

Segmentasi Pelanggan Berbasis Clustering Menggunakan LRFM dan Variabel Transaksi Untuk Mendukung Strategi Pemasaran

Andini Pramesti^{1)*}, I Kadek Dwi Nuryana²⁾

^{1,2)} Universitas Negeri Surabaya, Fakultas Teknik, Teknik Informatika, Program Studi Sistem Informasi

Received: 28 April 2026

Accepted: 10 May 2026

Published: 1 June 2026



*andini.22092@mhs.unesa.ac.id

Kata Kunci: Segmentasi Pelanggan, LRFM, KDD, Clustering, Customer Loyalty Matrix, Dashboard.

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : CV. Restu Tani Jaya merupakan salah satu pelaku usaha yang bergerak di sektor pertanian yang berlokasi di Nganjuk, Jawa Timur. Berdasarkan hasil wawancara CV. Restu Tani Jaya mengalami permasalahan dalam pengelolaan data dan strategi pemasaran yang kurang tepat sasaran. Data penjualan online CV. Restu Tani Jaya saat ini digunakan untuk arsip laporan penjualan saja dan belum diolah secara mendalam untuk menghasilkan informasi penting dalam memahami perilaku pelanggan. Oleh karena itu, diperlukan segmentasi pelanggan berdasarkan Length, Recency, Frequency, dan Monetary (LRFM) dan variabel perilaku transaksi tambahan. Data yang digunakan yaitu data penjualan online dari bulan Januari 2022 hingga September 2025 dengan total 2.393 baris. Penelitian dilakukan dengan mengikuti tahapan Knowledge Discovery in Database (KDD) dengan membandingkan tiga algoritma clustering yaitu K-Means, K-Medoids, dan Agglomerative Hierarchical Clustering (AHC). Hasil algoritma yang menghasilkan performa yang paling stabil adalah algoritma K-Means yaitu 2 cluster dengan nilai Silhouette Score adalah 0.51161 dan Davies Bouldin Index adalah 0.77046 dengan berdasarkan Customer Loyalty Matrix dan Customer Value Matrix cluster 0 merupakan kelompok pelanggan Core-Spender dengan strategi yang diberikan strategi retensi, sedangkan kelompok pelanggan pada cluster 1 merupakan Lost-Frequent dengan strategi reaktivasi. Hasil dari segmentasi disajikan dalam visualisasi dashboard untuk mempermudah perusahaan dalam pengambilan keputusan.

PENDAHULUAN

Perkembangan teknologi yang pesat saat ini memberikan pengaruh besar pada kehidupan manusia, bukan hanya membuat perubahan dalam hal sosial dan budaya, tetapi juga dalam berbisnis. CV. Restu Tani Jaya merupakan salah satu pelaku usaha yang bergerak di sektor pertanian yang didirikan pada tahun 2020 bertempat di Nganjuk, Jawa Timur. Usaha ini menjual berbagai sarana produksi pertanian seperti pupuk, insektisida, fungisida, herbisida, nutrisi untuk tanaman dan lain-lain. Namun, berdasarkan hasil wawancara CV. Restu Tani Jaya mengalami permasalahan dalam pengelolaan data pelanggan dan strategi pemasaran. Data penjualan online CV. Restu Tani Jaya saat ini digunakan untuk arsip laporan penjualan saja dan belum diolah secara mendalam untuk menghasilkan informasi penting dalam memahami perilaku pelanggan. Selain itu, strategi pemasaran seperti pemberian diskon dan ongkos kirim dilakukan secara umum, di mana semua pelanggan diperlakukan sama. CV. Restu Tani Jaya tidak memahami perilaku pelanggan yang mengakibatkan risiko kehilangan pelanggan dan pemborosan biaya pemasaran. Padahal mempertahankan pelanggan adalah tantangan suatu usaha berkembang karena mendapatkan pelanggan baru lebih sulit dan mengeluarkan biaya yang lebih besar [1].

Salah satu solusi yang dapat diterapkan yaitu segmentasi pelanggan dengan analisis LRFM pengembangan dari model RFM. Penambahan *Length* dikarenakan pada RFM klasik tidak mempertimbangkan loyalitas pelanggan dikarenakan RFM tidak menghitung lamanya hubungan antara pelanggan dan perusahaan [2]. Dalam penelitian ini juga diperlukan variabel lain yang digunakan seperti untuk melihat perilaku sensitivitas terhadap diskon, jumlah barang per transaksi, keragaman produk yang dibeli, dan beban ongkos kirim yang ditanggung. Pengelompokan pelanggan dilakukan dengan metode *data mining* bertujuan untuk menemukan pola dan informasi yang tidak terlihat pada data kompleks yang nantinya dapat memberikan wawasan mendalam untuk bisnis [3]. Teknik analisis *data mining* yang digunakan yaitu *clustering*, dimana teknik ini digunakan untuk melakukan segmentasi pelanggan yang otomatis dengan mengelompokkan data yang ada untuk mengetahui karakteristik tiap kelompok pelanggan [4]. Algoritma *clustering* yang dipakai adalah K-Means, K-Medoids, dan Agglomerative Hierarchical Clustering. Beberapa penelitian berikut menggunakan data mining untuk segmentasi pelanggan salah satunya penelitian oleh Syukron et al. melakukan segmentasi pelanggan reseller kecantikan dengan menggunakan tiga algoritma yaitu K-Means, K-Medoids, dan Fuzzy C-Means berdasarkan LRFM algoritma terbaik K-means dibuktikan dengan nilai DBI 0,167 pada jumlah 6 kluster [5]. Berikutnya penelitian oleh Ramadhan et al. melakukan segmentasi pelanggan retail dengan menggunakan dua algoritma K-Means dan Agglomerative Clustering berdasarkan RFM dengan algoritma terbaik AHC dibuktikan dengan nilai silhouette score 0,6363 terbentuk 3 kluster [6].

Sejumlah penelitian sebelumnya sudah melakukan perbandingan beberapa metode algoritma clustering untuk segmentasi pelanggan, tetapi hanya sebatas membandingkan performanya. Pada penelitian ini tidak hanya melakukan evaluasi tetapi juga menggabungkan variabel LRFM dan variabel perilaku transaksi agar segmentasi yang dihasilkan lebih kontekstual. Penelitian ini juga melakukan interpretasi hasil segmentasi menggunakan Customer Loyalty Matrix dan Customer Value Matrix kemudian diterjemahkan ke dalam rekomendasi strategi pemasaran serta adanya visualisasi dashboard untuk mendukung pengambilan keputusan.

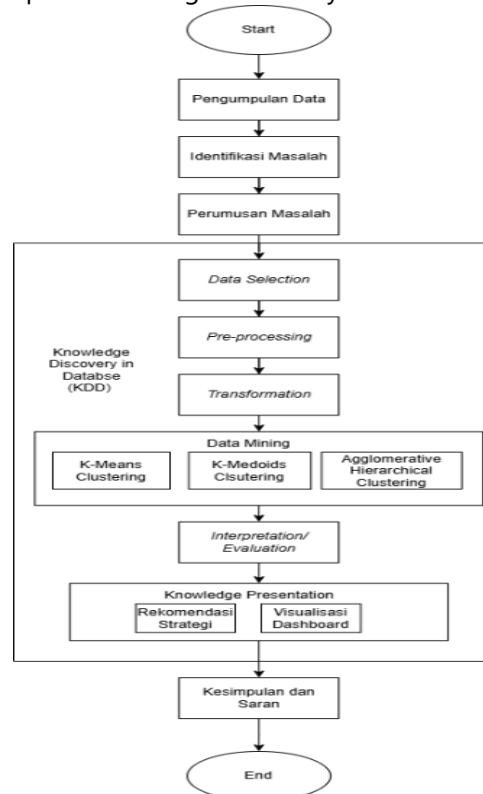
TINJAUAN LITERATUR

Pada penelitian sebelumnya segmentasi pelanggan dengan LRFM dengan *clustering* banyak digunakan diantaranya penelitian oleh Syukron et al. mengelompokkan pelanggan reseller kecantikan dengan LRFM dan membandingkan K-Means, K-Medoids, dan Fuzzy C-Means [5] untuk menentukan metode terbaik untuk segmentasi pelanggan, sedangkan penelitian oleh Ramadhan et al. membandingkan dua algoritma K-Means dan Agglomerative Clustering menggunakan metrik evaluasi Silhouette Score. Selain evaluasi algoritma penelitian sebelumnya juga telah melakukan analisis karakteristik dari cluster yang terbentuk yaitu penelitian oleh Nikmah et al. penelitian ini terbentuk empat *cluster* pelanggan berdasarkan tingkat loyalitasnya yaitu *Premium Loyalty*, *Inertia Loyalty*, *Latent Loyalty* dan *No Loyalty*. Masing-masing *cluster* memiliki karakteristik, diantaranya *Premium Loyalty* memiliki hubungan dan pembelian yang tinggi, *Inertia Loyalty* tingkat loyalitasnya rendah, tetapi pembelian ulangnya besar, *Latent Loyalty* memiliki loyalitas tinggi dan pembelian ulang yang rendah, dan *No Loyalty* memiliki loyalitas dan pembelian ulang yang rendah [7]. Berbeda dengan penelitian yang lain, penelitian oleh Zhou et al. melakukan analisis lajutan seperti demografi, geografis, dan produk setiap kluster yang akan lebih mudah memahami karakteristik pelanggan yang mendalam, sehingga penyusunan strategi pemasaran tepat sasaran [8]. Penelitian oleh Viandari et al. telah menghubungkan hasil clustering dengan Customer Loyalty Matrix dan Customer Value Matrix untuk menginterpretasi hasil clustering dengan memetakan pelanggan berdasarkan loyalitas dan nilai ekonominya yang digunakan [9].

Berdasarkan dari penelitian sebelumnya, masih jarang penelitian lain yang mengombinasikan LRFM dengan variabel transaksi tambahan dan menerjemahkan hasil clustering menjadi rekomendasi strategi pemasaran dan visualisasi dashboard. Oleh karena itu, penelitian ini melakukan segmentasi pelanggan dengan mengombinasikan variabel LRFM dan variabel perilaku tambahan serta menginterpretasikan hasil dengan Customer Loyalty Matrix dan Customer Value Matrix untuk menyusun strategi pemasaran dan menyajikan hasil dalam bentuk visualisasi dashboard.

METODE PENELITIAN

Penelitian mengikuti tahapan Knowledge Discovery in Database (KDD) yang terdiri enam fase



Gambar 1. Alur Penelitian

utama, berikut penjelasan tahapan alur penelitian diatas:

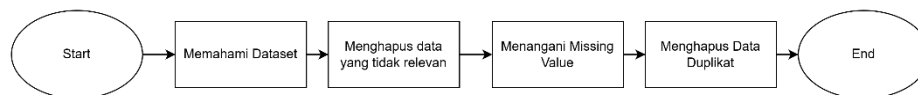
1. Pengumpulan Data

Pengumpulan data dilakukan wawancara dengan pihak CV. Restu Tani Jaya untuk mengidentifikasi berbagai kendala dalam pengelolaan pelanggan dan kebutuhan informasi dalam pengambilan keputusan. Data yang dipakai untuk analisis adalah data yang disesuaikan dengan topik dan batasan permasalahan pada penelitian ini

2. Data Selection

Pada data transaksi penjualan CV. Restu Tani Jaya terdapat banyak informasi dan variabel. Pada tahap ini dilakukan seleksi data untuk memilih dan menentukan variabel-variabel yang sesuai dengan kebutuhan penelitian yaitu LRFM (Length, Recency, Frequency, Monetary) dan variabel untuk perilaku tambahan. Variabel yang tidak relevan akan dihilangkan.

3. Preprocessing Data



Gambar 2. Alur Data Cleaning

Pada tahap *Preprocessing* data dilakukan *data cleaning* untuk memastikan dataset transaksi pelanggan mempunyai kualitas yang baik dan bebas dari masalah yang dapat mengganggu analisis. Proses *data cleaning* pada Gambar 2.

4. Transformation

Setelah data bersih melanjutkan langkah transformasi data dengan Python di Google Colab yaitu membentuk variabel LRFM dan membentuk variabel perilaku transaksi yaitu rata-rata diskon per

order, rata-rata ongkir per order, rata-rata jumlah iter per order, jumlah produk unik. Variabel ini digunakan untuk memperkaya informasi yang belum tercakup LRFM.

Tabel 1. Tahap Pembentukan LRFM

Recency	Frequency	Monetary	Length
Selisih waktu sejak transaksi terakhir pelanggan	Menghitung jumlah transaksi pelanggan	Menghitung Total Pembayaran	Selisih antara waktu terakhir dengan waktu pertama kali pelanggan membeli

Tabel 2. Deskripsi Variabel Perilaku Transaksi

Rata-rata diskon per order	Rata-rata ongkir per order	Rata-rata jumlah iter per order	Jumlah Produk Unik
Rata-rata diskon yang diterima pelanggan per order	Rata-rata biaya pengiriman pelanggan per order	Rata-rata jumlah barang yang dibeli per order	Jumlah jenis produk yang dibeli pelanggan

Sebelum clustering, data kemudian dinormalisasi dengan *Min-Max* sehingga diperoleh keseimbangan nilai antara data asli dan data yang sudah diproses dalam normalisasi [10]. Rumus dari Normalisasi *Min-Max* sebagai berikut:

$$x_{new} = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Keterangan :

- x_{new} = Nilai Setelah Normalisasi
- x_i = Nilai Sebelum Normalisasi
- $\max(x)$ = Nilai Maksimum dalam data
- $\min(x)$ = Nilai Minimum dalam data

5. Data Mining

Setelah semua variabel memiliki nilai yang setara, pada penelitian ini dilakukan proses data mining dengan menggunakan tiga algoritma *clustering* yaitu Algoritma K-Means, K-Medoids, dan AHC. Tiga algoritma dipilih berdasarkan perbedaan karakteristiknya, seperti K-Means memiliki waktu komputasi yang efisien untuk data besar, K-Medoids lebih tahan terhadap outlier, serta AHC membentuk struktur hierarki data.

a. K-Means Clustering

Algoritma K-Means bekerja dengan meminimalkan variasi dalam *cluster* dan memaksimalkan variasi data pada *cluster* lain [11]. Berikut langkah-langkah algoritma K-Means [12]:

1. Menentukan jumlah cluster sebanyak misal k
2. Menentukan centroid awal dari *cluster* secara acak
3. Menghitung jarak antara titik data dengan pusat centroid dengan jarak Euclidean pada persamaan (2) [3], masukkan titik ke cluster yang jaraknya terdekat.

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

Keterangan :

- $d_{(x,y)}$ = Jarak Euclidean antara data x & y
 - x_i = nilai data pertama dimensi ke- i
 - y_i = nilai data kedua pada dimensi ke- i
 - n = jumlah dimensi data
4. Menghitung centroid baru dengan rata-rata data dalam *cluster*
 5. Mengulangi langkah c dan d hingga centroid tidak berubah secara signifikan
 6. Memperbarui nilai centroid untuk menentukan ada tidaknya perubahan nilai centroid, jika tidak ada perhitungan centroid dihentikan.

b. K-Medoids Clustering

Algoritma K-Means menggunakan centroid, sedangkan K-Medoids menggunakan medoid yaitu titik asli data, sehingga algoritma ini lebih tahan outlier daripada algoritma K-Means [13]. Berikut langkah-langkah algoritma K-Medoids [14]:

1. Menentukan jumlah *cluster* (k).
2. Memilih pusat *cluster* awal atau *medoids* dengan acak.
3. Menghitung jarak tiap titik data ke *medoids* dengan jarak Euclidean pada persamaan (2).
4. Kelompokkan titik data pada *cluster* terdekat
5. Menguji kandidat *medoids* baru pada tiap *cluster*
6. Menghitung S_{total} dari selisih antara total jarak kandidat *medoids* baru dan total jarak *medoids* lama
7. Jika $S_{total} < 0$, maka kandidat *medoids* baru sebagai *medoids* baru, jika $S_{total} > 0$, iterasi berhenti
8. Ulangi langkah d-g sampai tidak terjadi perubahan *medoids* atau nilai $S_{total} > 0$

c. Agglomerative Hierarchical Clustering

AHC pengelompokan data yang dilakukan secara *bottom-up* yaitu dengan mengelompokkan setiap data dimulai sebagai klaster tunggal dan secara bertahap digabung menjadi satu *cluster* sampai semua data terkelompokkan [15]. Ada beberapa langkah algoritma *Agglomerative Hierarchical Clustering*, antara lain [16]:

1. Menghitung matrik jarak antar data yang ada.
2. Menggabungkan 2 klaster yang memiliki jarak terdekat.
3. Melihat kedekatan antara klaster baru dan klaster yang masih ada dengan memperbarui matrik jarak pada persamaan (3) [17].

$$L(C_1, C_2) = \frac{n_{c1}n_{c2}}{n_{c1}+n_{c2}} d(c_{c1} - c_{c2})^2$$

Keterangan:

n_{c1}, n_{c2} = jumlah anggota pada *cluster* C_1 dan C_2

c_{c1}, c_{c2} = centroid dari C_1 dan C_2

$|c_{c1} - c_{c2}|^2$ = jarak Euclidean kuadrat antar centroid

4. Mengulangi langkah b dan c sampai 1 klaster yang tersisa.

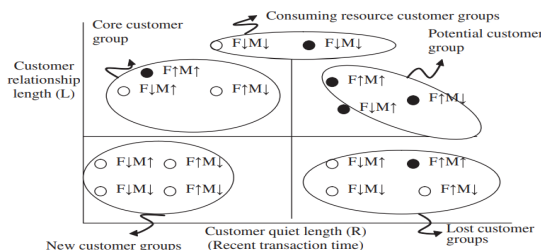
6. Interpretation/Evaluation

a. Evaluasi Kualitas Cluster

Penelitian ini dilakukan dengan dua metrik evaluasi diantaranya *Silhouette Coefficient* memiliki skor yaitu antara -1 hingga 1, jika semakin tinggi nilai maka menunjukkan klaster semakin baik [12]. Selain itu, evaluasi dengan *Davies-Bouldin Indeks* (DBI), Kebalikan dari *Silhouette Score* nilai DBI terendah adalah nilai terbaik artinya *cluster* tersebut berhasil padu di dalam dan terpisah di luar.

b. Interpretasi Variabel LRFM

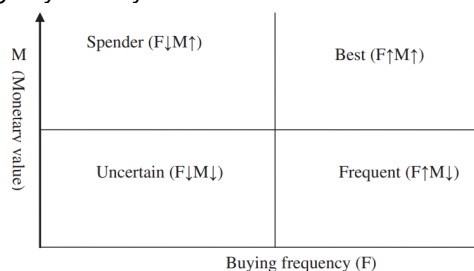
Dalam menjelaskan hasil *clustering* dan mendapatkan gambaran mengenai perilaku pelanggan dari model LRFM perlu integrasi *Customer Value Matrix* dan *Customer Loyalty Matrix* [18] [19].



Gambar 3. Customer Loyalty Matrix Chang & Tsay (2004) dalam Li et al. (2011)

Customer Loyalty Matrix adalah parameter untuk menentukan tingkat loyalitas suatu pelanggan [20].

Terdapat dua indikator yang berhubungan dengan loyalitas pelanggan diantaranya *Length* dan *Recency* [19]. Jadi semakin panjang hubungan pelanggan *L* tinggi dan semakin baru waktu transaksi *R* rendah, maka semakin tinggi loyalitasnya.



Gambar 4. Customer Value Matrix Marcus (1998) dalam Li et al. (2011)

Customer Value Matrix adalah parameter untuk menentukan nilai pelanggan bagi perusahaan [20]. *Customer Value Matrix* ditentukan oleh *Frequency* dan *Monetary* [19]. Berikut arti simbol $\uparrow\downarrow$ pada *customer value-loyalty matrix* [19]:

Tabel 3. Simbol Pada Customer Value-Loyalty Matrix

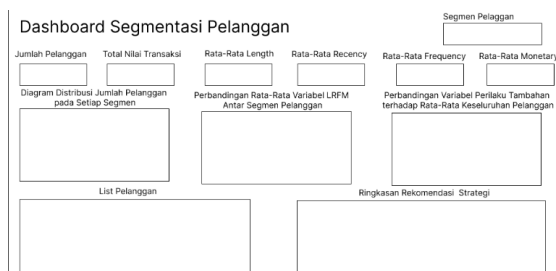
Simbol	Makna
$\uparrow$	Rata-rata atribut tiap <i>cluster</i> lebih besar daripada Rata-rata keseluruhan atribut
$\downarrow$	Rata-rata atribut tiap <i>cluster</i> lebih kecil daripada Rata-rata keseluruhan atribut

7. Knowledge Presentation

a. Rekomendasi Strategi

Pada tahap sebelumnya telah dilakukan analisis karakteristik tiap *cluster*, tahap selanjutnya perumusan rekomendasi strategi pemasaran berdasarkan analisis tersebut.

b. Visualisasi Dashboard



Gambar 5. Gambaran Dashboard CV Restu Tani Jaya

Langkah akhir setelah memperoleh hasil selanjutnya akan divisualisasikan dalam bentuk *dashboard* BI agar informasi mudah dipahami oleh pihak CV. Restu Tani Jaya. Dashboard dirancang berdasarkan dari kebutuhan bisnis utama yang diperoleh saat wawancara dengan pihak CV Restu Tani Jaya.

HASIL PENELITIAN

Tabel 4. Kebutuhan Informasi

Permasalahan Bisnis	Kebutuhan Informasi
Semua Pelanggan diperlakukan sama	Mebutuhkan segmentasi pelanggan berbasis data
Pelanggan penting hanya berdasarkan ingatan	Mebutuhkan Pemetaan pelanggan prioritas berdasar nilai dan aktivitas transaksi
Diskon belum dianalisis efektivitasnya	Mebutuhkan sensitivitas pelanggan terhadap diskon
Ongkos kirim sering menjadi penyebab pembatalan	Mebutuhkan analisis pengaruh ongkos kirim
Pelanggan jarang membeli banyak variasi	Analisis keragaman dan pola pembelian barang
Strategi yang dilakukan masih bersifat umum	Rekomendasi startegi pemasaran yang tersegmentasi

Sebagaimana pada Tabel 4 hasil wawancara dengan CV. Restu Tani Jaya terlihat bahwa perusahaan

kurang memahami karakteristik pelanggan. Oleh karena itu dilakukan segmentasi pelanggan untuk mendukung strategi pemasaran yang lebih tepat sasaran.

Dataset yang digunakan adalah data transaksi penjualan online dari platform Shopee. Data diberikan dalam bentuk file excel (.xlsx) dan jumlah data awal yaitu 2.393 baris dengan 49 atribut setelah proses seleksi menjadi 11 atribut yang disesuaikan dengan kebutuhan penelitian yaitu LRFM dan variabel untuk perilaku tambahan. Data mentah kemudian dilakukan preprocessing dan transformasi data untuk memastikan data mempunyai kualitas yang sesuai agar untuk proses selanjutnya berjalan dengan optimal.

Tabel 5. Hasil Transformasi Data ke Variabel LRFM dan Variabel Tambahan

Username (Pembeli)	Recency	Frequency	Monetary	Length	Rata_diskon	Rata_Jumlah_Item	Jumlah_Produk_Unik	Rata_Ongkir
.sarno	861	1	153510	0	0.0	1.0	1	1000
02102012rara	1110	3	635420	100	8800.0	3.3	1	12833
0.27.2012	1141	1	300000	0	0.0	4.0	1	36000
083wcld_hv	804	1	352000	0	0.0	4.0	1	92500

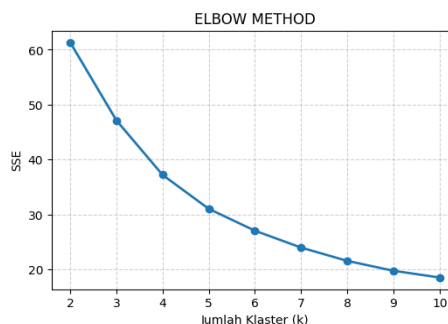
Variabel yang telah terbentuk akan dilakukan penyamaan nilai *Min-Max* dengan rentang 0-1. Pada Gambar 9 adalah cuplikan hasil dari normalisasi data.

Tabel 6. Hasil Normalisasi Data

Username (Pembeli)	Recency	Frequency	Monetary	Length	Rata_diskon	Rata_Jumlah_Item	Jumlah_Produk_Unik	Rata_Ongkir
.sarno	0.635625	0.000000	0.033570	0.000000	0.000000	0.000000	0.000000	0.000645
02102012rara	0.819660	0.058824	0.146297	0.076104	0.050286	0.048611	0.000000	0.015904
08.27.2012	0.842572	0.000000	0.067836	0.000000	0.000000	0.062500	0.000000	0.045777
083wcld_hv	0.593496	0.000000	0.080000	0.000000	0.000000	0.062500	0.000000	0.118633

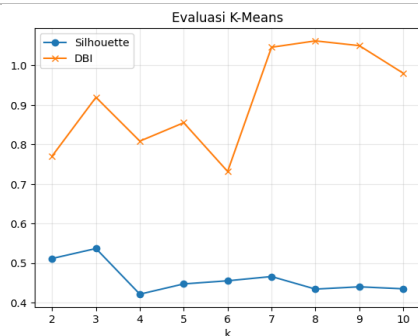
Selanjutnya dilakukan clustering dengan tiga algoritma yaitu:

1. K-Means Clustering



Gambar 6. Grafik Elbow Method

Pada grafik terjadi penurunan yang signifikan pada jumlah k=2 sampai k=4. Setelah k=4 penurunan SSE cenderung melandai yang menandakan penurunan perbedaan sudah sangat kecil dengan menambah cluster akan memiliki hasil yang mirip antar cluster. Oleh karena itu, menurut elbow jumlah cluster optimal yang dipilih ada k=4, penentuan jumlah klaster dievaluasi dengan menggunakan *Silhouette Score* dan DBI.



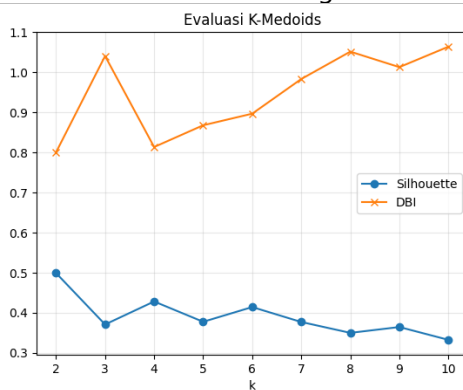
Gambar 7. Evaluasi Silhouette Score dan DBI K-Means

Tabel 7. Nilai Silhouette Score dan DBI K-Means

K-Means		
Cluster (k)	Silhouette Score	Davies-Bouldin Index
2	0.51161	0.77046
3	0.53689	0.91918
4	0.42147	0.80846
5	0.44744	0.85500
6	0.45533	0.73176
7	0.46607	1.04581
8	0.43430	1.06193
9	0.44010	1.04988
10	0.43505	0.98018

Dari grafik keseimbangan kualitas terlihat pada 2 klaster untuk K-Means Clustering karena masih menghasilkan skor Silhouette yang masih tinggi yaitu 0.51161 dan DBI yang cukup rendah yaitu 0.77046.

2. K-Medoids Clustering



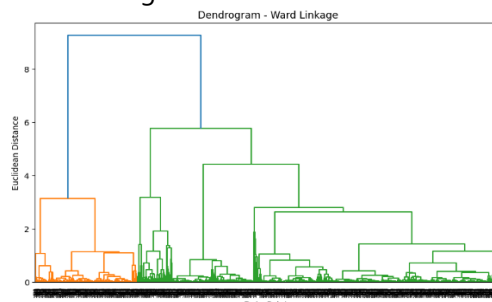
Gambar 8. Evaluasi Silhouette Score dan DBI K-Medoids

Tabel 8. Nilai Silhouette Score dan DBI K-Medoids K-Means

K-Medoids		
Cluster (k)	Silhouette Score	Davies-Bouldin Index
2	0.49886	0.80006
3	0.37060	1.04023
4	0.42794	0.81324
5	0.37728	0.86749
6	0.41402	0.89642
7	0.37677	0.98288
8	0.34962	1.05119
9	0.36420	1.01253
10	0.33220	1.06353

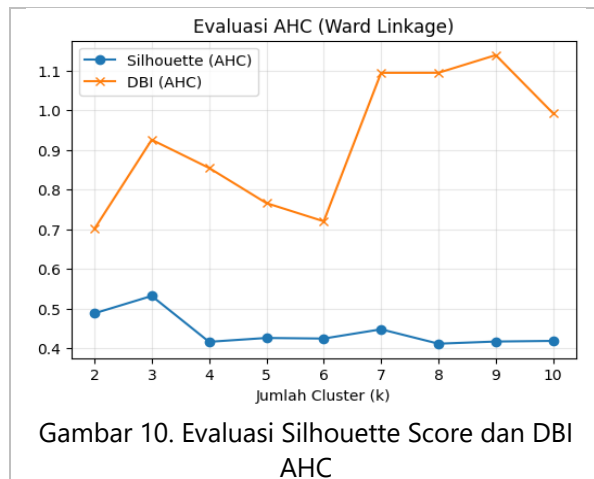
Berdasarkan gambar 10 dan table 8 nilai *Silhouette Score* tertinggi pada k=2 yaitu dengan nilai 0.49886, sedangkan nilai DBI terendah juga pada k=2 yaitu 0.80006. Sehingga, pada Silhouette dan DBI sama-sama menunjukkan hasil yang konsisten bahwa jumlah cluster pada K-Medoids adalah k=2.

3. Agglomerative Hierarchical Clustering



Gambar 9. Visualisasi Dendrogram

Dendrogram pada gambar diatas dilakukan penggabungan dengan Ward Linkage. Pada gambar tersebut ada lonjakan jarak penggabungan yang cukup signifikan pada jarak 8-9 artinya adanya dua cluster yang gabung dengan perbedaan yang besar. Sehingga, pemotongan dendrogram dilakukan sebelum lonjakan yaitu 2 cluster menurut dendrogram.



Tabel 9. Nilai Silhouette Score dan DBI AHC

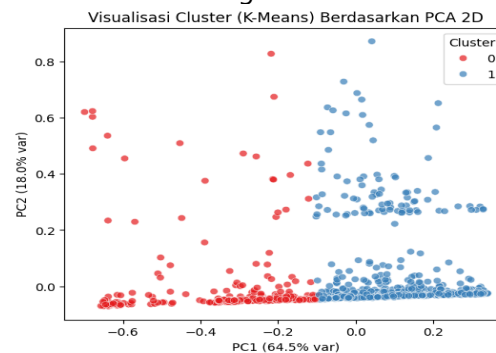
AHC		
Cluster (k)	Silhouette Score	Davies-Bouldin Index
2	0.48804	0.70068
3	0.53193	0.92579
4	0.41632	0.85478
5	0.42602	0.76607
6	0.42450	0.72016
7	0.44783	1.09458
8	0.41161	1.09468
9	0.41704	1.13921
10	0.41863	0.99241

Gambar 10 dan table 9 menunjukkan pada k=3 memiliki Silhouette Score yang tinggi dengan DBI yang tinggi juga, hal ini menandakan bahwa k=3 memiliki cluster yang kurang kompak dan rapi. Oleh karena itu, jumlah kluster yang dipilih adalah k=2 karena antara pemisahan kluster pada Silhouette dan kekompakan kluster pada DBI memiliki keseimbangan terbaik.

Tabel 10. Perbandingan Hasil Evaluasi Algoritma Clustering

Metode	Silhouette (k=2)	DBI (k=2)
K-Means	0.511610	0.770462
K-Medoids	0.498856	0.800059
AHC	0.488040	0.700679

Perbandingan hasil evaluasi diatas, algoritma K-Means memiliki nilai Silhouette Score tertinggi artinya pemisahan cluster lebih baik dari algoritma lain, tetapi bukan pemilik nilai DBI terendah. Pada penelitian ini, K-Means dipilih karena memiliki keseimbangan hasil dari kedua metrik evaluasi.



Gambar 11. Visualisasi Pemisahan Cluster K-Means Berdasarkan PCA 2D

Pada gambar 11 menunjukkan kedua cluster terpisah jelas, dimana cluster 0 cenderung berada di sisi kiri dan cluster 1 berada di sisi kanan artinya model berhasil mengelompokkan pelanggan berdasarkan kemiripan pola perilaku. Dalam memberikan makna bisnis suatu cluster yang terbentuk, dilakukan proses pelabelan dengan memanfaatkan Customer Loyalty Matrix dan Customer Value Matrix.

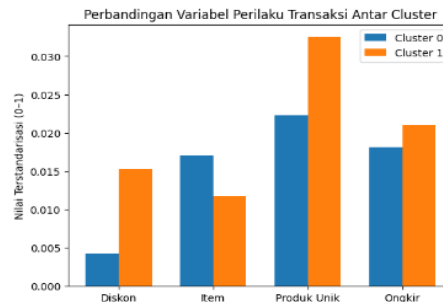
Tabel 11. Hasil Pelabelan Integrasi Customer Loyalty-Value Matrix

Cluster	Length	Recency	Frequency	Monetary	Simbol	Label
0	40.919786	481.467914	1.189840	186446.617647	L↑R↓F↑M↑	Core-Spender
1	0.14.318227	1047.164532	1.299507	.168803.225616	L↓R↑F↑M↓	Lost-Frequent
Rata-Rata Keseluruhan	21.48	894.85	1.27	173553.86		

Berdasarkan Tabel 6, menunjukkan bahwa cluster 0 ialah Core-Spender. Jadi cluster Core-Spender adalah pelanggan yang mempunyai hubungan jangka panjang yang baik dengan perusahaan dan baru melakukan transaksi dalam waktu dekat ini. Selain itu, mempunyai nilai transaksi yang besar, namun tidak terlalu sering dalam melakukan pembelian. Pada cluster ini merupakan cluster pelanggan yang mempunyai

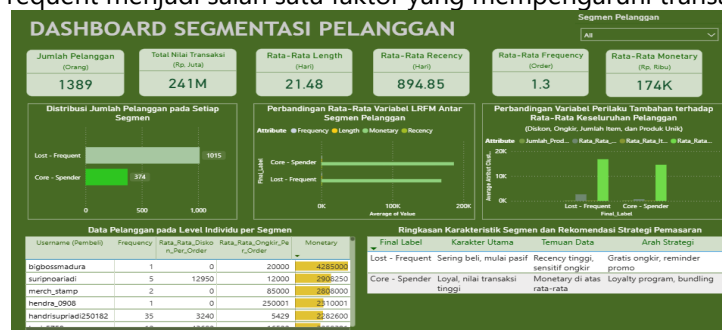
nilai ekonomi yang tinggi dan perlu dipertahankan.

Sedangkan cluster Lost-Frequent adalah kelompok pelanggan yang relatif sering melakukan pembelian tetapi memiliki Recency yang tinggi artinya pelanggan pernah aktif bertransaksi namun sekarang sudah lama tidak bertransaksi. Cluster ini memiliki nilai transaksi yang rendah daripada cluster lain. Hal ini menunjukkan cluster ini memiliki karakteristik pelanggan yang kurang loyal dan perlu didorong kembali agar melakukan transaksi.



Gambar 12. Perbandingan Variabel Perilaku Transaksi Antar Cluster

Dari gambar 12 menunjukkan bahwa: Rata-rata diskon per order cluster 1 Lost-Frequent memiliki nilai lebih besar daripada cluster 0 Core-Spender artinya cluster 1 Lost-Frequent cenderung melakukan transaksi saat ada diskon/promosi; Rata-rata jumlah item per order cluster 0 Core-Spender memiliki nilai lebih tinggi cluster 1 Lost-Frequent, sehingga Core-Spender lebih cenderung membeli lebih banyak produk dalam sekali transaksi jadi cocok untuk strategi bundling; Jumlah produk unik cluster 1 Lost-Frequent memiliki nilai yang lebih tinggi daripada cluster 0 Core-Spender artinya cluster 0 cenderung membeli produk yang berulang dan cluster 1 lebih cenderung membeli berbagai jenis produk; Rata-rata ongkir per order cluster 1 Lost-Frequent memiliki nilai rata-rata yang lebih tinggi daripada cluster 0 Core-Spender, sehingga cluster Lost-Frequent cenderung membayar ongkir yang relatif lebih besar. Tingginya rata-rata ongkir pada cluster Lost-Frequent menjadi salah satu faktor yang mempengaruhi transaksi *cluster* ini.



Gambar 13. Dashboard Segmentasi Pelanggan Berdasarkan LRFM dan Variabel Perilaku Tambahan

Hasil segmentasi pelanggan disajikan dalam bentuk dashboard, sehingga memudahkan pihak CV. Restu Tani Jaya dalam memahami perbedaan karakteristik antar *cluster* seperti tingkat frekuensinya, nilai pembelian, lama hubungan pelanggan, pola pembelian, serta sensitivitas pada diskon dan ongkir dengan cepat. Penyajian insight serta rekomendasi strategi dapat membantu CV. Restu Tani Jaya dalam proses pengambilan keputusan berdasarkan data.

PEMBAHASAN

Kedua cluster yang terbentuk memiliki karakteristik yang berbeda dan perlu strategi yang sesuai dengan karakteristiknya. Berikut penyusunan strategi masing-masing cluster berdasarkan karakteristik perilaku pelanggannya:

1. Strategi Pemasaran pada kelompok Core-Spender

- Melakukan strategi retensi. Strategi bisa berupa pelayanan prioritas yang konsisten dan lebih responsif, memberikan ucapan terima kasih, serta pemberian reward voucher eksklusif untuk menjaga hubungan jangka panjang dengan kelompok pelanggan yang bernilai tinggi.

- b. Melakukan penawaran produk bundling. Kelompok pelanggan ini memiliki rata-rata jumlah item per order yang tinggi, sehingga strategi ini untuk meningkatkan nilai *Monetary* dengan merekomendasikan produk pelengkap atau produk yang nilainya lebih tinggi.
 - c. Melakukan pendekatan secara personalisasi. Contohnya seperti memberikan rekomendasi produk berdasarkan pola riwayat pembelian untuk meningkatkan peluang pembelian ulang.
 - d. Mengoptimalkan layanan. Kelompok pelanggan ini memberikan pendapatan yang besar pada CV. Restu Tani Jaya, sehingga perlu adanya optimalisasi layanan seperti kecepatan dalam pengiriman dan menjaga kualitas produk.
2. Strategi Pemasaran pada kelompok Lost-Frequent
 - a. Memberikan strategi reaktivasi. Tujuannya untuk mendorong pembelian ulang ke toko seperti komunikasi personal melalui WhatsApp Blast dan fitur Shopee Chat dengan memberikan informasi promo musiman atau informasi produk baru.
 - b. Optimalisasi program promosi berbasis harga. Pada kelompok pelanggan ini lebih tertarik dengan potongan harga dan program promosi, sehingga dapat memberikan voucher dengan minimal pembelian tertentu atau berlaku terbatas melalui fitur Voucher Toko atau *Flash Sale* untuk menarik minat pelanggan.
 - c. Pengelolaan biaya pengiriman. Kelompok pelanggan ini memiliki rata-rata ongkos kirim yang tinggi yang menunjukkan biaya ongkir menjadi hambatan, sehingga strategi program gratis ongkir atau subsidi ongkir dengan minimum pembelian dapat mendorong kembali pelanggan untuk bertransaksi.

KESIMPULAN

Penelitian berhasil melakukan segmentasi pelanggan pada CV. Restu Tani Jaya dengan menggunakan pendekatan LRFM dan diperkaya dengan variabel perilaku transaksi tambahan yaitu rata-rata diskon per order, rata-rata jumlah item per order, rata-rata ongkir per order, dan jumlah produk unik. Berdasarkan evaluasi dengan Silhouette Score dan Davies-Bouldin Index, algoritma K-Means mempunyai kualitas kluster yang paling stabil, jadi proses interpretasi dan pelabelan cluster menggunakan hasil pengelompokan algoritma K-Means. Terbentuk dua segmen yaitu Core-Spender dan Lost-Frequent. Karakteristik dari Core-Spender memiliki nilai transaksi yang lebih tinggi, loyal, ketergantungan pada diskon dalam pembelian rendah sehingga memerlukan strategi retensi, melakukan penawaran produk bundling, melakukan pendekatan secara personalisasi, dan mengoptimalkan layanan, sedangkan segmen Lost-Frequent memiliki karakteristik relatif sering bertransaksi pada masa lalu, sebagian besar saat ini tidak aktif, dan lebih responsif terhadap diskon. memerlukan strategi reaktivasi, melakukan optimalisasi program promosi berbasis harga, dan pengelolaan biaya pengiriman. Hasil analisis disajikan dalam bentuk visualisasi dashboard tujuannya untuk mendukung proses pengambilan keputusan CV. Restu Tani Jaya yaitu dalam memahami profil pelanggan, perbedaan antar segmen, dan rekomendasi strategi yang dapat dilakukan.

REFERENCES

- [1] M. Amirulhaq Iskandar, U. Latifa, U. Singaperbangsa Karawang, and J. H. Ronggo Waluyo, "WEBSITE PREDIKSI CUSTOMER CHURN UNTUK MEMPERTAHANKAN PELANGGAN PADA PERUSAHAAN TELEKOMUNIKASI," 2023.
- [2] F. Marisa, S. S. S. Ahmad, Z. I. M. Yusof, Fachrudin, and T. M. A. Aziz, "Segmentation model of customer lifetime value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM model," *International Journal of Integrated Engineering*, vol. 11, no. 3, pp. 169–180, 2019, doi: 10.30880/ijie.2019.11.03.018.
- [3] I. Gede, I. Sudipa, and M. Darmawiguna, "BUKU AJAR DATA MINING," 2024. [Online]. Available: <https://www.researchgate.net/publication/377415198>
- [4] A. Ayu, D. Sulistyawati, and M. Sadikin, "SISTEMASI: Jurnal Sistem Informasi Penerapan Algoritma K-Medoids untuk Menentukan Segmentasi Pelanggan." [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>

- [5] J. Homepage, H. Syukron, M. Fauzi Fayyad, F. Junita Fauzan, Y. Ikhsani, and U. Rizkya Gurning, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Comparison K-Means K-Medoids and Fuzzy C-Means for Clustering Customer Data with LRFM Model "Perbandingan K-Means K-Medoids dan Fuzzy C-Means untuk Pengelompokan Data Pelanggan dengan Model LRFM," vol. 2, pp. 76–83, 2022.
- [6] G. Ramadhan, P. Analisis Algoritma K-means, and Y. Astuti, "Perbandingan Kinerja Algoritma K-means dan Agglomerative Clustering Untuk Segmentasi Penjualan Online Pada Customer Retail," vol. 9, no. 1, 2024.
- [7] T. L. Nikmah, N. H. S. Harahap, G. C. Utami, and M. M. Razzaq, "Customer Segmentation Based on Loyalty Level Using K-Means and LRFM Feature Selection in Retail Online Store," *Jurnal ELTIKOM*, vol. 7, no. 1, pp. 21–28, Jun. 2023, doi: 10.31961/eltikom.v7i1.648.
- [8] J. Zhou, J. Wei, and B. Xu, "Customer segmentation by web content mining," *Journal of Retailing and Consumer Services*, vol. 61, Jul. 2021, doi: 10.1016/j.jretconser.2021.102588.
- [9] N. P. V. Viandari, I. M. A. D. Suarjaya, and I. N. Piarsa, "Pemetaan Pelanggan dengan LRFM dan Two Stage Clustering untuk Memenuhi Strategi Pengelolaan," *Jurnal RESTI*, vol. 6, no. 1, pp. 130–139, Feb. 2022, doi: 10.29207/resti.v6i1.3778.
- [10] P. Henderi, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *IJIS: International Journal of Informatics and Information Systems*, vol. 4, pp. 13–20, Mar. 2021, doi: 10.47738/ijis.v4i1.73.
- [11] L. Rahmawati, J. Informatika, S. W. Sihwi, and E. Suryani, "ANALISA CLUSTERING MENGGUNAKAN METODE K-MEANS DAN HIERARCHICAL CLUSTERING (STUDI KASUS: DOKUMEN SKRIPSI JURUSAN KIMIA, FMIPA, UNIVERSITAS SEBELAS MARET)," 2024.
- [12] J. Kiwasi Wororomi *et al.*, "DATA MINING (MEMAHAMI POLA DI BALIK ANGKA) PENERBIT CV.EUREKA MEDIA AKSARA," 2024.
- [13] C. K. Reddy and B. Vinzamuri, "data_clustering," 2014.
- [14] N. Mirantika, T. S. Syamfithriani, and R. Trisudarmo, "Implementasi Algoritma K-Medoids Clustering Untuk Menentukan Segmentasi Pelanggan," *JURNAL NUANSA INFORMATIKA*, vol. 17, pp. 2614–5405, Jan. 2023, doi: 10.25134/nuansa.
- [15] R. P. Justitia, N. Hidayat, and E. Santoso, "Implementasi Metode Agglomerative Hierarchical Clustering Pada Segmentasi Pelanggan Barbershop (Studi Kasus: RichDjoe Barbershop Malang)," 2021. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [16] Z. Muhammad, "Hierarchical Clustering 4," 2024.
- [17] A. Gere, "Recommendations for validating hierarchical clustering in consumer sensory projects," *Curr. Res. Food Sci.*, vol. 6, p. 100522, 2023, doi: <https://doi.org/10.1016/j.crfs.2023.100522>.
- [18] H. H. Chang and S.-F. Tsay, "INTEGRATING OF SOM AND K-MEAN IN DATA MINING CLUSTERING: AN EMPIRICAL STUDY OF CRM AND PROFITABILITY EVALUATION," 2004. [Online]. Available: <https://api.semanticscholar.org/CorpusID:167554134>
- [19] 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 business," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7186–7191, Jun. 2011, doi: 10.1016/j.eswa.2010.12.041.
- [20] Y. Afilda, "Pemetaan Pelanggan Menggunakan Model Length, Recency ,Frequency dan Monetary Untuk Mendukung Strategi Pengelolaan Pelanggan," 2021.