pelunasan berdasarkan kode pelanggan.

	jum_trans	total	dp	pelunasan
0	7	14420500	7200000	7220500
1	8	17228700	10300000	6928700
2	4	10191900	3050000	7141900
3	7	14517900	8700000	5817900
4	7	19517700	5850000	13667700
..	...	...	...	...
68	10	19593200	9750000	9843200
69	8	20907900	12500000	8407900
70	7	18604000	5550000	13054000
71	7	19231900	5750000	13481900
72	8	23483200	7300000	16183200

Gambar 3. Hasil Pemilihan Features

d. Normalisasi dengan metode StandardScaler

Setelah pemilihan features, data akan dinormalisasi menggunakan metode StandardScaler.

Hasil Normalisasi:

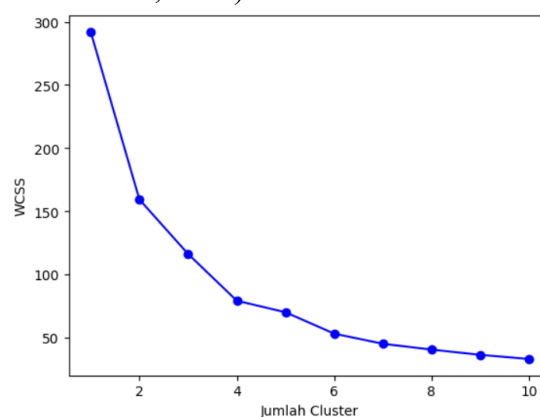
	jum_trans	total	dp	pelunasan
0	-0.372160	-0.901719	-0.315883	-0.972185
1	0.245288	-0.316839	0.736904	-1.055639
2	-2.224504	-1.782433	-1.725259	-0.994665
3	-0.372160	-0.881433	0.193530	-1.373322
4	-0.372160	0.159904	-0.774355	0.871680
..	...	...	...	...
68	1.480183	0.175629	0.550120	-0.222107
69	0.245288	0.449449	1.484044	-0.632595
70	-0.372160	-0.030398	-0.876237	0.696165
71	-0.372160	0.100379	-0.808315	0.818542
72	0.245288	0.985821	-0.281922	1.591100

Gambar 4. Hasil Normalisasi Data

3. Penentuan Jumlah Kluster

a. Metode Elbow

Metode Elbow digunakan untuk mendapatkan nilai cluster optimal untuk menghasilkan segmen yang efektif (Atira & Nurina Sari, 2023).



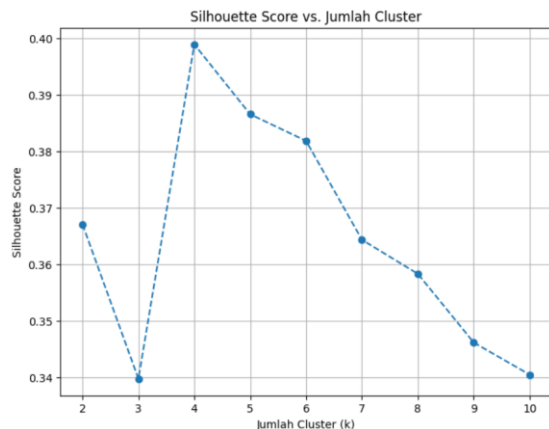
Gambar 5. Grafik Elbow Method

Pada gambar 5 terlihat bahwa penurunan nilai WCSS cukup tajam di awal, namun laju penurunannya mulai melambat setelah mencapai jumlah cluster tertentu. Titik siku (elbow point) yang menunjukkan jumlah cluster optimal tampak pada $K = 4$. Setelah titik ini, penambahan cluster tidak memberikan pengurangan WCSS yang berarti, sehingga $K = 4$ dipilih sebagai jumlah cluster yang paling sesuai untuk analisis ini.

b. Silhouette Score

Silhouette Score digunakan untuk mengevaluasi keakuratan hasil dari Metode Elbow. Skor ini menunjukkan seberapa dekat data dalam satu cluster dibandingkan

dengan jaraknya terhadap data di cluster lain.



Gambar 6. Grafik Silhouette Score

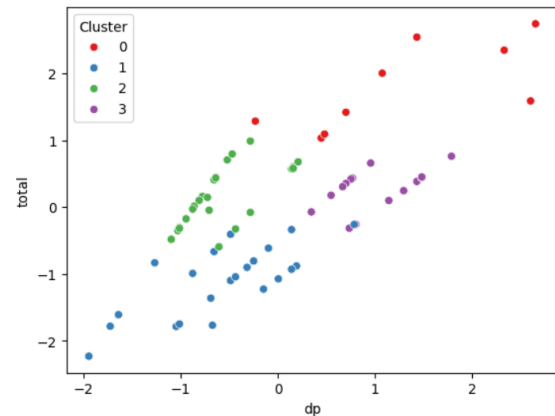
Berdasarkan Gambar 6, dapat dilihat bahwa nilai Silhouette Score tertinggi diperoleh saat $K = 4$. Temuan ini sejalan dengan hasil dari metode Elbow, yang menunjukkan bahwa membagi pelanggan ke dalam 4 segmen merupakan pendekatan paling optimal dalam membedakan pola transaksi mereka.

4. Penerapan K-Means Clustering

Setelah ditetapkan bahwa jumlah cluster optimal adalah 4 ($K = 4$), algoritma K-Means Clustering digunakan untuk mengelompokkan data transaksi pelanggan. Proses ini mencakup:

- Inisialisasi centroid sebanyak 4 buah sesuai jumlah cluster
- Pengelompokan data berdasarkan jarak terdekat ke masing-masing centroid
- Iterasi yang dilakukan hingga posisi centroid stabil atau jumlah iterasi maksimum tercapai.

Hasil dari penerapan K-Means Clustering menunjukkan bahwa data pelanggan berhasil dibagi ke dalam empat kelompok berbeda. Informasi mengenai cluster ini kemudian ditambahkan ke dataset sebagai kolom baru dengan nama Cluster.



Gambar 7. Visualisasi Hasil K-Means

Gambar 7 menunjukkan hasil penerapan algoritma K-Means Clustering terhadap data transaksi pelanggan, dengan dua variabel utama yaitu DP (uang muka) pada sumbu horizontal dan total transaksi pada sumbu vertikal. Setiap titik dalam grafik merepresentasikan satu pelanggan, dan warnanya menunjukkan keanggotaan cluster berdasarkan hasil pengelompokan K-Means. Terdapat empat cluster yang terbentuk, masing-masing diwakili oleh warna merah, biru, hijau, dan ungu. Pola penyebaran data menunjukkan adanya hubungan positif antara DP dan total transaksi, di mana pelanggan dengan DP lebih besar cenderung memiliki total transaksi yang lebih tinggi. Keempat cluster tampak terpisah cukup jelas, menandakan bahwa algoritma berhasil membagi pelanggan ke dalam segmen yang memiliki karakteristik berbeda. Informasi ini dapat dimanfaatkan untuk analisis segmentasi pelanggan dan pengambilan keputusan strategis, seperti pemberian penawaran khusus yang disesuaikan dengan perilaku masing-masing kelompok pelanggan.

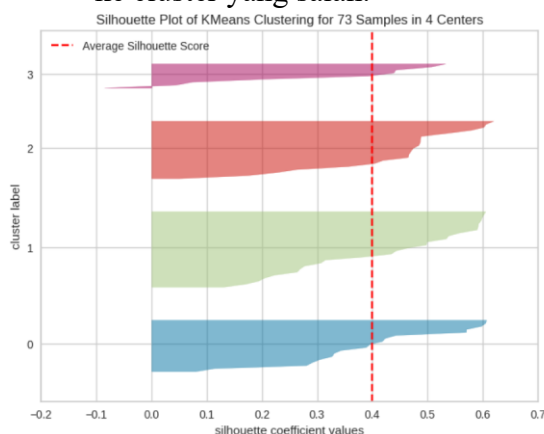
5. Visualisasi dan Interpretasi

a. Evaluasi Silhouette Plot

Evaluasi terhadap hasil clustering bertujuan untuk mengukur seberapa baik kualitas pengelompokan yang dihasilkan oleh algoritma K-Means. Metrik ini menilai

sejauh mana suatu data berada lebih dekat ke anggota dalam cluster yang sama dibandingkan dengan anggota di cluster berbeda. Silhouette Score memiliki rentang nilai dari -1 hingga 1, dengan penjelasan sebagai berikut:

- 1) Nilai mendekati 1 mengindikasikan bahwa pengelompokan sangat baik, di mana data dalam satu cluster cenderung seragam dan terpisah jauh dari cluster lainnya.
- 2) Nilai sekitar 0 menunjukkan bahwa data berada di area perbatasan antara dua cluster.
- 3) Nilai negatif menandakan bahwa data mungkin telah dikelompokkan ke cluster yang salah.



Gambar 8. Silhouette Plot Average

Gambar di atas merupakan Silhouette Plot yang digunakan untuk mengevaluasi kualitas hasil klusterisasi K-Means terhadap 73 data yang dibagi ke dalam 4 cluster. Grafik ini menggambarkan nilai koefisien Silhouette untuk masing-masing data pada setiap cluster, yang mengukur sejauh mana sebuah data cocok berada dalam clusternya sendiri dibandingkan dengan cluster lain.

Setiap warna horizontal merepresentasikan satu cluster, dan panjang batang menggambarkan nilai koefisien Silhouette untuk setiap data dalam cluster tersebut. Nilai koefisien

berkisar antara -1 hingga 1. Dalam plot ini, sebagian besar nilai berada di atas 0, yang berarti data lebih dekat ke cluster-nya sendiri dibanding cluster lain. Garis merah putus-putus menunjukkan rata-rata nilai Silhouette keseluruhan, yang berada sekitar 0.399. Nilai ini termasuk dalam kategori cukup baik, menunjukkan bahwa struktur kluster cukup terbentuk.

b. Analisa hasil Cluster

Setelah pengelompokan data pelanggan menggunakan algoritma K-Means dilakukan, tahap selanjutnya adalah menganalisis karakteristik dari setiap cluster. Tujuan dari analisis ini adalah untuk memahami pola transaksi yang muncul dalam masing-masing kelompok, seperti rata-rata jumlah pembelian, harga rata-rata, serta total nilai transaksi. Tabel 1 menyajikan ringkasan karakteristik utama tiap cluster.

Tabel 2. Karakteristik Cluster

Cluster	0	1	2	3
Transaksi	92	130	201	132
Penjualan	245,736,000	295,508,900	505,867,100	321,634,300
DP	107,050,000	142,150,000	165,900,000	178,400,000
Pelunasan	138,686,000	153,358,900	339,967,100	143,234,300
Persentase Pelanggan (%)	12.33	30.14	35.62	21.92

Hasil pengelompokan data pelanggan menghasilkan empat cluster dengan karakteristik yang berbeda. Berdasarkan hasil clustering, pelanggan toko sembako dikelompokkan menjadi empat klaster dengan karakteristik yang berbeda. Klaster 2 merupakan kelompok paling aktif, dengan jumlah transaksi tertinggi (201 transaksi) dan total penjualan terbesar, mencapai lebih dari 505 juta rupiah. Mereka menyumbang sekitar 35,62% dari total pelanggan dan menunjukkan komitmen tinggi dalam pembayaran, dengan pelunasan sebesar 339 juta rupiah. Klaster 1 juga menunjukkan performa yang baik, dengan 130 transaksi dan kontribusi

penjualan lebih dari 295 juta rupiah, serta presentase pelanggan sebesar 30,14%. Klaster 3 menempati posisi ketiga, dengan jumlah transaksi dan pelunasan yang lebih rendah, meskipun penjualannya tetap signifikan. Sementara itu, klaster 0 merupakan kelompok paling kecil dengan persentase pelanggan terendah (12,33%) dan total transaksi serta pelunasan yang lebih rendah dibanding klaster lainnya.

c. Rekomendasi Pemasaran

Berdasarkan hasil segmentasi pelanggan menggunakan algoritma K-Means, strategi pemasaran dapat disesuaikan dengan karakteristik masing-masing cluster. Strategi pemasaran dapat disesuaikan dengan karakteristik masing-masing klaster. Untuk klaster 2, yang merupakan pelanggan paling potensial, strategi loyalitas seperti pemberian diskon eksklusif atau program poin dapat diterapkan guna menjaga dan meningkatkan retensi. Klaster 1 dapat dijadikan sasaran promosi produk baru atau upselling, karena menunjukkan minat beli tinggi. Klaster 3 memerlukan pendekatan personal seperti reminder pembayaran atau penawaran angsuran yang lebih fleksibel untuk meningkatkan pelunasan. Sementara itu, klaster 0 sebaiknya menjadi fokus dalam program edukasi produk dan promosi khusus untuk mendorong lebih banyak transaksi serta membangun loyalitas jangka panjang. Pendekatan ini memungkinkan peningkatan performa penjualan secara keseluruhan berdasarkan segmentasi pelanggan yang tepat sasaran.

KESIMPULAN

1. Metode K-Means Clustering berhasil diterapkan untuk segmentasi pelanggan dalam rangka optimalisasi penjualan sembako. Proses dimulai dengan pengolahan data transaksi pelanggan,

termasuk pembersihan dan normalisasi data menggunakan StandardScaler. Penentuan jumlah cluster optimal dilakukan dengan pendekatan Elbow Method dan Silhouette Score, yang menunjukkan bahwa empat segmen merupakan pemisahan yang paling representatif. Dengan pembagian ini, pelanggan dikelompokkan berdasarkan kesamaan dalam perilaku transaksi seperti jumlah pesanan, total penjualan, uang muka, dan pelunasan. Hasil segmentasi memberikan gambaran yang jelas mengenai variasi karakteristik pelanggan dan dapat dijadikan acuan dalam strategi bisnis.

2. Strategi pemasaran yang relevan untuk setiap segmen dapat dirancang berdasarkan karakteristik masing-masing cluster. Cluster 2, yang memiliki kontribusi paling tinggi dalam hal jumlah transaksi dan total penjualan, dapat difokuskan dalam program loyalitas atau penawaran eksklusif guna mempertahankan pelanggan. Cluster 1 dapat dijadikan sasaran promosi produk baru atau upselling, karena menunjukkan minat beli tinggi. Klaster 3 memerlukan pendekatan personal seperti reminder pembayaran atau penawaran angsuran yang lebih fleksibel untuk meningkatkan pelunasan. Sementara itu, Cluster 0 yang memiliki kontribusi paling rendah, fokus dalam program edukasi produk dan promosi khusus untuk mendorong lebih banyak transaksi serta membangun loyalitas jangka panjang. Pendekatan pemasaran yang disesuaikan ini akan mendorong peningkatan efektivitas penjualan secara keseluruhan.

Saran

1. Untuk meningkatkan efektivitas pemasaran, pemilik usaha disarankan untuk menerapkan pendekatan yang disesuaikan dengan masing-masing segmen pelanggan. Misalnya, program loyalitas dan insentif eksklusif dapat diberikan kepada pelanggan di Cluster 3 untuk mempertahankan keterlibatan mereka, sementara strategi promosi atau diskon bisa difokuskan pada pelanggan di Cluster 0 untuk mendorong peningkatan transaksi.
2. Penelitian ini masih terbatas pada empat variabel transaksi. Untuk penelitian selanjutnya, disarankan agar mempertimbangkan variabel tambahan seperti frekuensi kunjungan, jenis produk yang dibeli, atau waktu transaksi, agar segmentasi pelanggan dapat dilakukan dengan lebih komprehensif dan mendalam.

DAFTAR PUSTAKA

- Atira, A., & Nurina Sari, B. (2023). *Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk Penentuan Cluster Optimum dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan*. Jurnal Ilmiah Wahana Pendidikan, 9(17).
- Awalina, E. F. L., & Rahayu, W. I. (2023). *Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail*. Jurnal Teknologi Dan Informasi, 13(2). <https://doi.org/10.34010/jati.v13i2.10090>
- Bangoria, B., Mankad, N., & Pambhar, V. (2013). *A survey on Efficient Enhanced K-Means Clustering Algorithm*. IJSRD -International Journal for Scientific Research & Development, 1(9online).
- Fauzi, A., & Priati. (2017). *Data Mining dengan Teknik Clustering Menggunakan Algoritma K-Means pada Data Transaksi Superstore*. Seminar Nasional Informatika Dan Aplikasinya (SNIA), September.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. In *Data Mining: Concepts and Techniques*. <https://doi.org/10.1016/C2009-0-61819-5>
- Kastiawan, N., Huda, B., Novalia, E., & Nurapriani, F. (2024). *Klasterisasi Data Obat dengan Algoritma K-Means (Kasus pada UPTD Puskesmas Curug)*. Jurnal Sains Komputer & Informatika (J-SAKTI, 8(1), 120–130.
- Kotler, P., & Keller, K. L. (2021). *Marketing Management 16th Global Edition*. Pearson Education Limited.
- Lia Hananto, A., Assiroj, P., Priyatna, B., Nurhayati, Fauzi, A., Yuniar Rahman, A., & Shofiah Hilabi, S. (2021). *Analysis of Drug Data Mining with Clustering Technique Using K-Means Algorithm*. Journal of Physics: Conference Series, 1908(1). <https://doi.org/10.1088/1742-6596/1908/1/012024>
- Perdana, S. A., Florentin, S. F., & Santoso, A. (2022). *Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift*. Sebatik, 26(2). <https://doi.org/10.46984/sebatik.v26i2.1991>
- Pramudiansyah, A., & Munte, H. (2021). *Segmentasi Pelanggan Menggunakan Algoritma K-Means Berdasarkan Model Recency Frequency Monetary*. Jurnal Ilmiah Ilmu Komputer, 7(2), 06–19. <https://doi.org/10.35329/jiik.v7i2.201>
- Qtaish, A., Braik, M., Albashish, D., Alshammari, M. T., Alreshidi, A., & Alreshidi, E. J. (2024). *Optimization Of K-Means Clustering Method Using Hybrid Capuchin Search Algorithm*.

- Journal of Supercomputing, 80(2).
<https://doi.org/10.1007/s11227-023-05540-5>
- Rachmawati, F., Jaenudin, J., Ginting, N. B., & Laksono, P. (2024). *Machine Learning for the Model Prediction of Final Semester Assessment (FSA) using the Multiple Linear Regression Method*. Jurnal Teknik Informatika, 17(1), 1–9.
<https://doi.org/10.15408/jti.v17i1.28652>
- Ramadhan, A. G. (2023). *Data Mining untuk Segmentasi Pelanggan dengan Algoritma K-Means: Studi Kasus pada Data Pelanggan di Toko Retail*. Syntax Literate ; Jurnal Ilmiah Indonesia, 8(10).
<https://doi.org/10.36418/syntax-literate.v8i10.13746>
- Rumapea, A. Y. N., Pratiwi, D., & Sari, S. (2024). *Analisis Segmentasi Pelanggan Ritel Online Menggunakan K-Means Clustering Berdasarkan Model Recency , Frequency , Monetary (RFM)*. Jurnal Sains Dan Teknologi, 6(3), 292–299.
<https://doi.org/DOI:https://doi.org/10.55338/saintek.v6i3.4607>
- Shahapure, K. R., & Nicholas, C. (2020). Cluster Quality Analysis Using Silhouette Score. Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020.
<https://doi.org/10.1109/DSAA49011.2020.00096>
- Sugiyono. (2019). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif dan R&D*. In Bandung: CV Alfabeta.
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). *Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster*. IOP Conference Series: Materials Science and Engineering, 336(1).
<https://doi.org/10.1088/1757-899X/336/1/012017>
- Tukino, T., & Huda, B. (2019). *Penerapan Algoritma K-Means Untuk Mendukung Keputusan Dalam Pemilihan Tema Tugas Akhir Pada Prodi Sistem Informasi Universitas Buana Perjuangan Karawang*. Techno Xplore : Jurnal Ilmu Komputer Dan Teknologi Informasi, 4(1).
<https://doi.org/10.36805/technoexplo.v4i1.542>