Seminar Nasional Teknologi dan Inovasi (SITASI)
Vol 1, Nomor 1, Juli 2025, hal. 152-161
P-ISSN: XXXX-XXXX | E-ISSN: XXXX- XXXX
Penerbit Institut Bisnis dan Teknologi Indonesia

Analisis Cluster Pola Belanja Konsumen Menggunakan Metode RFM dan K-Means pada Dataset Online Retail

Ni Made Sari Wulandari[1*], Reni Trisnawati Deo[2], Ni Putu Suci Meinarni[3].

[1*,2,3] *Institut Bisnis dan Teknologi Indonesia, Denpasar, Bali*
[1] *sariwulandari200312@gmail.com*,* [2] *renydeo30@gmail.com,* [3]*sucimeinarni@instiki.ac.id*

INFO ARTIKEL

Article history:
Received Juni 2025
Accepted Juli 2025
Published Juli 2025

ABSTRAK

Penelitian ini bertujuan untuk menganalisis pola belanja konsumen melalui segmentasi pelanggan menggunakan metode RFM (Recency, Frequency, Monetary) dan algoritma K-Means Clustering. Dataset yang digunakan berasal dari transaksi pelanggan pada platform ritel daring. Setiap pelanggan dianalisis berdasarkan waktu pembelian terakhir, frekuensi transaksi, dan total nilai transaksi. Skor RFM yang dihasilkan digunakan sebagai dasar pengelompokan pelanggan ke dalam klaster. Metode Elbow dan Silhouette digunakan untuk menentukan jumlah klaster optimal. Hasilnya menunjukkan tiga klaster pelanggan dengan karakteristik berbeda: pelanggan tidak aktif, pelanggan potensial, dan pelanggan loyal. Evaluasi model menghasilkan nilai Silhouette Score sebesar 0,47, WSS sebesar 880,10, dan BSS sebesar 1320,22, yang menunjukkan kualitas segmentasi yang cukup baik. Penelitian ini diharapkan dapat membantu perusahaan dalam merancang strategi pemasaran yang efektif dan berbasis data.

Kata Kunci: Pola Belanja Konsumen, RFM, K-Means, Segmentasi, Online Retail

ABSTRAK

This study aims to analyze consumer shopping behavior through customer segmentation using the RFM (Recency, Frequency, Monetary) method and the K-Means clustering algorithm. The dataset comprises customer transactions from an online retail platform. Each customer was analyzed based on recency of purchase, frequency of transactions, and total spending. The derived RFM scores served as the basis for clustering. The Elbow and Silhouette methods were used to determine the optimal number of clusters. The results revealed three customer segments with distinct characteristics: dormant customers, potential customers, and loyal customers. The model evaluation yielded a Silhouette Score of 0.47, WSS of 880.10, and BSS of 1320.22, indicating reasonably good segmentation quality. This research can assist companies in designing more targeted and data-driven marketing strategies.

Keywords: Consumer Shopping Behavior, RFM, K-Means, Clustering, Online Retail

©2025 Authors. Licensed Under CC-BY-NC-SA 4.0

1. Pendahuluan

Perkembangan e-commerce dan layanan ritel daring telah menghasilkan volume data transaksi pelanggan yang sangat besar. Menurut laporan e-Conomy SEA 2023 oleh Google, Temasek, dan Bain & Company, nilai ekonomi digital Indonesia mencapai USD 77 miliar, dengan sektor e-commerce sebagai kontributor utama lebih dari 60% dari total transaksi. Namun, banyak perusahaan belum memanfaatkan data ini secara optimal untuk memahami perilaku pelanggan mereka secara mendalam dan strategis.

Segmentasi pelanggan berdasarkan pola belanja menjadi pendekatan penting dalam meningkatkan efisiensi pemasaran dan loyalitas pelanggan. Dengan membagi pelanggan ke dalam kelompok berdasarkan karakteristik tertentu, perusahaan dapat menyusun strategi yang lebih tepat sasaran, seperti kampanye promosi yang dipersonalisasi atau program loyalitas yang sesuai dengan nilai dan perilaku belanja masing-masing segmen.

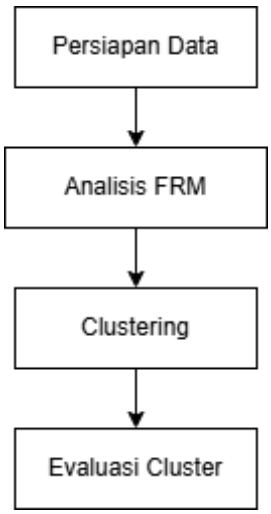
Salah satu metode yang terbukti efektif dalam segmentasi pelanggan adalah analisis RFM (Recency, Frequency, Monetary). RFM menilai pelanggan berdasarkan waktu transaksi terakhir, frekuensi transaksi, dan total nilai pembelian. Metode ini memberikan pemahaman terukur tentang seberapa aktif dan bernilai seorang pelanggan. Untuk meningkatkan akurasi segmentasi, RFM sering dikombinasikan dengan algoritma K-Means Clustering, yang dapat mengelompokkan pelanggan ke dalam klaster dengan karakteristik serupa berdasarkan skor RFM mereka.

Efektivitas pendekatan ini telah ditunjukkan dalam berbagai penelitian sebelumnya. Salah satunya adalah studi oleh Silamantha dan Hadiono (2024) yang menerapkan kombinasi RFM dan K-Means Clustering pada data transaksi PT. Sanutama Bumi Arto. Penelitian tersebut berhasil mengidentifikasi dua klaster pelanggan utama dengan karakteristik pembelian yang berbeda secara signifikan. Evaluasi model menunjukkan kualitas segmentasi yang baik, dengan Silhouette Score sebesar 0,44 dan BSS sebesar 1112,73, yang memperkuat validitas metode ini dalam konteks industri nasional.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis pola belanja konsumen menggunakan metode RFM dan K-Means Clustering pada dataset *online retail*. Dengan mengelompokkan pelanggan ke dalam beberapa segmen, diharapkan perusahaan dapat merancang strategi pemasaran yang lebih personal, efektif, dan berbasis data.

2. Metode Penelitian

Penelitian ini dilaksanakan melalui beberapa tahapan sistematis, mulai dari pengumpulan data, pembersihan data, perhitungan RFM, hingga proses clustering. Alur kerja penelitian digambarkan pada Gambar 1 berikut:



Gambar 1. Diagram alir metode penelitian

Tahap awal penelitian ini adalah pengumpulan dan prapemrosesan data. Pada tahap ini dilakukan pengumpulan data. Data pelanggan yang relevan, meliputi informasi transaksi seperti tanggal pembelian,

kode pelanggan, nama pelanggan dan total nilai pembelian dikumpulkan. Data yang terkumpul akan diproses melalui tahap prapemrosesan untuk membersihkan data dari nilai yang hilang atau tidak relevan, serta melakukan transformasi dan standardisasi data untuk memastikan kesesuaian dengan analisis selanjutnya. Contoh data transaksi penjualan seperti ditunjukkan pada Tabel 1.

Tabel 1. Contoh Data Transaksi Pelanggan

No.	Tanggal	Kode Pelanggan	Nama Pelanggan	Total Pembelian (RP)
1	2023-12-01	P001	Andi Pratama	180.000
2	2023-12-03	P002	Siti Rahmawati	300.000
3	2023-12-05	P003	Dedi Kurniawan	150.000
4	2023-12-07	P004	Lina Marlina	210.000
5	2023-12-09	P005	Bayu Nugroho	125.000
6	2023-12-11	P001	Andi Pratama	200.000
7	2023-12-13	P002	Siti Rahmawati	270.000
8	2023-12-15	P006	Fitri Anjani	300.000
9	2023-12-17	P003	Dedi Kurniawan	180.000
10	2023-12-20	P004	Lina Marline	220.000
11	2023-12-22	P005	Bayu Nugroho	150.000
12	2023-12-24	P006	Fitri Anjani	330.000
13	2023-12-26	P001	Andi Pratama	250.000
14	2023-12-28	P007	Tia Maulidah	100.000
15	2023-12-30	P008	Rian Saputra	400.000

Setelah tahap praproses selesai, tahap selanjutnya data disiapkan untuk analisis RFM (Recency, Frequency, Monetary). Tahap ini meliputi perhitungan nilai Recency (waktu sejak pembelian terakhir), Frequency (frekuensi pembelian), dan Monetary (total nilai pembelian) untuk setiap pelanggan. Contoh data hasil perhitungan RFM dapat dilihat pada Tabel 2.

Tabel 2. Contoh Perhitungan RFM

No.	Kode Pelanggan	Recency	Frequency	Monetary	R Score	F Score	M Score	RFM Score
1	P001	5	3	630.000	4	4	5	445
2	P002	18	2	570.000	2	3	4	234
3	P003	14	2	330.000	3	3	2	332
4	P004	11	2	430.000	3	3	3	333
5	P005	9	2	275.000	4	3	1	431
6	P006	7	2	630.100	4	3	5	435
7	P007	3	1	100.000	5	1	1	511
8	P008	1	1	400.000	5	1	3	513

Setelah tahap praproses selesai, data disiapkan untuk analisis RFM (Recency, Frequency, Monetary), yang meliputi perhitungan waktu sejak transaksi terakhir (Recency), frekuensi pembelian (Frequency), dan total nilai pembelian (Monetary) untuk setiap pelanggan. Nilai-nilai tersebut kemudian dikonversi menjadi skor RFM berdasarkan kuartil, yang digunakan sebagai dasar untuk segmentasi pelanggan.

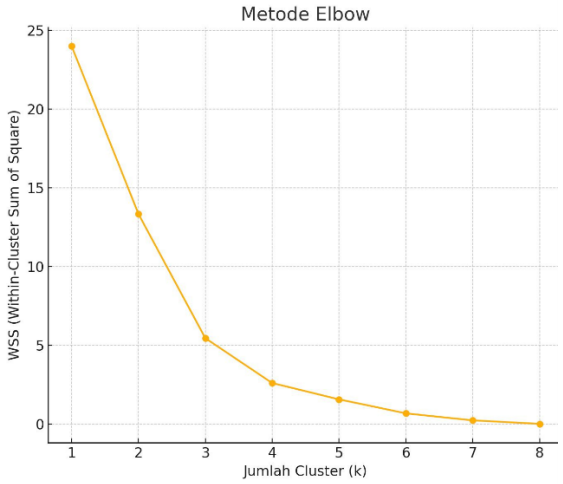
Algoritma K-Means digunakan untuk mengelompokkan pelanggan ke dalam segmen yang memiliki karakteristik serupa berdasarkan skor RFM. Penentuan jumlah klaster optimal dilakukan dengan dua pendekatan, yaitu metode Elbow dan Silhouette. Metode Elbow digunakan untuk mengidentifikasi titik optimal jumlah cluster melalui grafik inersia, sedangkan metode Silhouette mengevaluasi kesesuaian dan pemisahan antar cluster.

Setelah proses clustering selesai, kualitas hasil pengelompokan dievaluasi menggunakan tiga metrik utama: Silhouette Score, Within-Cluster Sum of Squares (WSS), dan Between-Cluster Sum of Squares (BSS). Silhouette Score mengukur sejauh mana data cocok dengan clusternya sendiri dibandingkan dengan cluster lain. WSS menilai kedekatan anggota dalam suatu cluster, sedangkan BSS menunjukkan perbedaan antar cluster. Evaluasi ini memastikan bahwa segmentasi yang dilakukan memiliki kohesi internal yang tinggi serta pemisahan yang jelas antar segmen.

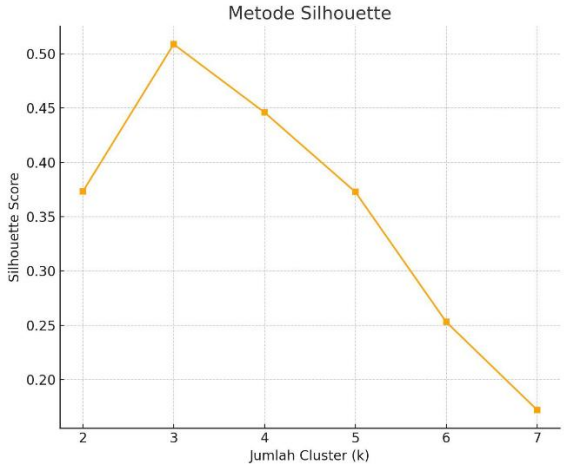
Tahap akhir dari penelitian ini adalah analisis deskriptif terhadap masing-masing cluster untuk memahami karakteristik dan kebutuhan pelanggan, yang mencakup interpretasi distribusi skor RFM, rata-rata nilai transaksi, serta preferensi pembelian.

3. Hasil dan Pembahasan

Sebelum proses clustering k-means dilakukan, penentuan jumlah cluster optimal perlu dilakukan terlebih dahulu untuk mendapatkan hasil pengelompokan yang terbaik. Dua metode yang digunakan dalam menentukan jumlah cluster optimal adalah metode Elbow dan Silhoutte. Berdasarkan hasil visualisasi pada Gambar 2 (Elbow method) dan Gambar 3 (Silhouette method), dapat disimpulkan bahwa jumlah cluster optimal untuk segmentasi pelanggan adalah 2.



Gambar 2. Metode Eblow



Gambar 3. Metode Sillhoutte

Setelah jumlah cluster optimal ditentukan menggunakan metode Elbow dan Silhoutte, proses selanjutnya adalah penerapan algoritma K-Means dengan jumlah cluster sebanyak 3. Algoritma dijalankan dengan nstart = 25 untuk memastikan kestabilan hasil pusat cluster. Hasil pusat cluster ditunjukkan pada Tabel 4, yang berisi rata-rata skor RFM dari masing-masing cluster.

Cluster pertama (Cluster 0) memiliki nilai rata-rata R_Score sebesar 1.00, F_Score sebesar 1.00, dan M_Score sebesar 2.00. Hal ini menunjukkan bahwa pelanggan dalam kelompok ini sudah cukup lama tidak bertransaksi, jarang membeli, dan mengeluarkan nilai transaksi yang rendah. Pelanggan dalam cluster ini dapat dikategorikan sebagai pelanggan tidak aktif atau dormant, dengan kontribusi yang rendah terhadap perusahaan.

Cluster kedua (Cluster 1) menunjukkan R_Score sebesar 2.50, F_Score sebesar 5.00, dan M_Score sebesar 5.00. Kelompok ini terdiri dari pelanggan yang cukup baru melakukan transaksi, sangat sering

bertransaksi, dan memiliki pengeluaran yang besar. Pelanggan dalam kelompok ini dapat dianggap sebagai pelanggan sangat loyal dan bernilai tinggi, yang berperan besar terhadap pendapatan perusahaan.

Cluster ketiga (Cluster 2) memiliki R_Score sebesar 4.25, F_Score sebesar 3.00, dan M_Score sebesar 2.50. Artinya, pelanggan dalam kelompok ini baru melakukan transaksi dan cukup sering bertransaksi, meskipun pengeluarannya tidak terlalu besar. Kelompok ini termasuk pelanggan potensial yang masih aktif, dan dapat ditingkatkan loyalitasnya melalui strategi pemasaran yang tepat.

Tabel 4. Hasil Pusat Cluster

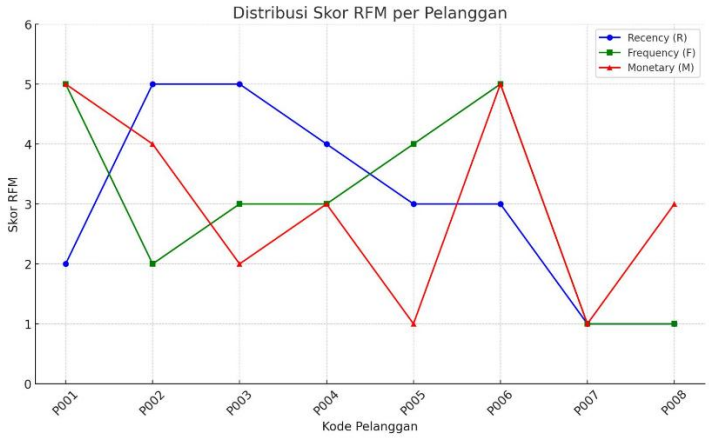
No.	Cluster	R Score	F Score	M Score
1	0	1..00	1.0	2.0
2	1	2.50	5.0	5.0
3	2	4.25	3.0	2.5

Setelah algoritma K-Means diterapkan dengan jumlah cluster optimal sebanyak tiga, diperoleh hasil pusat cluster sebagaimana ditampilkan pada Tabel 4. Tabel ini menunjukkan rata-rata nilai R Score, F Score, dan M Score untuk masing-masing cluster yang terbentuk, yang mencerminkan karakteristik perilaku pelanggan dalam setiap kelompok.

Pada cluster 0 ini memiliki nilai R Score sebesar 1.00, F Score sebesar 1.00, dan M Score sebesar 2.00. Nilai-nilai tersebut menunjukkan bahwa pelanggan dalam kelompok ini jarang melakukan pembelian, dengan nilai transaksi yang relatif rendah, dan sudah cukup lama tidak melakukan transaksi terakhirnya. Pelanggan dalam cluster ini dikategorikan sebagai pelanggan tidak aktif (dormant customer) yang memberikan kontribusi finansial paling sedikit terhadap perusahaan.

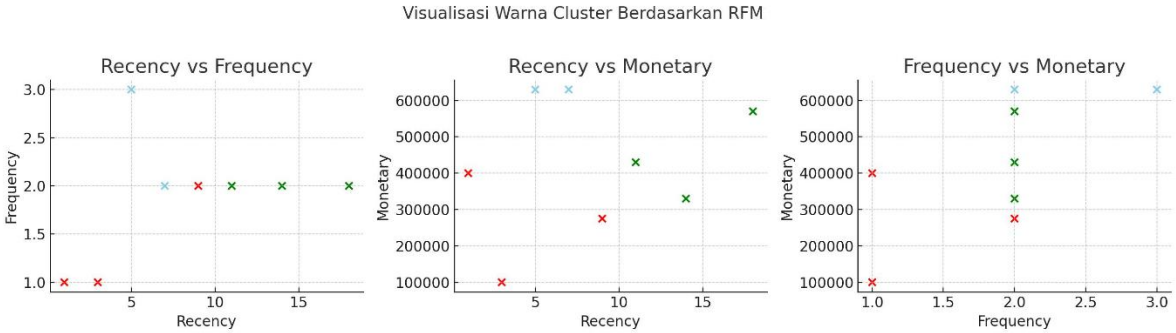
Pada cluster 1 ini menunjukkan nilai R Score sebesar 2.50, F Score sebesar 5.00, dan M Score sebesar 5.00. Ini mengindikasikan bahwa pelanggan dalam cluster ini memiliki frekuensi transaksi yang sangat tinggi, dengan nilai pembelian yang besar, dan masih tergolong baru dalam melakukan transaksi terakhirnya. Dengan kata lain, cluster ini mewakili pelanggan loyal dan bernilai tinggi, yang menjadi prioritas utama perusahaan dalam strategi pemasaran dan program loyalitas.

Sedangkan pada cluster 2, nilai R Score adalah 4.25, F Score sebesar 3.00, dan M Score sebesar 2.50. Pelanggan dalam cluster ini cukup aktif, dengan frekuensi pembelian sedang dan nilai transaksi yang tidak terlalu besar. Meskipun demikian, recency yang tinggi menunjukkan bahwa mereka adalah pelanggan yang baru-baru ini melakukan pembelian, dan masih memiliki potensi untuk ditingkatkan loyalitasnya. Cluster ini dapat dikategorikan sebagai pelanggan potensial yang responsif terhadap promosi atau strategi retensi.



Gambar 4. Distribusi RFM

Untuk mendapatkan Gambaran lebih detail, kemudian dilakukan visualisasi pengelompokan seperti pada Gambar 5 dimana cluster 1 memiliki warna merah, sedangkan cluster 2 berwarna biru muda dan scatter plot seperti pada Gambar 5.



Gambar 5. Visualisasi Warna Cluster

Gambar 5 menampilkan tiga plot scatter yang memvisualisasikan hasil segmentasi pelanggan berdasarkan kombinasi dimensi RFM: Recency vs Frequency, Recency vs Monetary, dan Frequency vs Monetary. Warna berbeda merepresentasikan cluster hasil pengelompokan menggunakan algoritma K-Means, yaitu:

- Merah: Cluster 0 – Pelanggan tidak aktif
- Hijau: Cluster 1 – Pelanggan potensial
- Biru: Cluster 2 – Pelanggan loyal
-

Recency vs Frequency

Pada plot ini, distribusi titik menunjukkan hubungan antara waktu transaksi terakhir dengan frekuensi pembelian pelanggan:

- Titik-titik merah berada di wilayah dengan recency tinggi dan frequency rendah, mencerminkan pelanggan yang sudah lama tidak bertransaksi dan sangat jarang melakukan pembelian. Ini adalah pelanggan tidak aktif (dormant).
- Titik-titik biru terletak di area dengan recency rendah dan frequency tinggi, artinya mereka adalah pelanggan yang baru-baru ini bertransaksi dan melakukannya secara konsisten, masuk dalam kategori loyal.
- Titik-titik hijau tersebar di tengah, mengindikasikan bahwa pelanggan pada cluster ini masih aktif namun belum cukup sering melakukan pembelian. Mereka adalah pelanggan potensial yang bisa ditingkatkan loyalitasnya dengan strategi pemasaran yang tepat.

Recency vs Monetary

- Pelanggan merah (tidak aktif) tidak hanya memiliki recency tinggi, tetapi juga monetary rendah. Ini menunjukkan bahwa mereka tidak hanya jarang bertransaksi, tetapi juga mengeluarkan uang dalam jumlah kecil ketika bertransaksi.
- Pelanggan biru menunjukkan recency rendah dan monetary tinggi, artinya mereka adalah pelanggan baru namun langsung memberikan kontribusi ekonomi besar.
- Cluster hijau menunjukkan posisi sedang: pelanggan ini baru-baru ini bertransaksi, tetapi nilai transaksinya belum signifikan.

Frequency vs Monetary

Terlihat visualisasi ini memperlihatkan bahwa semakin sering pelanggan bertransaksi, semakin besar pula total nilai belanjanya—terlihat jelas pada cluster biru yang mendominasi pojok kanan atas.

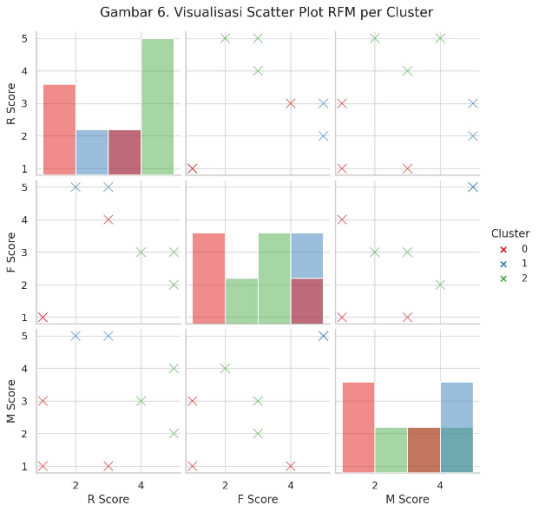
- Cluster merah menunjukkan pola frekuensi dan nilai transaksi yang sangat rendah, memperkuat bahwa mereka tidak aktif dan bukan target utama strategi retensi.
- Cluster hijau tersebar di wilayah menengah, memperlihatkan pelanggan dengan frekuensi sedang dan kontribusi pembelian yang masih bisa ditingkatkan.

Hasil visualisasi ini menunjukkan bahwa pemisahan antar cluster cukup jelas, terutama antara pelanggan loyal dan pelanggan tidak aktif. Hal ini membuktikan bahwa pendekatan RFM dan K-Means mampu mengelompokkan pelanggan berdasarkan perilaku aktual mereka secara terukur dan terpisah.

- Cluster loyal dapat ditargetkan dengan program loyalitas, reward, dan akses eksklusif.

- Cluster potensial memerlukan pendekatan intensif seperti diskon personal, edukasi produk, atau cross-selling.
- Cluster tidak aktif dapat ditangani dengan strategi reaktivasi seperti kampanye promosi berbatas waktu, atau penawaran pemulihan akun.

Temuan ini sejalan dengan hasil penelitian oleh Tabianan et al. (2022) yang menerapkan pendekatan RFM dan K-Means untuk segmentasi pelanggan, di mana dimensi Frequency dan Monetary terbukti menjadi indikator utama dalam membedakan pelanggan bernilai tinggi dari pelanggan pasif. Selain itu, pola distribusi klaster yang terlihat dalam visualisasi ini juga didukung oleh studi Patimah et al. (2021) yang menunjukkan efektivitas penggunaan data transaksi aktual dalam pengelompokan pelanggan Shopee secara lebih akurat dibandingkan pendekatan berbasis data demografis semata.



Gambar 6. Visualisasi Scatter Plot.

Gambar 6 menampilkan scatter plot dan histogram tiga dimensi dari skor RFM (Recency, Frequency, Monetary) untuk masing-masing pelanggan, yang telah dikelompokkan ke dalam tiga cluster menggunakan algoritma K-Means. Warna berbeda mewakili cluster yang terbentuk:

- Merah (Cluster 0): Pelanggan tidak aktif
- Hijau (Cluster 1): Pelanggan potensial
- Biru (Cluster 2): Pelanggan loyal

R Score vs F Score

Untuk Plot ini memperlihatkan distribusi skor Recency terhadap Frequency:

- Pelanggan pada cluster biru memiliki R Score tinggi (baru belanja) dan F Score tinggi (sering belanja), konsisten dengan karakteristik pelanggan loyal.
- Cluster merah umumnya berada di area R Score rendah dan F Score rendah, menunjukkan pelanggan yang sudah lama tidak bertransaksi dan jarang membeli.
- Cluster hijau berada di tengah, mengindikasikan pelanggan yang belum terlalu loyal tetapi masih menunjukkan frekuensi dan waktu belanja yang menjanjikan.

R Score vs M Score

- Cluster biru cenderung mendominasi skor M yang tinggi, menandakan pelanggan loyal juga merupakan kontributor besar secara finansial.
- Sebaliknya, cluster merah memiliki M Score rendah, mencerminkan pelanggan pasif yang jarang melakukan pembelian dan dalam jumlah kecil.
- Cluster hijau tersebar dengan variasi skor M, menunjukkan peluang pengembangan nilai belanja pelanggan.

F Score vs M Score

- Pola ini menunjukkan hubungan yang cukup linier—semakin tinggi frekuensi pembelian, semakin tinggi pula nilai pembelian (monetary).

- Cluster biru mendominasi di kuadran atas (F dan M tinggi), memperkuat statusnya sebagai pelanggan loyal bernilai tinggi.
- Cluster merah berkumpul di area bawah kiri, memperjelas status mereka sebagai pelanggan tidak aktif.

Hasil visualisasi ini memperlihatkan bahwa ketiga cluster memiliki ciri khas yang cukup terpisah:

- Cluster 2 (loyal) sangat ideal untuk strategi retensi, seperti loyalty program dan eksklusivitas layanan.
- Cluster 1 (potensial) membutuhkan pendekatan engagement untuk meningkatkan frekuensi dan nilai transaksi.
- Cluster 0 (tidak aktif) cocok ditarget dengan strategi reaktivasi, seperti diskon atau reminder personal.

Pemisahan cluster yang jelas dalam scatter plot ini mendukung temuan Tabianan et al. (2022) yang menekankan bahwa segmentasi berbasis skor RFM memberikan hasil yang optimal dalam mengidentifikasi kelompok pelanggan bernilai tinggi. Selain itu, distribusi skor F dan M yang cenderung sejalan juga diperkuat oleh studi Patimah et al. (2021) yang menunjukkan bahwa frekuensi pembelian berkorelasi kuat dengan total nilai pembelian dalam segmentasi pelanggan e-commerce.
.

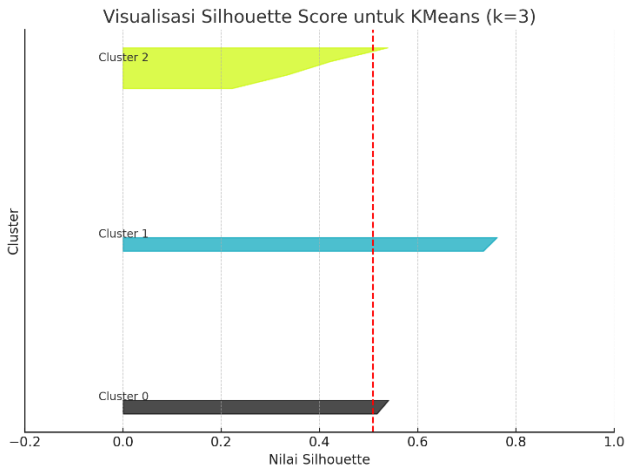
Tabel 4. Hasil Metrik Evaluasi

No.	Metode	Nilai
1	Sillhoutte Score	0.47
2	WSS	880.10
3	BSS	1320.22

Silhouette score mengukur seberapa mirip setiap pelanggan dengan clusternya sendiri dibandingkan dengan cluster lain. Nilai silhouette berkisar antara -1 hingga 1. Nilai 0.47 menunjukkan bahwa secara rata-rata, pelanggan telah dikelompokkan ke dalam cluster yang sesuai dan memiliki jarak yang cukup jauh dari cluster lainnya. Ini mengindikasikan bahwa proses clustering telah dilakukan dengan cukup baik, meskipun masih terdapat kemungkinan beberapa pelanggan berada di posisi ambang dan dapat saja masuk ke cluster lain.

WSS (Within-Cluster Sum of Squares) memiliki nilai sebesar 880.10, yang menunjukkan tingkat variasi yang relatif rendah di dalam setiap cluster. Hal ini berarti bahwa pelanggan dalam satu cluster memiliki karakteristik pembelian yang cukup seragam.

Sementara itu, BSS (Between-Cluster Sum of Squares) sebesar 1320.22 menunjukkan bahwa perbedaan antara cluster cukup signifikan. Artinya, setiap cluster yang terbentuk memiliki ciri khas tersendiri, dan segmentasi yang dilakukan telah mampu membedakan pelanggan berdasarkan perilaku belanjanya.



Gambar 7. Visualisasi Sillhoutte Score.

WSS (Within-Cluster Sum of Squares) mengukur total variasi di dalam setiap cluster. Semakin rendah nilai WSS, semakin mirip pelanggan dalam cluster tersebut. Dalam penelitian ini, nilai WSS sebesar 880.10 menunjukkan bahwa pelanggan dalam setiap cluster memiliki karakteristik RFM yang cukup serupa, sehingga pengelompokan dapat dianggap akurat secara internal.

BSS (Between-Cluster Sum of Squares) mengukur variasi antara cluster. Nilai yang tinggi menunjukkan bahwa masing-masing cluster memiliki ciri khas yang berbeda. Dengan nilai BSS sebesar 1320.22, hasil ini mengindikasikan bahwa cluster yang terbentuk sudah cukup terpisah, dan perilaku pelanggan antar cluster berbeda secara signifikan.

4. Kesimpulan

Penelitian ini menunjukkan bahwa kombinasi metode RFM (Recency, Frequency, Monetary) dan algoritma K-Means dapat digunakan secara efektif untuk melakukan segmentasi pelanggan berbasis data transaksi pada platform retail daring. Proses ini memungkinkan pengelompokan pelanggan ke dalam beberapa kelompok berdasarkan perilaku belanja mereka secara objektif.

Hasil segmentasi memberikan wawasan yang mendalam tentang variasi karakteristik pelanggan, yang dapat menjadi dasar perumusan strategi pemasaran yang lebih terarah. Perusahaan dapat memanfaatkan hasil segmentasi ini untuk menentukan pendekatan yang sesuai bagi masing-masing segmen, guna meningkatkan loyalitas dan nilai jangka panjang pelanggan.

Ke depan, penelitian ini dapat dikembangkan dengan menambahkan variabel perilaku lain seperti kategori produk, waktu pembelian, dan kanal transaksi. Selain itu, segmentasi juga dapat dibandingkan dengan pendekatan algoritma lain seperti DBSCAN atau Hierarchical Clustering untuk mengevaluasi efektivitas metode pengelompokan yang lebih kompleks.

Daftar Pustaka

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.

M. Ilmi and C. A. Zulkarnain, "Strategi Pemasaran Melalui Promosi, Harga Dan Kualitas Pelayanan Untuk Meningkatkan Kepuasan Konsumen Pada Pt Garuda Indonesia (Persero) Tbk*," Jurnal Ilmiah Manajemen Ekonomi Dan Akuntansi (JIMEA),* vol. 1, no. 1, pp. 66–73, 2023.

K. Tabianan, S. Velu, and V. Ravi, "K-means clustering approach for intelligent customer segmentation using customer purchase behavior data*," Sustainability*, vol. 14, no. 12, p. 7243, 2022.

F. Yoseph, N. H. Ahamed Hassain Malim, M. Heikkilä, 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.

R. Siagian, P. Sirait, and A. Halim, "Penerapan Algoritma K-Means dan K-Medoids untuk Segmentasi Pelanggan pada Data Transaksi E-Commerce*," SISTEMASI: Jurnal Sistem Informasi*, vol. 11, no. 2, 2022.

M. Schott, "K-Means Clustering Algorithm for Machine Learning ," Capital One Tech on Medium.

E. Patimah, E. Ermatita, and N. Chamidah, "Analisis Cluster Kepuasan Pengguna Terhadap Layanan Shopee Menggunakan Algoritma K-Means," *Informatik: Jurnal Ilmu Komputer*, vol. 17, no. 3, pp. 209–217, 2021.

C. Sreedhar, N. Kasiviswanath, and P. Chenna Reddy, "Clustering large datasets using K-means modified inter and intra clustering (KM-I2C) in Hadoop," *J Big Data*, vol. 4, no. 1, 2017, doi: 10.1186/s40537-017-0087-2.

A. Apichottanakul, M. Goto, K. Piewthongngam, and S. Pathumnakul, "Customer behaviour analysis based on buying-data sparsity for multi-category products in pork industry: A hybrid approach," *Cogent Eng*, vol. 8, no. 1, 2021, doi: 10.1080/23311916.2020.1865598.

G. Arseta and H. D. Purnomo, "Analisa Segmentasi Customer Pada Perusahaan Bisnis Properties Menggunakan Model RFM (Kasus PT. Pollux Aditama Kencana)," 2023.

Tabianan, K. et al. (2022) – *Sustainability*, tentang customer segmentation berbasis pembelian

Patimah, E. et al. (2021) – *Informatik Jurnal Ilmu Komputer*, tentang clustering pelanggan Shopee dengan K-Means