

Pemasaran Efektif melalui Optimasi *RFM Analysis* Berdasarkan *K-Means Clustering*

Mas'ud Ghania Al-Fathoni *, Teti Sofia Yanti

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

ghaniakevin@gmail.com, tetisofiyanti@unisba.ac.id

Abstract. The digitalization trend demands an accurate understanding of customer behavior. RFM (Recency, Frequency, Monetary) Analysis with K-Means Clustering offers objective segmentation for big data, overcoming the limitations of traditional scoring methods. This study aims to identify customer purchasing patterns of PT Maha Nagari Nusantara and formulate effective marketing strategies. This quantitative-exploratory study analyzed 86,605 customer transactions of PT Maha Nagari Nusantara on Shopee (2024). The data was processed into RFM, then Min-Max normalization was performed. Exploration of k values with values of 2 to 6 with evaluation of results using the Silhouette Coefficient, choosing $k = 2$ (0.5906) as the most objective data structure. K-Means identified two main segments, namely "Active & Potential Customers" as many as 40,404 customers with low Recency an average of 100 days with Frequency-Monetary values above average. Prioritize strategies with loyalty building, upselling, cross-selling. "Passive & Risky Customers" totaling 46,201 customers with a very high Recency of 270 days on average indicates that they have been inactive for a long time, at risk of churn. Strategy priority with reactivation campaigns and increased engagement. A differentiated dual marketing strategy, focusing on active retention and passive reactivation, is essential company to optimize budgets and increase retention.

Keywords: *Marketing Analytics, RFM Analysis, K-Means Clustering.*

Abstrak. Tren digitalisasi menuntut pemahaman perilaku pelanggan yang akurat. RFM (Recency, Frequency, Monetary) Analisis dengan K-Means Clustering menawarkan segmentasi objektif untuk data besar, mengatasi keterbatasan metode skoring tradisional. Penelitian ini bertujuan mengidentifikasi pola pembelian pelanggan PT Maha Nagari Nusantara dan merumuskan strategi pemasaran efektif. Penelitian kuantitatif-eksploratif ini menganalisis 86.605 transaksi pelanggan PT Maha Nagari Nusantara di Shopee (2024). Data diolah menjadi RFM, lalu dilakukan normalisasi Min-Max. Eksplorasi nilai k dengan nilai 2 hingga 6 dengan evaluasi hasil menggunakan Silhouette Coefficient, memilih $k=2$ (0,5906) sebagai struktur data paling objektif. K-Means mengidentifikasi dua segmen utama yakni "Pelanggan Aktif & Berpotensi" sebanyak 40.404 pelanggan dengan Recency rendah rata-rata 100 hari dengan nilai Frequency-Monetary di atas rata-rata. Prioritas strategi dengan pembinaan loyalitas, upselling, cross-selling. "Pelanggan Pasif & Berisiko" berjumlah 46.201 pelanggan dengan Recency sangat tinggi rata-rata 270 hari menandakan sudah lama tidak aktif, berisiko churn. Prioritas strategi dengan kampanye reaktivasi dan peningkatan keterlibatan. Strategi pemasaran ganda yang terdiferensiasi, berfokus pada retensi aktif dan reaktivasi pasif, esensial bagi perusahaan untuk optimasi anggaran dan peningkatan retensi.

Kata Kunci: *Analisis Pemasaran, RFM Analisis, K-Means Clustering.*

A. Pendahuluan

Adanya fleksibilitas menjadi alasan bagi masyarakat untuk beralih belanja secara digital dengan melalui berbagai e-commerce. Pada tahun 2023, total nilai transaksi sektor e-commerce di Indonesia tercatat mencapai Rp 453 triliun. Meskipun data resmi untuk tahun 2024 belum tersedia, Menteri Perdagangan, Budi Santoso, memperkirakan bahwa angka tersebut akan meningkat sebesar 2,8 persen, mencapai Rp 487 triliun (Pramudhiaz, 2024). Proyeksi ini mencerminkan semakin luasnya adopsi platform e-commerce sebagai sarana utama bagi masyarakat dalam berbelanja. Pemasaran efektif dapat dilakukan jika adanya pendekatan juga pemahaman pelanggan, tentu caranya saat ini akan berbeda dengan konvensional (Sheth, 2021). Menganalisa perilaku belanja pelanggan di era digital ini menjadi sangat mudah karena memperoleh informasi mengenai pasar hingga perilaku pelanggan dari adanya nilai pasti dari metrik-metrik yang disajikan.

Di sinilah peran analisis statistika menjadi sangat krusial dalam mendukung pengambilan keputusan yang berbasis data. Salah satu pendekatan yang bisa digunakan dalam segmentasi pelanggan adalah penggunaan analisis *Recency*, *Monetary*, *Frequency*, yang mengevaluasi pelanggan berdasarkan tiga dimensi utama: mengukur interval waktu sejak transaksi terakhir dilakukan (*Recency*), mencerminkan tingkat frekuensi pembelian yang dilakukan oleh pelanggan dalam periode tertentu (*Frequency*), dan merepresentasikan total nilai transaksi yang telah dikeluarkan oleh pelanggan (*Monetary*). Pengelompokan pelanggan dengan berdasarkan skor RFM tidak lagi menjadi efektif dikala data yang digunakan sangatlah besar. Oleh karena itu, menentukan segmentasi pelanggan menjadi sangat penting. Salah satu metode yang dapat digunakan untuk segmentasi adalah algoritma K-Means Clustering. Dalam konteks analisis RFM, K-Means memungkinkan pembentukan Cluster pelanggan berdasarkan kedekatan spasial dalam ruang multidimensi (Rumiarti & Budi, 2017).

Customer Relationship Management (CRM) adalah pendekatan strategis dalam bisnis yang bertujuan untuk mengelola interaksi dengan pelanggan guna meningkatkan loyalitas, kepuasan, dan nilai pelanggan dalam jangka panjang. CRM mulai digunakan tahun 1990, saat itu istilah ini sering digunakan untuk menggambarkan solusi berbasis teknologi yang dirancang untuk meningkatkan interaksi dan layanan pelanggan (Buttle & Makklan, 2015). Dengan semakin ketatnya persaingan di berbagai industri, CRM menjadi elemen kunci dalam strategi pemasaran yang berorientasi pada pelanggan, memungkinkan bisnis untuk memberikan layanan yang lebih relevan dan responsif terhadap kebutuhan mereka (Guerola-Navarro et al., 2021).

Analisis pemasaran membantu perusahaan memahami dinamika pasar secara lebih mendalam, sehingga dapat merancang strategi yang lebih tepat sasaran (Iacobucci et al., 2019). Selain itu, analisis pemasaran juga bertujuan untuk meningkatkan efisiensi anggaran pemasaran dengan memastikan bahwa setiap investasi yang dilakukan memberikan hasil yang optimal (France & Ghose, 2018).

RFM adalah teknik segmentasi pelanggan yang digunakan untuk mengukur dan menganalisis nilai pelanggan berdasarkan tiga aspek utama, yakni seberapa baru pelanggan melakukan transaksi (*Recency*), seberapa sering mereka melakukan pembelian (*Frequency*), dan berapa banyak uang yang telah pelanggan belanjakan (*Monetary*) (Sabuncu et al., 2020). Metode ini menjadi alat yang sangat berguna dalam dunia pemasaran karena memberikan pemahaman yang lebih dalam tentang perilaku pelanggan, memungkinkan bisnis untuk mengelompokkan mereka berdasarkan kebiasaan pembelian dan tingkat loyalitas.

Metode normalisasi perlu digunakan dalam penelitian ini karena skala dari ketiga variabel RFM sangat berbeda. Normalisasi yang digunakan adalah normalisasi min-max. Normalisasi Min-Max merupakan salah satu teknik penskalaan data yang bertujuan untuk mengubah rentang nilai suatu variabel ke dalam skala tertentu, umumnya antara 0 dan 1 (Sholeh et al., 2022). Dalam konteks K-Means Clustering, normalisasi Min-Max memungkinkan setiap nilai memiliki kontribusi yang seimbang dalam proses pembentukan Cluster, sehingga hasil klusterisasi menjadi lebih representatif dan tidak terpengaruh oleh perbedaan unit pengukuran.

K-Means clustering adalah metode unsupervised learning yang digunakan untuk mengelompokkan data berdasarkan kemiripan pola dalam sekumpulan dataset (Rumiarti & Budi, 2017). K-Means memiliki karakteristik utama yang menjadikannya pilihan dalam klusterisasi, seperti kemampuannya dalam mengelompokkan data ke dalam sejumlah Cluster yang ditentukan sebelumnya, bekerja secara iteratif untuk menyesuaikan posisi centroid, serta memiliki fleksibilitas dalam menangani data dengan dimensi yang kompleks (Pham et al., 2005). Salah satu manfaat utama dari metode ini adalah kemampuannya dalam mengidentifikasi pola yang tidak terlihat secara eksplisit

dalam dataset mentah, sehingga dapat digunakan untuk memahami perilaku pelanggan dan merancang strategi bisnis yang lebih efektif.

Evaluasi hasil clustering dilakukan untuk memastikan bahwa pengelompokan yang dihasilkan sudah optimal dan sesuai dengan tujuan analisis. Silhouette Coefficient digunakan dalam evaluasi pengelompokan. Silhouette Coefficient dihitung dengan mempertimbangkan jarak intra dan inter Cluster (Rousseeuw, 1987). Mengukur konsistensi dalam satu cluster dibandingkan dengan cluster lainnya, dengan nilai berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan klastering yang baik (Arbelaitz et al., 2013).

Penelitian ini berangkat dari pertanyaan-pertanyaan masalah seperti bagaimana nilai *Recency*, *Frequency*, dan *Monetary* menggambarkan perilaku pelanggan di platform Shopee tahun 2024, bagaimana algoritma K-Means Clustering dapat diintegrasikan dengan model RFM untuk mengoptimalkan segmentasi pelanggan, bagaimana karakteristik dari setiap segmen pelanggan yang terbentuk, dan Rekomendasi strategi pemasaran apa yang dapat diberikan berdasarkan hasil segmentasi tersebut agar memperoleh pemasaran efektif.

Berdasarkan permasalahan tersebut penelitian ini bertujuan untuk Menganalisis penerapan model *Recency*, *Frequency*, *Monetary* (RFM) dalam mengidentifikasi pola pembelian pelanggan pada perusahaan retail, Mengaplikasikan algoritma K-Means Clustering dalam mengelompokkan pelanggan berdasarkan hasil analisis RFM, Mendeskripsikan karakteristik setiap segmen pelanggan yang terbentuk, dan memberikan rekomendasi strategi pemasaran yang relevan berdasarkan hasil segmentasi untuk meningkatkan retensi pelanggan dan menghasilkan pemasaran yang efektif. Setidaknya akan bermanfaat untuk kontribusi terhadap pengembangan ilmu pengetahuan dalam bidang CRM, Analisis Pemasaran, serta khususnya terkait integrasi model RFM dan algoritma K-Means Clustering untuk segmentasi pelanggan juga membantu perusahaan ritel untuk mengidentifikasi segmentasi pelanggan yang lebih akurat, sehingga dapat meningkatkan efektivitas kampanye pemasaran.

B. Metode

Tuliskan Penelitian ini dilakukan dengan pendekatan kuantitatif dan eksploratif. Data yang digunakan merupakan data sekunder dari Torch.id, sebuah perusahaan retail fashion di Kota Bandung. Data yang digunakan adalah sebanyak 86.605 transaksi sepanjang tahun 2024 pada platform *e-commerce* Shopee, dengan mencakup informasi: Kode Pembeli, Harga Setelah Diskon dan Waktu Pesanan.

Dapat dilengkapi dengan gambar atau bagan yang berisi mengenai tahapan penelitian. Seperti contoh berikut ini.



Gambar 1. Tahapan Penelitian

Penelitian ini mengadopsi berbagai metode analisis data untuk memahami perilaku pelanggan dan mengoptimalkan strategi pemasaran perusahaan. Metode analisis yang digunakan seperti: (1) Analisis Deskriptif, (2) Analisis RFM, (3) Normalisasi Min-Max, (4), K-Means Clustering, (5) Evaluasi hasil Silhouette Coefficient, dan (6) Interpretasi hasil dan pemberian label untuk setiap Cluster yang terbentuk.

Analisis Deskriptif

Analisis deskriptif digunakan sebagai langkah awal guna mendeskripsikan dan memvisualisasi karakteristik pelanggan berdasarkan data yang tersaji. Memberikan gambaran menyeluruh tentang kebiasaan pembelian mereka.

Analisis RFM

1. *Recency*

Variabel ini untuk mengukur seberapa baru pelanggan melakukan transaksi. Dengan rumus sebagai berikut:

$$Recency(i) = Max(t) - t_i \forall i, n = 0, 1, 2, 3, \dots n \quad \dots (1)$$

2. *Frequency*

Variabel ini mengukur seberapa sering pelanggan melakukan transaksi. Dengan rumus sebagai berikut:

$$Frequency(i) = \sum k_t(i) \quad \dots (2)$$

3. *Monetary*

Variabel ini menghitung berapa jumlah yang dikeluarkan oleh setiap pelanggan dalam seluruh transaksinya. Dengan rumus sebagai berikut:

$$Monetary(i) = \sum_{n=1}^p H_p \times m_p(i), \forall i, n = 1, 2, 3, \dots n \quad \dots (3)$$

Normalisasi Min-Max

Secara matematis, normalisasi Min-Max didefinisikan sebagai berikut:

$$X_{ci} = \frac{O_i^X - \min^X}{\max^X - \min^X} \quad \dots (4)$$

di mana O_i^X adalah data awal dari variabel X, sedangkan $\max^X$ dan $\min^X$ masing-masing merupakan nilai maksimum dan minimum dari variabel tersebut (Virgiawan A, 2015).

K-Means Clustering

K-Means meminimalkan variabilitas Cluster dan meningkatkan pemisahan antar Cluster, sehingga pola tersembunyi dalam data dapat diungkap lebih jelas. Eksplorasi nilai k dengan melihat segmentasi perusahaan yang sudah mempunyai konsep 6 segmentasi pelanggan. Sehingga, eksplorasi nilai k dengan nilai 2 hingga 6. Hal ini, untuk menyesuaikan jumlah cluster agar relevan secara praktis dan bisnis. Proses kerja algoritma selanjutnya melibatkan dalam penentuan secara acak K titik awal dari data, yang merupakan pusat dari masing-masing Cluster dengan rumus sebagai berikut:

$$c_i = \frac{\sum_{j=1}^M x_j}{M} \quad \dots (5)$$

Dalam algoritma ini, ukuran jarak yang digunakan secara umum adalah jarak Euclidean Menurut Rojlerjanya (2019) rumus perhitungan jarak Euclidean yang telah disesuaikan dengan nilai RFM adalah sebagai berikut:

$$d_{Euclidean}(x, y) = \sqrt{(R_x - R_y)^2 + (F_x - F_y)^2 + (M_x - M_y)^2} \quad \dots (6)$$

Evaluasi Hasil

Mengukur konsistensi dalam satu cluster dibandingkan dengan cluster lainnya, dengan nilai berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan klastering yang baik (Arbelaitz et al., 2013). Berikut rumus dari Silhouette Coefficient:

$$S = \frac{\sum_{i=1}^N \frac{b_i - a_i}{\max(a_i, b_i)}}{N} \quad \dots (7)$$

C. Hasil Penelitian dan Pembahasan

Bagian ini akan menyajikan ilustrasi perhitungan manual untuk memahami bagaimana nilai *Recency*, *Frequency*, dan *Monetary* (RFM) dihasilkan.

Analisis RFM

Analisis RFM membentuk sebuah tabel yang berisikan informasi transaksi dari setiap pelanggan sebagai berikut:

Tabel 1. Tabel RFM

Kode Pembeli	<i>Recency</i> (Hari)	<i>Frequency</i> (Jumlah)	<i>Monetary</i> (Ribu Rupiah)
100000001	366	1	103999
100000002	229	2	205998
100000003	304	2	200838
100000004	241	2	202998
100000005	366	1	121999
100000006	366	2	233998

Statistik deskriptif dari variabel *Recency*, *Frequency*, dan *Monetary* (RFM) memberikan gambaran umum mengenai perilaku belanja pelanggan.

Tabel 2. Statistik Deskriptif RFM

Statistik	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>
Jumlah	86605	86605	86605
Rata-rata	191,3727	1,217043	256726,6
Std	99,76226	0,733802	265881,3
Min	1	1	0
Q1	114	1	134999
Q2	197	1	168999
Q3	273	1	295999
Max	366	42	12170174

Nilai rata-rata *Recency* sebesar 191 hari, dengan nilai median 197 hari, mengindikasikan bahwa separuh dari basis pelanggan telah tidak aktif bertransaksi selama lebih dari enam bulan. Dari sisi *Frequency*, nilai rata-rata hanya sebesar 1,21 kali transaksi per tahun, di mana 75% pelanggan dilihat dari kuartil ketiga hanya melakukan satu kali pembelian. Pola ini menunjukkan bahwa mayoritas pelanggan cenderung bersifat transaksional. Sementara itu, nilai rata-rata *Monetary* per pelanggan adalah sebesar Rp 256.726.

Normalisasi Min-Max

Tahapan selanjutnya adalah normalisasi data mengingat perbedaan skala pada ketiga variabel RFM. Normalisasi dilakukan dengan metode Min-Max. Berikut pada tabel 4.9 adalah data yang telah dilakukan normalisasi pada masing-masing variabel:

Tabel 3. Data Setelah Normalisasi Min-Max

Kode Pembeli	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>
100000001	1	0	0,008545
100000002	0,624658	0,02439	0,016926
100000003	0,830137	0,02439	0,016502
⋮	⋮	⋮	⋮
100086604	0	0,02439	0,033524
100086605	0	0	0,014215
100086606	0	0,02439	0,043138

Semua nilai telah diubah ke rentang 0–1 agar setiap variabel memiliki bobot yang seimbang.

Penerapan Algoritma K-Means

Dalam penentuan nilai k , peneliti melakukan pengujian dengan eksplorasi nilai k dari 2 hingga 6. Proses iterasi akan terus diulang dengan menghitung jarak setiap pelanggan ke centroid baru, mengelompokkan kembali, dan memperbarui posisi centroid. Iterasi akan berhenti jika tidak ada lagi pelanggan yang berpindah Cluster. Ini menandakan bahwa pengelompokan sudah stabil dan optimal atau telah mencapai konvergensi. Pada titik ini, setiap pelanggan telah secara definitif masuk ke dalam salah satu Cluster, dan hasil akhir segmentasi telah terbentuk.

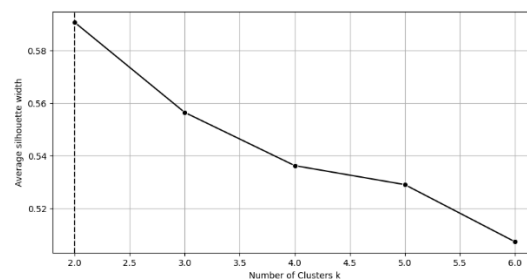
Evaluasi Hasil

Algoritma K-Means dijalankan dengan jumlah Cluster yang berbeda, mulai dari $k=2$ hingga $k=6$. Nilai Silhouette Coefficient dihitung untuk setiap hasil. Nilai yang lebih tinggi mendekati 1 menunjukkan bahwa objek dalam satu Cluster memiliki kemiripan yang tinggi dan terpisah dengan baik dari Cluster lainnya.

Tabel 4. Nilai Silhouette

Jumlah Cluster	Nilai Silhouette Coefficient
2	0,5906
3	0,5564
4	0,5361
5	0,5289
6	0,5072

Berdasarkan tabel di atas, dapat disimpulkan bahwa $k=2$ menghasilkan nilai Silhouette Coefficient tertinggi, yaitu 0,5906. Menurut kriteria interpretasi pada Tabel 2.3, nilai ini masuk dalam kategori "klasterisasi cukup baik".



Gambar 2. Visualisasi Silhouette Coefficient

Visualisasi dari skor ini juga dapat dilihat pada diatas, di mana titik tertinggi pada grafik berada pada saat Jumlah Cluster (k) = 2, yang secara visual mengonfirmasi bahwa 2 Cluster adalah pilihan yang paling optimal secara statistik.

Pemilihan Segmentasi Final

Meskipun model $k=6$ menawarkan granularitas yang lebih tinggi, penelitian ini memutuskan untuk melanjutkan analisis mendalam menggunakan model $k=2$. Keputusan ini didasarkan pada pertimbangan mengutamakan objektivitas data karena tujuan utama penggunaan K-Means Clustering dalam penelitian ini adalah untuk menemukan segmentasi yang paling objektif berdasarkan data. Nilai Silhouette Coefficient untuk $k=6$ tergolong rendah (0,5072). Hal ini mengindikasikan bahwa jika data dipaksa terbagi menjadi 6 kelompok, batas antar segmen kemungkinan besar tidak jelas dan tumpang tindih. Ini berisiko menciptakan segmen-segmen yang perbedaannya tidak signifikan secara praktis.

Analisis Karakteristik Cluster

Setelah memilih model segmentasi optimal dengan $k=2$, tahap selanjutnya adalah menganalisis secara mendalam karakteristik dari setiap Cluster yang terbentuk.

- a) Cluster 1: Pelanggan Aktif dan Berpotensi

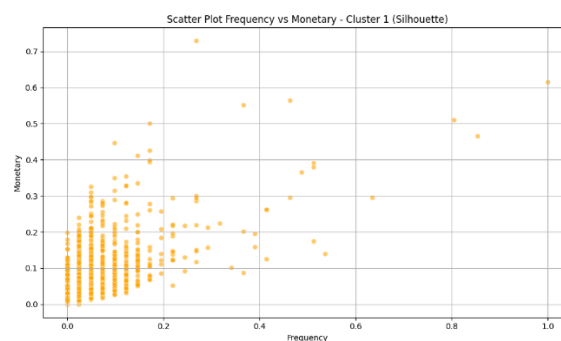
Cluster ini memiliki jumlah 40.437 pelanggan, dapat dianggap sebagai inti dari basis pelanggan

yang aktif saat ini. Berikut statistik deskriptif dari Cluster 1:

Tabel 5. Statistik Deskriptif Cluster 1

Statistik	Recency	Frequency	Monetary
Jumlah	40404	40404	40404
Rata-rata	100,82593	1.24406	269090,6
Std	55,37551	0,84326	291227,5
Min	1	1	0
Q1	51	1	134999
Q2	105	1	168999
Q3	149	1	299999
Max	186	42	8875987

Karakteristik paling menonjol dari Cluster ini adalah nilai Recency rata-rata yang rendah. Ini berarti mereka adalah kelompok pelanggan yang telah melakukan transaksi relatif sekitar 100 hari dengan median sebesar 105 hari. Hal ini menunjukkan bahwa sebagian besar pelanggan di dalam cluster ini masih cukup aktif melakukan transaksi dalam kurun waktu yang relatif dekat. Sebaran Recency pada cluster ini berkisar antara 1 hingga 186 hari, yang mengindikasikan bahwa hampir seluruh pelanggan melakukan transaksi dalam enam bulan terakhir. Pada aspek Frequency, nilai maksimum tercatat sebanyak 42 kali transaksi, meskipun median-nya tetap di angka 1. Artinya, mayoritas pelanggan hanya melakukan satu kali pembelian, namun terdapat beberapa pelanggan yang melakukan transaksi secara berulang, bahkan hingga puluhan kali. Dari sisi Monetary, rata-rata total nilai pembelian adalah sebesar Rp269.090, dengan sebaran nilai hingga Rp8.875.987. Nilai kuartil atas (Q3) sebesar Rp299.999 menunjukkan bahwa 75% pelanggan dalam cluster ini memiliki total belanja di bawah Rp300.000. Untuk melihat sebaran pelanggan pada cluster 1, bisa kita lihat pada gambar dibawah ini:



Gambar 3. Scatter Plot Cluster 1

Pada Scatter plot diatas antara Frequency dan Monetary untuk Cluster 1 menunjukkan adanya konsentrasi titik pada rentang frekuensi 1–5 kali dan nilai transaksi di bawah Rp500.000. Secara umum, pelanggan di Cluster 1 dapat dikategorikan sebagai pelanggan aktif dengan tingkat loyalitas yang baik, namun dengan kecenderungan untuk membeli produk dalam rentang harga menengah ke bawah.

b) Cluster 0: Pelanggan Pasif dan Berisiko (*At-Risk*)

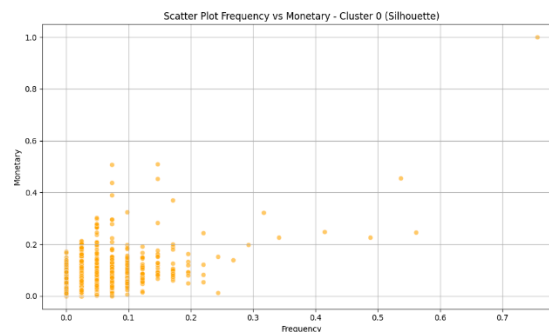
Cluster ini merupakan kelompok mayoritas dengan jumlah 46,168 pelanggan. Mereka adalah segmen yang memerlukan perhatian khusus dari perusahaan. Berikut statistik deskriptif dari klaster 0:

Tabel 6. Statistik Deskriptif Cluster 0

Statistik	Recency	Frequency	Monetary
Jumlah	46201	46201	46201
Rata-rata	270,55817	1,19342	245914
Std	50,34008	0,62154	241028,9
Min	186	1	199

Q1	228	1	134999
Q2	268	1	166999
Q3	311	1	295999
Max	366	32	12170174

Rata-rata Recency pada cluster ini adalah sebesar 270 hari, dengan median sebesar 268 hari. Hal ini menunjukkan bahwa sebagian besar pelanggan dalam cluster ini sudah cukup lama tidak melakukan pembelian, dengan rentang waktu terakhir transaksi antara 186 hingga 366 hari. Berdasarkan sebaran tersebut, pelanggan dalam cluster ini tergolong tidak aktif, atau bahkan mendekati kategori hilang (*churn*). Dari aspek Frequency, nilai rata-rata adalah 1,19 kali transaksi dengan median 1, yang berarti sebagian besar pelanggan hanya melakukan satu kali pembelian. Meskipun demikian, terdapat outlier dengan frekuensi pembelian hingga 32 kali, yang menandakan adanya segelintir pelanggan dengan loyalitas tinggi. Dari sisi Monetary, rata-rata total pembelian adalah sebesar Rp245.914, dengan nilai maksimum mencapai Rp12.170.000. Kuartil atas (Q3) sebesar Rp295.999 menunjukkan bahwa sebagian besar pelanggan dalam cluster ini memiliki total pembelian yang masih dalam kategori menengah ke bawah. Untuk melihat sebaran pelanggan pada cluster 0, bisa kita lihat pada gambar dibawah ini:



Gambar 4. Scatter Plot Cluster 0

Scatter plot diatas antara Frequency dan Monetary menunjukkan bahwa titik-titik pelanggan sebagian besar terkonsentrasi pada nilai frekuensi 1–3 kali dengan nilai belanja di bawah Rp500.000. sebagian besar pelanggan hanya membeli satu kali dalam jumlah kecil, ada sejumlah pelanggan yang memiliki kontribusi besar terhadap total pendapatan. Oleh karena itu, pelanggan dalam Cluster 0 dapat digolongkan sebagai pelanggan pasif dengan kontribusi nilai yang beragam, dan sebagian dari mereka merupakan big spender tidak loyal.

D. Kesimpulan

Mayoritas pelanggan PT Maha Nagari Nusantara di platform Shopee menunjukkan perilaku transaksional. Hal ini ditandai dengan frekuensi pembelian rata-rata yang sangat rendah sebanyak 1,21 kali per tahun dan nilai median Recency yang tinggi dengan nilai 197 hari, mengindikasikan bahwa lebih dari separuh pelanggan cenderung tidak aktif setelah melakukan satu kali pembelian. Pola ini menyoroti tantangan fundamental dalam hal retensi dan pembangunan loyalitas pelanggan.

Integrasi algoritma K-Means Clustering terbukti mampu mengoptimalkan analisis RFM dengan menghasilkan segmentasi yang lebih objektif dibandingkan metode skoring tradisional yang subjektif. Berdasarkan evaluasi statistik menggunakan Silhouette Coefficient, model paling optimal membagi basis pelanggan menjadi dua Cluster ($k=2$) yang paling homogen di dalam dan berbeda di luar.

Penerapan K-Means berhasil mengidentifikasi dua persona pelanggan yang sangat berbeda secara perilaku, yakni *Cluster 1* dengan label "Pelanggan Aktif dan Berpotensi" sebanyak 40.404 pelanggan. Kelompok ini memiliki *Recency* yang rendah dengan rata-rata 100 hari, menandakan mereka baru saja bertransaksi, serta memiliki nilai *Frequency* dan *Monetary* sedikit di atas rata-rata. Mereka adalah kelompok inti yang menopang bisnis saat ini. Kemudian, *Cluster 0* dengan label "Pelanggan Pasif dan Berisiko" dengan jumlah pelanggan 46.201. Ini adalah kelompok mayoritas yang ditandai dengan *Recency* sangat tinggi dengan rata-rata 270 hari, menunjukkan bahwa mereka adalah

pelanggan yang sudah lama tidak aktif, kemungkinan besar pembeli satu kali, dan berada pada risiko tinggi untuk hilang secara permanen (*churn*).

Perusahaan perlu mengadopsi pendekatan strategi ganda yang terdiferensiasi. Satu set strategi harus berfokus pada pembinaan dan retensi untuk "Pelanggan Aktif dan Berpotensi", dan satu set strategi lain harus berfokus pada reaktivasi untuk "Pelanggan Pasif dan Berisiko".

Mengingat "Pelanggan Pasif dan Berisiko" adalah segmen terbesar, upaya reaktivasi menjadi sangat krusial. Disarankan untuk merancang kampanye "*win-back*" yang agresif. Untuk segmen "Pelanggan Aktif dan Berpotensi", fokus utama adalah mengubah mereka menjadi pelanggan setia. Disarankan untuk mengimplementasikan program pembinaan melalui email setelah pembelian dan memperkenalkan program loyalitas berbasis poin untuk mendorong pembelian berulang.

Perilaku pelanggan bersifat dinamis. Disarankan agar analisis segmentasi ini dilakukan secara rutin misalnya, per kuartal atau semester untuk memantau pergerakan pelanggan antar segmen dan menyesuaikan strategi pemasaran sesuai dengan perubahan yang terjadi.

Ucapan Terimakasih

Segala puji dan syukur kehadirat Allah SWT atas limpahan rahmat dan karunia-Nya. Shalawat serta salam senantiasa tercurah kepada junjungan kita Nabi Muhammad SAW, beserta keluarga, sahabat, dan pengikutnya hingga akhir zaman. Penulis mengucapkan terima kasih kepada Kedua Orang Tua yang selalu mendukung penulis dalam hal apapun serta kepada M Wildan Habibi, Servina Nabila, dan Farras Anaqah Putri selaku mentor yang selalu membantu dan memberikan semangat kepada penulis selama proses kegiatan magang dan penelitian.

Daftar Pustaka

- Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46(1), 243–256. <https://doi.org/10.1016/j.patcog.2012.07.021>
- Buttle, F., & Makklan, S. (2015). *Customer Relationship Management*.
- Diyandini, S. M., & Kurniati, E. (2024). Analisis Maksimalisasi Utilitas Konsumen Melalui Grafik dan Persamaan Rumus. *DataMath: Journal of Statistics and Mathematics*, 2(1), 1–10.
- France, S. L., & Ghose, S. (2018). *Marketing Analytics: Methods, Practice, Implementation, and Links to Other Fields*. <https://doi.org/10.1016/j.eswa.2018.11.002>
- Guerola-Navarro, V., Oltra-Badenes, R., Gil-Gomez, H., & Gil-Gomez, J. A. (2021). Research model for measuring the impact of customer relationship management (CRM) on performance indicators. *Economic Research-Ekonomika Istrazivanja*, 34(1), 2669–2691. <https://doi.org/10.1080/1331677X.2020.1836992>
- Iacobucci, D., Petrescu, M., Krishen, A., & Bendixen, M. (2019). The state of marketing analytics in research and practice. *Journal of Marketing Analytics*, 7(3), 152–181. <https://doi.org/10.1057/s41270-019-00059-2>
- Khoeriyah, R. Y., & Hajarisman, N. (2021). Regresi Terboboti Geografis Semiparametrik (RTG-S) untuk Pemodelan Indeks Pembangunan Kesehatan Masyarakat Kabupaten/Kota di Sumatera Utara. *Jurnal Riset Statistika*, 1(1), 43–50. <https://doi.org/10.29313/jrs.v1i1.145>

- Novian, W., & Herlina, M. (2022). Faktor-faktor yang Mempengaruhi Perilaku Masyarakat terhadap Penggunaan Internet pada Desa Digital Sukaraja. *Jurnal Riset Statistika*, 161–168. <https://doi.org/10.29313/jrs.v2i2.1466>
- Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1), 103–119. <https://doi.org/10.1243/095440605X8298>
- Pramudhiaz, E. I. (2024, November 20). Nilai Transaksi E-commerce Indonesia Tahun Ini Diproyeksikan Tembus Rp 487 Triliun. *Tribunnews.Com*. <https://www.tribunnews.com/bisnis/2024/11/20/nilai-transaksi-e-commerce-indonesia-tahun-ini-diproyeksikan-tembus-rp-487-triliun>
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. In *Journal of Computational and Applied Mathematics* (Vol. 20).
- Rumiarti, C. D., & Budi, I. (2017). *SEGMENTASI PELANGGAN PADA CUSTOMER RELATIONSHIP MANAGEMENT DI PERUSAHAAN RITEL: STUDI KASUS PT GRAMEDIA ASRI MEDIA*.
- Sabuncu, İ., TÜRKAN, E., & POLAT, H. (2020). CUSTOMER SEGMENTATION AND PROFILING WITH RFM ANALYSIS. *Turkish Journal of Marketing*, 5(1), 22–36. <https://doi.org/10.30685/tujom.v5i1.84>
- Sheth, J. (2021). New areas of research in marketing strategy, consumer behavior, and marketing analytics: the future is bright. *Journal of Marketing Theory and Practice*, 29(1), 3–12. <https://doi.org/10.1080/10696679.2020.1860679>
- Sholeh, M., Andayati, D., Yuliana Rachmawati, R., Studi Informatika, P., & Teknologi Informasi dan Bisnis, F. (2022). *DATA MINING MODEL KLASIFIKASI MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DENGAN NORMALISASI UNTUK PREDIKSI PENYAKIT DIABETES DATA MINING MODEL CLASSIFICATION USING ALGORITHM K-NEAREST NEIGHBOR WITH NORMALIZATION FOR DIABETES PREDICTION*.
- Virgiawan A. (2015). *CUSTOMER SEGMENTATION USING RFM ANALYSIS AND FUZZY-C-MEANS ALGORITHM TO HELP CUSTOMER RELATIONSHIP MANAGEMENT AT PT. XYZ*.