

Perancangan Dashboard Analitik Untuk Pemantauan Segmentasi Pelanggan Dengan K-Means Clustering

Daniela Aedo ¹⁾ Dedi Trisnawarman ²⁾

¹⁾²⁾ Sistem Informasi, FTI, Universitas Tarumanagara Jl. Letjen S Parman no 1, Jakarta 11440 Indonesia

email: daniela.825210025@stu.untar.ac.id ¹⁾, dedit@fti.untar.ac.id ²⁾

ABSTRAK

Pelanggan merupakan aspek penting yang menentukan keberlanjutan bisnis perusahaan. Dengan memahami perilaku dan nilai pelanggan, perusahaan dapat mengembangkan strategi pemasaran dan mempertahankan retensi pelanggan, serta menciptakan keunggulan dalam pasar. Penelitian ini bertujuan untuk melakukan analisis segmentasi pelanggan dan menyajikan hasil tersebut ke dalam bentuk dashboard. Analisis segmentasi pelanggan dilakukan menggunakan K-means clustering dengan fitur LRFM model dan nilai CLV. Berdasarkan hasil elbow method dan silhouette score, jumlah kluster optimal yang dipilih adalah k=6, sehingga terbentuk enam segmen pelanggan yang memiliki karakteristik unik. Hasil clustering kemudian digunakan untuk merancang dashboard yang berisi ringkasan karakteristik setiap segmen pelanggan, serta menyajikan informasi mengenai pola perilaku dan nilai CLV yang dihasilkan oleh pelanggan. Penelitian ini berhasil mengintegrasikan analisis clustering dalam segmentasi pelanggan serta merancang dashboard interaktif yang menyajikan informasi relevan sesuai dengan kebutuhan bisnis perusahaan XYZ, yaitu membantu tim marketing dalam memahami perilaku dan potensi nilai pelanggannya secara lebih dalam.

Key words

Dashboard, Customer Lifetime Value, Segmentasi Pelanggan, LRFM Model, K-Means Clustering

1. Pendahuluan

Dalam dunia bisnis, pelanggan merupakan aspek penting yang menentukan keberlanjutan bisnis perusahaan. Tidak hanya sebagai sumber pendapatan, pelanggan juga berperan sebagai faktor utama yang memberikan kontribusi serta mencerminkan nilai yang dibawa untuk perkembangan perusahaan, baik melalui aktivitas, interaksi, *feedback*, dan loyalitas [1]. Setiap aktivitas yang dilakukan oleh pelanggan memberikan wawasan mendalam tentang perilaku dan preferensi pembelian. Pemahaman wawasan tersebut memberikan perusahaan kemampuan untuk mengoptimalkan strategi pemasaran, meningkatkan loyalitas pelanggan, serta mengembangkan dan menawarkan produk atau layanan sesuai dengan kebutuhan pelanggan. Data seperti aktivitas pelanggan atau transaksi mengandung informasi penting

yang dapat menjadi aset paling berharga bagi perusahaan dalam memahami wawasan tersebut untuk mendukung kesuksesan operasional bisnis dan proses pengambilan keputusan [2]. Informasi dari data aktivitas pelanggan juga membantu perusahaan dalam mengidentifikasi adanya pola atau tren yang dapat muncul dari interaksi pelanggan dengan bisnis.

Sekitar 80% dari keuntungan bisnis perusahaan berasal dari 20% pelanggan loyal atau pelanggan yang selalu melakukan transaksi dengan perusahaan [3]. Tentu, perusahaan ingin mengetahui siapa saja 20% pelanggan tersebut yang telah berkontribusi dan dianggap berharga ini. Salah satu pendekatan untuk mengetahui pelanggan-pelanggan tersebut adalah *Customer Lifetime Value* (CLV). CLV merupakan suatu metrik yang mengacu pada total nilai keuntungan yang dihasilkan oleh pelanggan sepanjang masa hubungannya dengan perusahaan [4]. CLV menjadi penting dalam bisnis karena kemampuan analisisnya yang membantu perusahaan dalam mengambil keputusan strategis berdasarkan nilai potensial pelanggan. Penggunaan CLV dapat menargetkan pelanggan yang menguntungkan saja atau pelanggan yang berpotensi untuk belanja lagi dan memberikan lebih banyak keuntungan untuk perusahaan di masa depannya [5]. Dengan menjaga dan mengelola hubungan yang baik dengan pelanggan tersebut membantu perusahaan meningkatkan keuntungan dan keunggulan dalam persaingan pasar. Perusahaan dapat menggunakan analisis ini untuk mengidentifikasi pelanggan yang paling berharga, menentukan alokasi sumber daya yang efektif, mengoptimalkan strategi pemasaran, menargetkan pelanggan yang tepat dalam pemasaran, serta mengembangkan strategi akuisisi dan retensi pelanggan. Analisis segmen pelanggan berdasarkan *lifetime value* yang efektif dapat membawa lebih banyak keuntungan bagi perusahaan [6].

Berdasarkan hasil survei dan wawancara pada ritel XYZ, saat ini perusahaan, khususnya tim marketing, mengalami kesulitan dalam memahami kualitas dan perilaku pelanggan, serta mengidentifikasi pelanggan yang bernilai tinggi. Selain itu, tim juga kesulitan mengetahui pelanggan mana yang perlu dipertahankan atau menjadi target dalam suatu kampanye pemasaran, sehingga strategi pemasaran seringkali bersifat umum dan kurang tepat sasaran. Hal ini dapat menyebabkan alokasi sumber daya menjadi tidak efisien, retensi pelanggan menjadi kurang optimal, dan hilangnya pelanggan yang

sebenarnya memiliki potensi besar terhadap bisnis di masa depan. Oleh karena itu, dibutuhkan solusi yang mampu membantu tim marketing dalam menganalisis data pelanggan secara lebih efektif dan efisien.

Penelitian-penelitian sebelumnya telah mengusulkan berbagai pendekatan untuk meningkatkan analisis CLV. Beberapa studi menggabungkan berbagai metode seperti *clustering*, klasifikasi, prediksi, *market basket analysis*, serta berbagai algoritma *machine learning* untuk meningkatkan akurasi dan relevansi analisis CLV [7] [8]. Sebagian besar penelitian tersebut masih berfokus pada analisis data tanpa menyediakan tools visual yang dapat langsung digunakan oleh tim marketing. Namun, beberapa di antaranya mengusulkan ide tentang visualisasi berbasis segmentasi pelanggan untuk mendukung performa penjualan atau rancangan dashboard khusus untuk keperluan marketing dalam memantau dan menentukan target segmen pelanggan berdasarkan pola perilaku dan preferensi produk [9] [10]. Rancangan dashboard dapat membantu tim marketing mengidentifikasi pelanggan potensial dan menyusun strategi untuk mendorong peningkatan transaksi. Dengan berbagai fitur penyaringan dan visualisasi data pelanggan berdasarkan berbagai dimensi, tim marketing dapat dengan mudah menentukan target pelanggan dan penawaran yang lebih terpersonalisasi.

Dengan demikian, penelitian ini terinspirasi untuk mengintegrasikan metode analitik untuk melakukan segmentasi pelanggan, yaitu *clustering* menggunakan K-Means, kemudian menyajikan hasil ke dalam bentuk dashboard yang interaktif dan mudah digunakan. Jurnal dengan judul “Perancangan Dashboard Analitik Untuk Pemantauan Segmentasi Pelanggan Dengan K-Means Clustering” ini disusun dengan tujuan untuk merancang dashboard yang dapat membantu tim marketing ritel XYZ dalam memahami perilaku dan nilai pelanggan terhadap perusahaan. Penulis berharap dengan adanya segmentasi pelanggan dan visualisasi data ini, perusahaan dapat lebih mudah memahami karakteristik pelanggan dan mengambil keputusan yang strategis untuk perkembangan bisnis kedepannya.

2. Metodologi

Langkah-langkah yang penulis lakukan dalam penelitian ini dapat dilihat pada **Gambar 1**, dimana terdapat enam langkah penting dengan penjelasan setiap langkah sebagai berikut:

2.1. Pengumpulan Data

Penelitian ini menggunakan data mart ritel XYZ, terutama data transaksi dan data pelanggan. Data pelanggan berisi informasi detail pelanggan, seperti ID pelanggan, nama pelanggan, email, nomor telepon, tanggal lahir, jenis kelamin, alamat, kota, tempat registrasi dan tanggal registrasi. Data transaksi berisi pencatatan pembelian oleh pelanggan dalam periode 30 Juni 2023 hingga 30 Juni 2024. Dalam kurun waktu 12 bulan tersebut, ritel XYZ tercatat memiliki sebanyak

987496 transaksi dan 131827 pelanggan, serta berisi beberapa variabel seperti *period*, *receive_no*, *cust_id*, *product_id*, *quantity*, *amount*, dan *store_id*. Data pelanggan digunakan pada tahap visualisasi dashboard untuk menampilkan informasi pelanggan, sedangkan data transaksi digunakan dalam analisis *clustering* untuk memahami pola perilaku pembelian pelanggan.



Gambar 1. Metode Penelitian

2.2. Data Cleaning

Sebelum melakukan analisis *clustering*, data transaksi perlu dibersihkan untuk memastikan kualitas data yang digunakan dalam analisis. Pada tahap ini, data transaksi yang tidak relevan, seperti transaksi yang dibatalkan, dihapus, format data seperti tanggal dan nilai numerik diperiksa dan disesuaikan, serta *missing values* atau data yang hilang ditangani.

2.3. Ekstraksi Fitur LRFM dan CLV

Langkah ini bertujuan untuk mengekstraksi data transaksi guna membentuk fitur model *length*, *recency*, *frequency*, *monetary* (LRFM) dan nilai CLV. Model LRFM terdiri dari 4 variabel dimana; *length* merujuk pada jarak antara transaksi pertama dan transaksi terakhir pelanggan. *Recency* merujuk pada jarak transaksi terakhir pelanggan terhadap tanggal referensi. *Frequency* merujuk pada jumlah transaksi yang dilakukan oleh pelanggan dalam periode waktu tertentu. Sedangkan *Monetary* merujuk pada jumlah uang yang akan dikeluarkan oleh pelanggan ketika bertransaksi dalam periode waktu

tertentu [11]. Selanjutnya, nilai CLV dapat dihitung dengan menggunakan komponen rata-rata nilai transaksi, frekuensi pembelian, dan lama hubungan pelanggan dengan bisnis. Dengan menggabungkan ketiga komponen tersebut, CLV memberikan gambaran tentang seberapa berharga seorang pelanggan bagi bisnis dalam jangka panjang [12]. Hasil perhitungan LRFM dan CLV kemudian digabungkan ke dalam satu dataset untuk digunakan sebagai input dalam proses *clustering*. Dataset ini akan menjadi dasar untuk mengelompokkan pelanggan berdasarkan perilaku dan nilainya.

2.4. Penerapan K-Means Clustering

Fitur LRFM dan CLV tersebut kemudian digunakan dalam proses penerapan algoritma K-Means menggunakan bahasa pemrograman python. K-Means merupakan algoritma *unsupervised learning* yang mengelompokkan data berdasarkan kesamaan karakteristik nilai data [13]. Dengan menggunakan metode iteratif, K-Means berusaha untuk meminimalkan jarak antara setiap titik data dengan titik pusat kluster yang mewakili kluster tersebut. Algoritma K-Means bekerja berdasarkan lima tahapan sebagai berikut [14]:

1. Menentukan jumlah kelompok atau kluster,
2. Menetapkan centroid awal secara acak berdasarkan jumlah kelompok,
3. Menghitung jarak antara data dengan centroid pada setiap kelompok menggunakan rumus *Euclidean Distance* (1),

$$d_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

4. Menghitung nilai rata-rata setiap nilai kelompok untuk memperbarui centroid, dan
5. Mengulangi Langkah 3 hingga nilai centroid tidak berubah dan objek kluster tidak berpindah ke kluster lain.

$$X = \frac{X - \mu}{\sigma} \quad (2)$$

Namun, sebelum menerapkan K-Means, data harus dinormalisasikan terlebih dahulu untuk memastikan semua fitur memiliki skala yang seragam. Normalisasi dilakukan menggunakan *Standard Scaler*. Metode ini melakukan penskalaan data dengan menyesuaikan nilai sehingga mean dan standar deviasi masing-masing memiliki nilai 0 dan 1, menggunakan rumus (2) [15]. Selanjutnya yaitu menentukan jumlah kluster yang optimal menggunakan *Elbow Method*. Metode ini didasarkan pada jarak kuadrat antara pusat kluster dan titik sampel setiap kluster untuk memberikan sejumlah nilai k [16]. *Elbow Method* bekerja dengan menghitung *Within-Cluster Sum-of-Squares* (WCSS) sebagai indikator untuk mengukur seberapa dekat titik data dalam kluster berada terhadap centroid, yang dinyatakan dengan rumus (3). Semakin kecil nilai WCSS, semakin baik

kluster tersebut, karena hal ini menunjukkan bahwa titik-titik data lebih terkonsentrasi di sekitar centroid [17].

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (3)$$

Selain itu, kualitas jumlah kluster juga divalidasi menggunakan *Silhouette Score*. Metode evaluasi ini menghitung setiap titik data untuk melihat seberapa baik titik-titik dalam suatu kluster saling dekat dan terpisah dari titik-titik di kluster lain, sesuai dengan rumus (4). Nilai *Silhouette Score* berkisaran dari -1 hingga 1, dimana nilai yang mendekati +1 menunjukkan data terpisah dengan jelas di setiap kluster. Sebaliknya, nilai yang mendekati -1 menunjukkan data ditempatkan di kluster yang tidak sesuai [18]. Setelah menentukan jumlah kluster yang optimal, algoritma K-Means diterapkan untuk mengelompokkan pelanggan ke dalam kluster yang sesuai. Hasil *clustering* ini dianalisis lebih lanjut untuk memahami karakteristik setiap kluster.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (4)$$

2.5. Pembuatan Dashboard

Tabel 1. Tabel KPI

KPI	Deskripsi	Target
Total member	Jumlah pelanggan yang aktif belanja	≥ 130 ribu
Total sales	Total penjualan yang dihasilkan dalam satu tahun	≥ 195 miliar
Recency	Jarak waktu sejak pembelian terakhir	≤ 120
Frequency	Jumlah transaksi yang dilakukan oleh pelanggan	≥ 3
Monetary	Rata-rata jumlah uang yang dikeluarkan setiap transaksi	≥ 500 ribu
CLV	Jumlah nilai CLV yang dihasilkan oleh seluruh pelanggan	≥ 48,750 miliar

Hasil *clustering* akan divisualisasikan dalam bentuk dashboard menggunakan platform Google Data Studio. Sumber data yang digunakan adalah hasil olah data *clustering* dan informasi pelanggan yang diambil dari tabel dimensi pelanggan. Perancangan dashboard diperlukan untuk menyajikan hasil analisis data dalam bentuk grafik dan tabel, sehingga pengguna dapat lebih mudah membaca dan memahami hasil segmentasi pelanggan. Dashboard ini dirancang tidak hanya untuk menampilkan hasil segmentasi, tetapi juga untuk menyajikan *Key Performance Indicators* (KPI) yang

relevan. KPI yang ditampilkan dalam dashboard dapat dilihat pada **Tabel 1**. Selain itu, dashboard ini juga memberikan *insight* yang berguna untuk memahami karakteristik nilai, kualitas, dan perilaku pelanggan di setiap segmen.

3. Hasil dan Pembahasan

3.1. Ekstraksi Fitur LRFM dan CLV

Tabel 2. LRFM dan CLV Setiap Pelanggan

cust_id	L	R	F	M	CLV
C00000003	195	149	5	886,927	2,092,337
C00000018	350	1	17	281,426	1,571,059
...
C000046854	159	144	5	195,520	152,290
C000236823	0	2	1	446,400	389,006

Pada tahap ini, perhitungan LRFM dan CLV dilakukan untuk setiap pelanggan. *Length*, *recency*, dan *frequency* dihitung berdasarkan variabel *period*, sedangkan *monetary* dihitung berdasarkan variabel *amount*. Untuk CLV menggunakan variabel *amount* dan *period*. Hasil penggabungan fitur LRFM dan nilai CLV setiap pelanggan dapat dilihat pada **Tabel 2**.

3.2. Penerapan K-Means Clustering

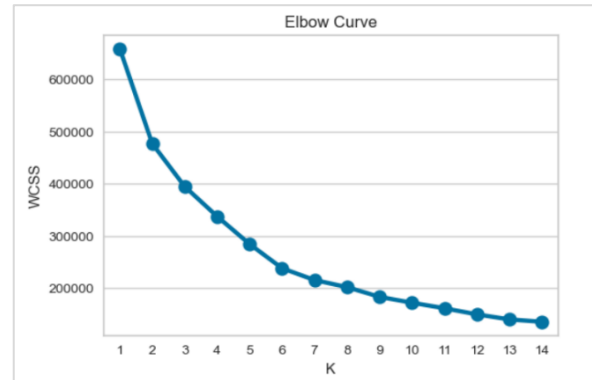
Tabel 3. Hasil Normalisasi Data

cust_id	L	R	F	M	CLV
C00000003	1.01	0.35	0.53	0.41	1.57
C00000018	2.49	-1.19	3.79	-0.38	1.17
...
C000046854	0.67	0.29	0.53	-0.49	-0.26
C000236823	-0.84	-1.18	-0.55	-0.16	-0.02

Setelah proses ekstraksi, nilai *length* dan *recency* masih dalam satuan hari, *frequency* dalam satuan jumlah transaksi, sedangkan *monetary* dan CLV menggunakan satuan mata uang. Perbedaan skala antar fitur ini dapat mempengaruhi analisis lebih lanjut, sehingga diperlukan proses normalisasi untuk menyamakan skala data. **Tabel 3** menunjukkan hasil normalisasi fitur LRFM dan nilai CLV setiap pelanggan.

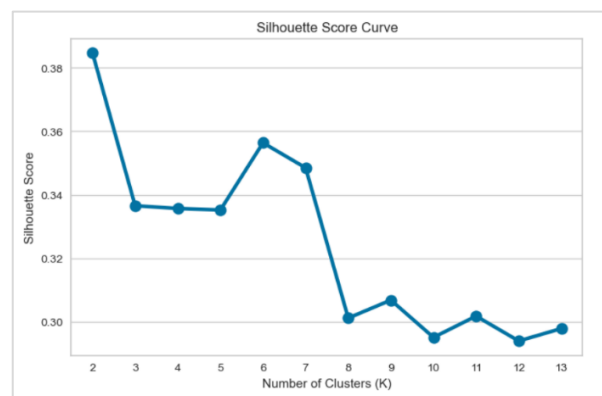
Untuk mendapatkan jumlah kluster yang memberikan hasil terbaik dalam analisis K-Means *clustering*, peneliti terlebih dahulu menggunakan *elbow method* untuk

menentukan titik optimal, kemudian melakukan validasi dengan *silhouette score* untuk memastikan kualitas pemisahan kluster. Setelah k telah ditentukan, nilai tersebut diterapkan dalam analisis kluster yang lebih mendalam.



Gambar 2. Hasil *Elbow Method*

Hasil dari *elbow method* ditunjukkan pada **Gambar 2** yang memvisualisasikan hubungan antara jumlah kluster dan nilai WCSS. Seiring dengan peningkatan jumlah kluster, nilai WCSS menurun, tetapi penurunan ini mulai melambat pada rentang k=6. Hal ini menunjukkan bahwa titik optimal berada di nilai tersebut, karena diatas titik k=6, penurunan nilai WCSS mulai melambat. Hasil perhitungan *silhouette score* juga menunjukkan bahwa k=6 memiliki nilai yang lebih tinggi yaitu sebesar 0.356, seperti yang ditampilkan pada **Gambar 3**. Nilai tersebut tergolong dalam kategori yang dapat diterima [19], mengindikasikan pemisahan kluster cukup baik. Selain itu, pemilihan k=6 juga didasarkan pada pertimbangan *trade-off* antara kualitas kluster dan interpretabilitas [20].



Gambar 3. Hasil *Silhouette Score*

Tabel 4. Statistik Deskriptif untuk Setiap Kluster

cluster	L	R	F	M	CLV
0	320	16	16	565,878	2,121,158
1	0	212	1	371,810	83,059
2	34	66	2	323,010	192,625
3	319	7	28	4,888,389	25,780,069
4	219	37	4	416,735	442,251
5	0	165	1	2,573,813	816,943

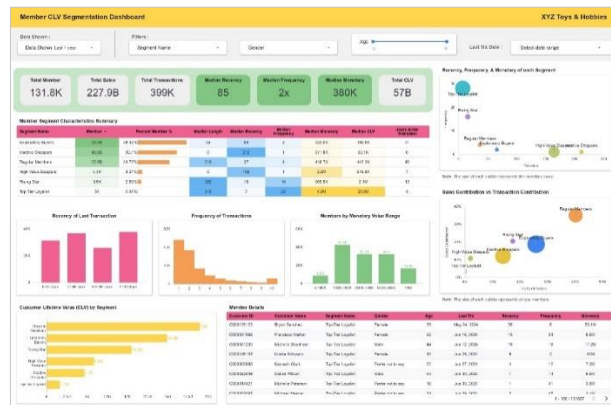
Pemilihan $k=6$ juga diputuskan bersama tim karena hasil *clustering* ini menunjukkan segmentasi yang logis dan bermakna, karena mampu menangkap karakteristik serta pola perilaku pelanggan dengan lebih jelas, serta relevan dengan strategi bisnis ritel XYZ. Dengan demikian, $k=6$ ini dianggap relevan untuk dijadikan dasar dalam analisis segmentasi pelanggan. Detail karakteristik setiap kluster dapat dilihat pada **Tabel 4**. Dalam analisis ini, median digunakan sebagai ukuran pemusatan data untuk menjelaskan karakteristik masing-masing kluster. Hal ini disebabkan oleh nilai *monetary* dan CLV memiliki *outlier* yang membuat mean tidak representatif. Median memberikan gambaran nilai tengah yang lebih akurat dan lebih robust terhadap *outlier* [21]. Hasil analisis ini memperoleh enam segmen pelanggan dengan karakteristik sebagai berikut:

1. Kluster 0 diberi label *Rising Star*. Pelanggan dalam kluster ini memiliki pola belanja yang aktif dan memiliki nilai CLV yang tergolong tinggi, serta memiliki potensi pertumbuhan yang menjanjikan dan dapat berkembang menjadi *loyalist*.
2. Kluster 1 diberi label *Inactive Shoppers*. Pelanggan dalam kluster ini sudah lama tidak aktif, memiliki nilai R yang tinggi dan nilai FM yang rendah dapat menurunkan potensi nilai jangka panjang terhadap bisnis. Kelompok ini berpotensi untuk *churn*.
3. Kluster 2 diberi label *Exploratory Buyers*. Pelanggan dalam kluster ini masih berada dalam tahap eksplorasi terhadap produk atau layanan, belum memiliki pola belum konsisten, namun dapat ditingkatkan loyalitasnya.
4. Kluster 3 diberi label *Top-Tier Loyalist*. Kelompok ini merupakan pelanggan paling setia dengan nilai RFM yang sangat baik, serta menjadi kontributor utama dalam bisnis, sehingga perlu untuk menjaga kepuasan dan mendorong loyalitas jangka panjang.
5. Kluster 4 diberi label *Regular Members*. Pelanggan dalam kluster ini memiliki pola belanja cukup konsisten dengan frekuensi belanja rutin. Stabilitas perilaku belanja mereka menjadikan kluster ini sebagai segmen yang layak untuk dipertahankan.
6. Kluster 5 diberi label *High-Value Sleepers*. Kluster ini berisi pelanggan yang memiliki nilai *monetary* yang tinggi, tapi sangat jarang atau bahkan hanya pernah belanja sekali dan sudah lama tidak berbelanja. Walaupun nilai CLV masih tergolong cukup besar, namun pelanggan dalam kluster ini berisiko tinggi untuk *churn*.

3.3. Perancangan Dashboard

Dashboard dirancang untuk membantu tim marketing dalam memahami pelanggannya berdasarkan pola perilaku dan nilai potensi, serta mendukung pengambilan keputusan untuk strategi retensi dan kampanye pemasaran. **Gambar 4** menunjukkan hasil rancangan

dashboard ritel XYZ yang berisi ringkasan karakteristik setiap segmen pelanggan, serta menyajikan informasi mengenai LRFM dan CLV yang dihasilkan oleh pelanggan. Visualisasi yang dipilih bertujuan untuk menunjukkan kontribusi tiap segmen terhadap bisnis, baik dari jumlah pelanggan, nilai transaksi, hingga potensi nilai pelanggan kedepannya.



Gambar 4. Hasil Rancangan Dashboard Ritel XYZ

Berikut merupakan informasi yang ditampilkan dalam bentuk *chart* dan filter pada dashboard:

1. Filter data. Dashboard dilengkapi dengan berbagai fitur filter yang membantu pengguna untuk menyesuaikan tampilan data sesuai kebutuhan analisis. Dengan adanya filter ini, pengguna dapat mempersempit cakupan data yang ditampilkan berdasarkan periode waktu yang digunakan, nama segmen pelanggan, dan jenis kelamin dalam bentuk *dropdown*, usia dalam bentuk *slider*, serta tanggal terakhir belanja dalam bentuk *date range control*. Fitur filter data dapat dilihat pada **Gambar 5**.



Gambar 5. Fitur Filter

2. Ringkasan KPI. Visualisasi berupa *scoreboard* ini menampilkan gambaran awal untuk memahami kondisi pelanggan secara keseluruhan, seperti jumlah pelanggan, total penjualan, total transaksi, median *recency*, *frequency*, dan *monetary*, serta total CLV. Ringkasan ini membantu pengguna untuk memahami performa bisnis secara cepat tanpa perlu menganalisis grafik yang lebih kompleks. **Gambar 6** menunjukkan ringkasan KPI pada dashboard.



Gambar 6. Ringkasan KPI

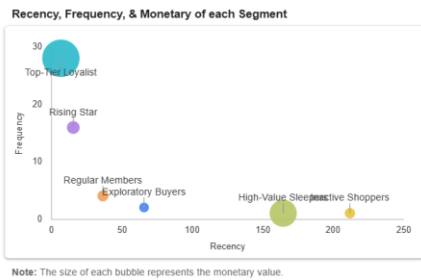
3. Ringkasan karakteristik setiap segmen pelanggan. **Gambar 7** menunjukkan tabel yang

berisi ringkasan metrik yang menggambarkan karakteristik setiap segmen pelanggan secara kuantitatif, seperti informasi jumlah pelanggan dalam setiap segmen, persentase kontribusinya terhadap populasi pelanggan, nilai median LRFM dan CLV, serta jarak antar transaksi. Informasi ini bertujuan untuk membandingkan performa antar segmen dan membantu tim marketing menentukan pendekatan pemasaran yang sesuai untuk setiap segmen pelanggan.

Segment Name	Members	Percent Member %	Median Recency	Median Frequency	Median Monetary	Median CLV	90th Order Threshold
Top-Tier Loyalist	11	0.5%	155	7	1,000	10,000	8
Rising Star	100	5.0%	100	15	500	2,500	15
Regular Members	1,000	50.0%	50	10	100	500	10
High-Value Sleepers	100	5.0%	150	5	1,000	5,000	12
Inactive Shoppers	1,000	50.0%	20	2	50	250	5
Exploratory Buyers	100	5.0%	30	8	200	1,000	8

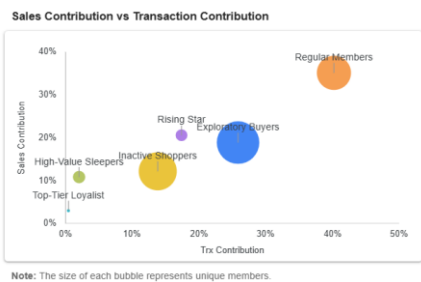
Gambar 7. Ringkasan Karakteristik Setiap Segmen

- RFM setiap segmen. Visualisasi ini menggambarkan karakteristik dan hubungan RFM pada setiap segmen dalam bentuk visual, seperti pada Gambar 8. Pengguna dapat lebih memahami sebaran nilai RFM pada setiap segmen pelanggan, serta mengidentifikasi perilaku dan kontribusi masing-masing kelompok pelanggan.



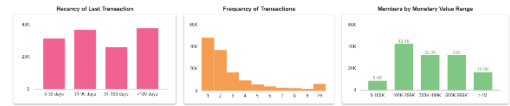
Gambar 8. Grafik RFM Setiap Segmen

- Kontribusi penjualan dan transaksi pelanggan. Gambar 9 menunjukkan visualisasi yang dirancang untuk memberikan gambaran seberapa besar kontribusi setiap segmen, dari sisi total penjualan dan jumlah transaksi. Tim marketing dapat mengidentifikasi segmen yang paling berpengaruh terhadap bisnis dan memahami apakah kontribusi segmen berasal dari jumlah pelanggan atau perilaku pembelian yang intensif.



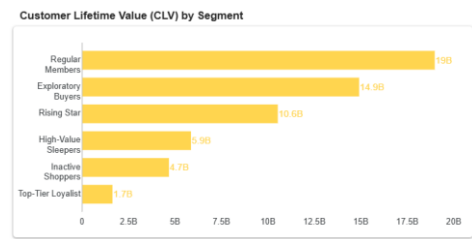
Gambar 9. Kontribusi Penjualan dan Transaksi Setiap Segmen

- Distribusi pelanggan berdasarkan RFM. Ketiga column chart ini menunjukkan jumlah pelanggan berdasarkan rentang *recency*, *frequency*, dan *monetary*. Informasi ini dapat membantu tim melihat pola penyebaran pelanggan dan mengaitkan sebaran tersebut dengan performa metrik. Gambar 10 menunjukkan ketiga grafik distribusi pelanggan berdasarkan RFM.



Gambar 10. Distribusi Pelanggan Berdasarkan RFM

- Nilai CLV yang dihasilkan oleh setiap segmen. Visualisasi ini dapat membantu tim mengidentifikasi segmen yang paling potensial dalam memberikan kontribusi bisnis ke depan, baik dari sisi loyalitas maupun nilai transaksi. Informasi ini membantu tim memprioritaskan segmen utama dan merancang strategi pemasaran yang lebih efektif. Gambar 11 menunjukkan grafik total nilai CLV setiap segmen pelanggan.



Gambar 11. Grafik Nilai CLV Setiap Segmen

- Detail Pelanggan. Gambar 12 menampilkan tabel yang berisi informasi yang lebih spesifik mengenai identitas, karakteristik dan perilaku setiap pelanggan, seperti ID pelanggan, nama pelanggan, nama segmen, jenis kelamin, usia, tanggal terakhir transaksi, nilai RFM, total pembelian, dan perkiraan nilai CLV. Tabel diurutkan berdasarkan nilai CLV tertinggi sehingga memudahkan tim marketing untuk mengenali pelanggan yang diperkirakan memiliki kontribusi terbesar ke depannya. Visualisasi ini bermanfaat ketika tim marketing ingin menelusuri atau melakukan analisis mendalam pelanggan secara individual.

Customer ID	Customer Name	Segment Name	Gender	Age	Last Tx	Recency	Frequency	Monetary
C00001010	Syam Prasasti	Top-Tier Loyalist	Female	30	May 19, 2024	96	2	10,000
C00001002	Yusuf Haniyati	Top-Tier Loyalist	Female	20	Jul 16, 2024	100	13	8,000
C00001020	Wahne Koolhaas	Top-Tier Loyalist	Male	44	Jul 15, 2024	10	10	17,200
C00000100	Nico Schrijver	Top-Tier Loyalist	Female	30	Jul 23, 2024	9	2	4,100
C00000000	Kenneth Cook	Top-Tier Loyalist	Female	19	Jul 17, 2024	4	0	7,000
C00000000	David Heston	Top-Tier Loyalist	Male	20	Jul 20, 2024	1	0	9,600
C00001400	Wahne Freeman	Top-Tier Loyalist	Female	19	Jul 20, 2024	1	41	9,600

Gambar 12. Tabel Detail Pelanggan

Dengan informasi dan *insight* yang disajikan, dashboard dapat membantu tim marketing dalam

memahami perilaku dan potensi nilai pelanggan secara lebih dalam, merancang strategi pemasaran yang sesuai, meningkatkan retensi pelanggan dan menentukan target pelanggan yang tepat berdasarkan perilaku dan nilai, serta mengidentifikasi pelanggan yang perlu dipertahankan atau memiliki potensi yang dapat ditingkatkan.

4. Kesimpulan

Penelitian ini berhasil mengintegrasikan analisis *clustering* dalam segmentasi pelanggan serta merancang dashboard interaktif yang menyajikan informasi relevan sesuai dengan kebutuhan bisnis perusahaan XYZ. Berdasarkan hasil *elbow method* dan *silhouette score*, jumlah kluster optimal yang dipilih adalah $k=6$ dengan nilai sebesar 0.356 sehingga terbentuk enam segmen pelanggan yang memiliki karakteristik unik. Hasil *clustering* kemudian digunakan untuk merancang dashboard yang berisi ringkasan karakteristik setiap segmen pelanggan, serta menyajikan informasi mengenai pola perilaku dan nilai CLV yang dihasilkan oleh pelanggan. Namun, dashboard yang dirancang masih memiliki fitur dan informasi yang terbatas. Oleh karena itu, pengembangan selanjutnya disarankan untuk menambahkan analisis lain seperti penggunaan *market basket analysis* untuk memahami pola pembelian produk atau penggunaan *machine learning* untuk memprediksi pola perilaku pelanggan, kemudian diintegrasikan ke dalam dashboard sehingga tim marketing dapat mengidentifikasi pelanggan potensial dari berbagai dimensi. Dengan demikian, penelitian ini menunjukkan bahwa kombinasi antara analisis data berbasis *machine learning* dan perancangan dashboard dapat membantu pengguna dalam memahami, mengeksplorasi, dan menganalisis perilaku belanja pelanggan.

REFERENSI

- [1] N. L. Rane, A. Achari and S. P. Choudhary, "Enhancing customer loyalty through quality of service: Effective strategies to improve customer satisfaction, experience, relationship, and engagement," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 5, no. 5, pp. 427-452, 2023.
- [2] N. I. Putri, R. Komalasari and Z. Munawar, "Pentingnya Keamanan Data dalam Intelijen Bisnis," *J-SIKA| Jurnal Sistem Informasi Karya Anak Bangsa*, vol. 1, no. 2, pp. 41-48, 2020.
- [3] F. J. Reh, "Understanding Pareto's principle-the 80-20 rule," *The Balance*, vol. 8, 2017.
- [4] Y. Sun, H. Liu and Y. Gao, "Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model," *Heliyon*, vol. 9, no. 2, 2023.
- [5] R. A. Megantara, F. Alzami, A. Akrom, R. A. Pramunendar, D. P. Prabowo, S. Wibowo and R. , "RFM Analysis for Customer Lifetime Value with PARETO/NBD Model in Online Retail Dataset," *Moneter: Jurnal Keuangan dan Perbankan*, vol. 11, no. 2, pp. 284-290, 2023.
- [6] F. Yoseph, N. H. Ahamed Hassain Malim, M. Heikkila, A. Brezulianu, O. Geman and N. A. Paskhal Rostam, "The impact of big data market segmentation using data mining and clustering techniques," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 5, pp. 6159-6173, 2020.
- [7] N. T. Lee, H. C. Lee, J. Hsin and S. H. Fang, "Prediction of customer behavior changing via a hybrid approach," *IEEE Open Journal of the Computer Society*, vol. 5, pp. 27-38, 2024.
- [8] A. M. Haddadi and H. Hamidi, "A hybrid model for improving customer lifetime value prediction using stacking ensemble learning algorithm," *Computers in Human Behavior Reports*, vol. 18, 2025.
- [9] R. P. Kuswidyawan, D. Trisnawarman and J. Hendryli, "Customer segmentation with K-means clustering algorithm for E-commerce sales visualization," *AIP Conference Proceedings*, vol. 2680, no. 1, 2023.
- [10] R. Zimmermann and P. Brandtner, "From Data to Decisions: Optimizing Supply Chain Management with Machine Learning-Infused Dashboards.," *Procedia Computer Science*, vol. 237, pp. 955-964, 2024.
- [11] F. Marisa, S. S. S. Ahmad, Z. I. M. Yusof, F. 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.
- [12] M. Singh, "Predictive Analytics in Evaluating Customer Lifetime Value: A Paradigm Shift in Modern Marketing," *International Journal of Science and Research (IJSR)*, vol. 12, no. 6, pp. 1771-1774, 2023.
- [13] B. C. Laksono and I. Y. Wulansari, "Pemodelan Dan Penerapan Metode Rfm Pada Estimasi Nilai Konsumen (Customer Lifetime Value) Menggunakan K-Means Clustering Machine Learning," *Seminar Nasional Official Statistics*, vol. 2020, no. 1, pp. 1277-1285, 2020.
- [14] T. M. Kodinariya and P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering," *International Journal*, vol. 1, no. 6, pp. 90-95, 2013.
- [15] V. A. Kherdekar and S. Naik, "Impact of Feature Normalization Techniques For Recognition of Speech for Mathematical Expression," in *Smart Trends in Computing and Communications*, Springer, 2024, pp. 109-117.
- [16] A. Et-Taleby, M. Boussetta and M. Benslimane, "Faults Detection for Photovoltaic Field Based on

K-Means, Elbow, and Average Silhouette Techniques through the Segmentation of a Thermal Image," *International Journal of Photoenergy*, vol. 2020, no. 1, 2020.

- [17] C. Yuan and H. Yang, "Research on K-value selection method of K-means clustering algorithm," *J*, vol. 2, no. 2, pp. 226-235, 2019.
- [18] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," *2020 IEEE 7th international conference on data science and advanced analytics (DSAA)*, pp. 747-748, 2020.
- [19] E. Savran, E. Karpat and F. Karpat, "Energy-Efficient Anomaly Detection and Chaoticity in Electric Vehicle Driving Behavior," *Sensors*, vol. 24, no. 17, 2024.
- [20] D. Bertsimas, A. Orfanoudaki and H. Wiberg, "Interpretable Clustering via Optimal Trees," *arXiv preprint arXiv:1812.00539*, 2018.
- [21] L. A. García-Escudero, A. Gordaliza, C. Matrán and A. Mayo-Isacar, "A review of robust clustering methods," *Advances in Data Analysis and Classification*, vol. 4, pp. 89-109, 2010.

Daniela Aedo, Mahasiswa S1 Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jakarta.

Dedi Trisnawarman, Dosen Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jakarta