

Implementasi *K-Means* untuk Segmentasi Pelanggan *Smartlock* Berdasarkan Pola Pembelian

Joni Prianto¹, Joko Priambodo²

^{1,2,3}Universitas Pamulang; Jl Surya Kencana No.1, Pamulang, Pamulang Barat, Pamulang Banten

^{1,2,3}Jurusan Teknik Informatika, Fakultas Teknik, Universitas Pamulang

e-mail: joniprianto60@gmail.com, dosen00276@unpam.ac.id

Abstrak

Persaingan pasar yang ketat di industri teknologi keamanan menuntut PT Digital Sukses Nasional untuk memahami karakteristik dan pola pembelian pelanggan secara mendalam guna merumuskan strategi pemasaran yang efektif. Penelitian ini bertujuan untuk mengembangkan dan mengimplementasikan sistem berbasis web yang mengintegrasikan model analisis RFM (Recency, Frequency, Monetary) dengan algoritma *K-Means Clustering* untuk melakukan segmentasi pelanggan produk *Smartlock* secara otomatis. Data yang digunakan terdiri dari 330 data transaksi dari 164 pelanggan unik produk *Smartlock* selama periode satu tahun, yang diproses melalui model RFM untuk mendapatkan skor aktivitas pelanggan. Penentuan jumlah kluster optimal (*K*) dibantu oleh metode *Elbow Method*, yang menghasilkan tiga kluster (*K*=3) sebagai jumlah kluster yang paling efisien. Hasil implementasi menunjukkan bahwa sistem berhasil mengelompokkan pelanggan ke dalam tiga segmen yang berbeda, yaitu Kluster 3 (Pelanggan Setia & Bernilai Tinggi), Kluster 2 (Pelanggan Pasif/Hilang), dan Kluster 1 (Pelanggan Baru/Potensial). Sistem yang dibangun juga berhasil menyediakan rekomendasi strategi pemasaran yang spesifik untuk setiap segmen kluster yang terbentuk. Hasil segmentasi ini memberikan dasar yang terukur bagi PT Digital Sukses Nasional untuk menyesuaikan pendekatan pemasaran yang lebih personal dan tepat sasaran, sehingga dapat mengoptimalkan retensi pelanggan dan meningkatkan total penjualan.

Kata Kunci: Segmentasi Pelanggan, *K-Means*, Pola Pembelian, *Smartlock*, RFM.

I. PENDAHULUAN

PT Digital Sukses Nasional merupakan perusahaan yang bergerak di bidang teknologi keamanan, salah satunya adalah produk *smartlock*. Dalam menghadapi persaingan pasar yang semakin ketat, perusahaan perlu memahami perilaku dan pola pembelian pelanggan untuk meningkatkan strategi pemasaran. Salah satu pendekatan yang efektif untuk menganalisis pola tersebut adalah segmentasi pelanggan.

Segmentasi pelanggan bertujuan untuk mengelompokkan pelanggan berdasarkan karakteristik tertentu sehingga perusahaan dapat memberikan layanan yang lebih personal dan tepat sasaran. Algoritma *K-Means* merupakan metode *clustering* yang mampu mengelompokkan data ke dalam beberapa kluster berdasarkan kesamaan karakteristik. Oleh karena itu, penelitian ini bertujuan mengimplementasikan algoritma *K-Means* dalam

mengelompokkan pelanggan produk *smartlock* berdasarkan pola pembelian mereka.

Berbagai penelitian sebelumnya menunjukkan bahwa algoritma *K-Means* telah banyak diterapkan dalam segmentasi pelanggan untuk mengoptimalkan strategi pemasaran. Penelitian oleh Hafidz Muhammad et al. (2024) membagi pelanggan berdasarkan pendapatan tahunan dan skor pengeluaran menggunakan metode *K-Means*, menghasilkan segmentasi yang membantu Perusahaan memahami perilaku belanja pelanggan. Maryana et al. (2018) menerapkan algoritma *K-Means* dengan model RFM (Recency, Frequency, Monetary) untuk mengelompokkan pelanggan CV Sinar Surya ke dalam dua kluster utama, yakni Everyday Shopper dan Golden Customer. Junaidi dan Madyartha (2024) juga menggunakan algoritma *K-Means* untuk mengelompokkan pelanggan berdasarkan usia, pendapatan, dan model RFM di Lantikya Store

Jombang, membantu toko menentukan strategi pemasaran yang lebih tepat sasaran.

Selain itu, Hadiwibowo (2023) menganalisis segmentasi pengunjung mall menggunakan K-Means *clustering* untuk memahami pola kunjungan dan meningkatkan strategi promosi. Terakhir, penelitian Karina Auliasari dan Mariza Kertaningtyas (2019) memanfaatkan *K-Means* dengan Bahasa pemrograman R untuk mengklasifikasikan konsumen berdasarkan karakteristik jenis usaha dan Lokasi, menunjukkan dominasi konsumen berjenis usaha PT di Pulau Jawa. Penelitian ini menjadi dasar yang kuat untuk menerapkan algoritma K-Means dalam segmentasi pelanggan produk smartlock di PT Digital Sukses Nasional, guna memahami pola pembelian pelanggan dan merancang strategi pemasaran yang lebih efektif.

II. METODE PELAKSANAAN

Penelitian ini mengadopsi pendekatan kuantitatif dengan memanfaatkan metodologi Data Mining, yang secara spesifik berfokus pada teknik klusterisasi. Data Mining dipilih karena kemampuannya untuk mengekstraksi pola yang bermakna dan tersembunyi dari himpunan data transaksi pelanggan yang besar, yang mana pola ini kemudian digunakan untuk mengelompokkan pelanggan ke dalam segmen-segmen homogen. Algoritma utama yang diimplementasikan adalah *K-Means Clustering*, sebuah metode non-hierarchical *clustering* yang efisien dalam mengelompokkan data berdasarkan jarak terdekat ke pusat kluster (*centroid*). *K-Means* menjadi sangat relevan dalam kasus segmentasi pelanggan karena dapat bekerja secara efektif pada data numerik yang dihasilkan dari model analisis RFM (*Recency, Frequency, Monetary*). Kombinasi RFM dan *K-Means* ini memungkinkan segmentasi tidak hanya berdasarkan data demografi, tetapi juga berdasarkan nilai ekonomi dan perilaku pembelian riil pelanggan, menjadikannya dasar yang kuat untuk perumusan strategi pemasaran yang ditargetkan.

A. Analisis Pengumpulan Data

Sebanyak 330 data transaksi digunakan sebagai dasar analisis, dengan atribut utama meliputi tanggal pembelian terakhir, frekuensi pembelian, jumlah produk yang dibeli, dan total nilai pembelian. Informasi ini kemudian diolah untuk melakukan pengelompokan pelanggan menggunakan model RFM (*Recency, Frequency, Monetary*)

Tabel 2.1 Data Transaksi PT Digital sukses nasional

NO	TAN GGA L	NAMA CUSTOM ER	ITEM	QT Y	TOTAL HARGA
1	06- Jan-24	Suzana Huang	Philips Alpha- VP- 5HWS	1	13500000
2	08- Jan-24	Rini	Philips DDL603E	1	6500000
3	28- Jan-24	Robert Anastasia	Philips 5100- 5HBKS	1	6500000
4	27- Jan-24	Steviani	Philips Alpha- VP- 5HWS	1	13500000
5	06- Jan-24	Bapak Dilen	Philips Alpha- 5HB	2	17600000
....
330	30- Dec- 24	Bapak Taslim	Philips DDL702E	1	9292429

B. Analisis RFM

Pada tahap ini, pelanggan dari produk *Smartlock* yang terjual akan dikelompokkan berdasarkan nilai transaksi mereka menggunakan konsep analisis RFM (*Recency, Frequency, Monetary*). Analisis ini bertujuan untuk mengidentifikasi perilaku pembelian pelanggan berdasarkan seberapa baru, seberapa sering, dan seberapa besar nilai transaksi yang dilakukan.

Dari data transaksi pada table 3.1 diatas langkah pertama adalah menghitung nilai *recency*, *frequency*, dan *monetary* untuk setiap pelanggan:

Rumus-Rumus RFM

Recency (R)

$$R_i = (T_{\text{analisis}} - T_{\text{transaksi terakhir}})$$

Frequency (F)

F_i = Jumlah transaksi selama periode analisis

Monetary (M)

M_i = Total belanja (Rp)

Contoh perhitungan:

Pada tahap ini penulis melakukan perhitungan untuk data *customer* Suzana Huang:

Tanggal analisis = 10-07-2025

Tanggal terakhir = 06-01-2024

Maka:

$R_1 = 649$ hari

$F_1 = 1$

$M_1 = 13.500.000$

Pada perhitungan RFM berikutnya dihitung menggunakan rumus yang sama, sehingga akan diperoleh nilai RFM masing-masing *customer* sebagai berikut:

Tabel 2.2 Nilai RFM Data Transaksi Pelanggan

id	customer_name	recency	frequency	monetary
1	Suzana Huang	649	1	13.500.000
2	Rini	647	1	6.500.000
3	Robert Anastasia	627	1	6.500.000
4	Steviani	628	1	13.500.000
5	Bapak Dilen	649	1	17.600.000
6	Ibu Juli Lie	643	1	15.900.000
7	PT Svarna Dipa	645	3	32.500.000
...
165	Bapak Taslim	302	1	9.292.429

C. K-Means Clustering

Analisis pengelompokan (*clustering*) menggunakan algoritma *K-Means* diterapkan untuk mengidentifikasi segmen pelanggan yang homogen berdasarkan nilai RFM (*Recency*, *Frequency*, *Monetary*) yang telah dinormalisasi. Sebelum *K-Means* dijalankan, data RFM harus melalui tahap Normalisasi Min-Max untuk menghilangkan bias yang

disebabkan oleh perbedaan skala pada variabel. Proses normalisasi ini mengubah setiap nilai variabel ke dalam rentang [0, 1].

Menggunakan formula:

$$X_{\text{Norm}} = \frac{X - X_{\text{min}}}{X_{\text{Max}} - X_{\text{Min}}}$$

Maka diperoleh:

$$R_{\text{Min}} = 306, R_{\text{Max}} = 739$$

$$F_{\text{Min}} = 1, F_{\text{Max}} = 79$$

$$M_{\text{Min}} = 210.000, M_{\text{Max}} = 789.679.600$$

Sehingga diperoleh semisal untuk pelanggan Suzana Huang memiliki nilai normalisasi :

$$R_{\text{Norm}} = \frac{R - R_{\text{min}}}{R_{\text{Max}} - R_{\text{Min}}} = \frac{665 - 306}{739 - 306} = \frac{359}{433} = 0,8291$$

$$F_{\text{Norm}} = \frac{F - F_{\text{min}}}{F_{\text{Max}} - F_{\text{Min}}} = \frac{1 - 1}{79 - 1} = \frac{0}{78} = 0,0000$$

$$M_{\text{Norm}} = \frac{M - M_{\text{min}}}{M_{\text{Max}} - M_{\text{Min}}} = \frac{13.500.000 - 210.000}{789.679.600} = \frac{13.290.000}{789.469.600} = 0,0168$$

Dan data yang sudah dinormalisasi tersaji dalam tabel 2.3

Tabel 2.3 Hasil Normalisasi

id	Customer ID	Customer name	Norm-R	Norm-F	Norm-M
1	229	Suzana Huang	0,829099	0	0,01683
2	226	Rini	0,82448	0	0,007965
3	227	Robert Anastasia	0,778291	0	0,007965
4	228	Steviani	0,7806	0	0,01683
5	96	Bapak Dilen	0,829099	0	0,022022
6	181	Ibu Juli Lie	0,815242	0	0,019869
.....
165	139	Bapak Taslim	0	0	0,011501

Selanjutnya, *K-Means* mengukur kemiripan antar pelanggan menggunakan metrik *Euclidean Distance* (Jarak *Euclidean*). Jarak ini dihitung sebagai jarak terpendek garis lurus antara setiap titik data dan pusat klaster (*centroid*) yang terdekat, dirumuskan sebagai:

$$d = \sqrt{(R_{Norm} - R_c)^2 + (F_{Norm} - F_c)^2 + (M_{Norm} - M_c)^2}$$

Maka misal untuk pelanggan Suzana Huang memiliki vektor ternormalisasi $P_i = (0.8291; 0.0000; 0.0168)$ dan Centroid Klaster 1 adalah $C_1 = (0.477; 0.005; 0.026)$. Sehingga *Euclidean Distance* (d) dihitung sebagai berikut:

$$d = \sqrt{(0,829 - 0,477)^2 + (0,000 - 0,005)^2 + (0,017 - 0,026)^2}$$

$$d = \sqrt{(0,352)^2 + (-0,005)^2 + (-0,009)^2}$$

$$d = \sqrt{0,1239 + 0,0000 + 0,0001} \approx 0,352$$

Pengelompokan ini bertujuan untuk meminimalkan jarak dalam klaster (*intra-cluster distance*) dan memaksimalkan jarak antar klaster (*inter-cluster distance*).

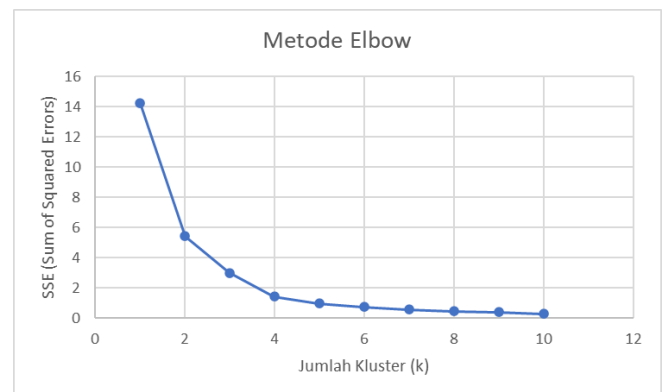
Tabel 2.4 Hasil Perhitungan *Euclidean Distance* (d)

Customer ID	Nama Pelanggan	Euclidean Distance	Cluster
229	Suzana Huang	0,352	1
226	Rini	0,0177	4
227	Robert Anastasia	0,1370	4
228	Steviani	0,0491	4
96	Bapak Dilen	0,0099	2
181	Ibu Juli Lie	0,1714	4
195	Ibu Rose	0,0665	4
113	Bapak Hoki	0,0058	4
122	Bapak Leo	0,0206	2
152	Bapak Yonathan	0,0007	2
149	Bapak Zulkifli	0,0324	2
136	Bapak Roni	0,0063	2
118	Bapak Kevin	0,1120	2
119	Bapak Lie Hendra	0,0968	1
220	Mr. Hendra	0,0061	4
213	Nyonya Fenny	0,6346	4
187	Ibu May	0,0726	2
153	Bapak Yonathan	0,1422	2
.....
111	Bapak Hendry	0,0790	2

Untuk menentukan jumlah klaster (k) yang optimal, digunakan metode *Elbow Method* dengan menghitung *Sum of Squared Errors* (SSE) untuk rentang nilai (k). SSE adalah total kuadrat jarak *Euclidean* dari setiap titik data ke centroid klaster terdekatnya. Semakin kecil nilai SSE, semakin padat klaster yang terbentuk. Nilai SSE bisa di lihat pada tabel 2.5.

Tabel 2.5 Hasil Perhitungan SSE

K	SSE (<i>Sum of Squared Errors</i>)
1	14,2216
2	5,43798
3	3,01305
4	1,45432
5	0,977765
6	0,768512
7	0,581764
8	0,466128
9	0,4028
10	0,3021



Gambar 2.1 Grafik *Elbow Method*

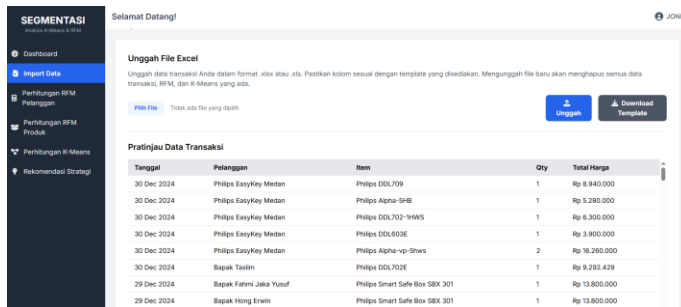
Hasil perhitungan SSE menunjukkan penurunan yang tajam pada nilai k awal dan mulai melandai pada $k = 3$. Titik pelandaian ini, yang dikenal sebagai "siku" (*elbow*), mengindikasikan bahwa penambahan klaster melebihi titik tersebut tidak lagi memberikan peningkatan signifikansi yang berarti, sehingga nilai $k = 3$ dipilih sebagai jumlah klaster optimal untuk segmentasi pelanggan.

III. HASIL DAN PEMBAHASAN

A. Implementasi Program

Implementasi program merupakan tahap penerapan dari hasil analisis dan perancangan sistem yang telah dibuat sebelumnya. Aplikasi dikembangkan sesuai dengan kebutuhan yang telah ditentukan, dengan tujuan agar pengguna dapat menjalankannya dengan mudah dan memperoleh pengalaman penggunaan yang optimal. Implementasi Alur Kerja Sistem.

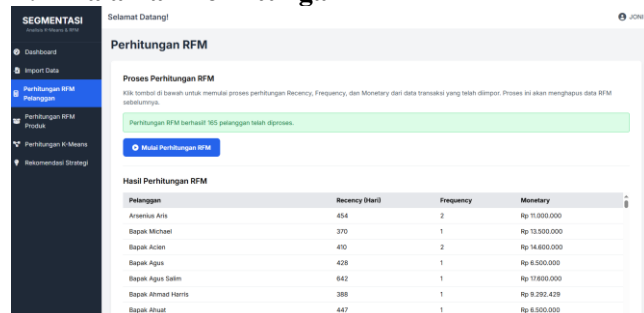
1. Halaman Import Data



Gambar 3.1 Halaman Import Data

Halaman *Import Data* Transaksi digunakan untuk mengunggah file transaksi penjualan dalam format *Excel* sesuai templat yang ditentukan. Setelah diunggah, sistem menampilkan pratinjau berisi data seperti tanggal, pelanggan, item, kuantitas, dan total harga untuk memastikan data berhasil dimuat sebelum dilakukan perhitungan RFM atau *K-Means*.

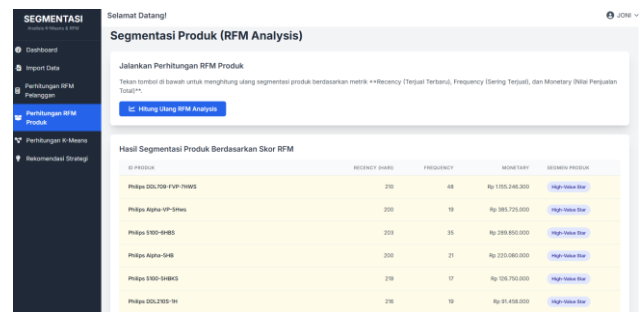
2. Halaman Perhitungan RFM



Gambar 3.2 Halaman Perhitungan RFM

Halaman Perhitungan RFM Pelanggan digunakan untuk menghitung nilai *Recency*, *Frequency*, dan *Monetary* berdasarkan data transaksi yang telah diimpor. Setelah tombol 'Mulai Perhitungan RFM' ditekan, sistem menampilkan hasil perhitungan RFM setiap pelanggan sebagai dasar untuk proses segmentasi menggunakan metode *K-Means*.

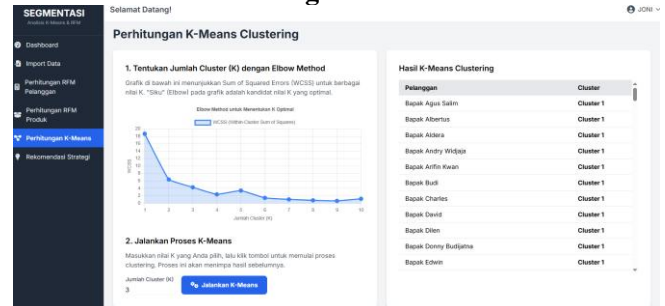
3. Halaman Perhitungan RFM Produk



Gambar 3.3 Halaman Perhitungan RFM Produk

Halaman Segmentasi Produk RFM yang digunakan untuk menghitung dan menampilkan nilai *Recency*, *Frequency*, dan *Monetary* dari setiap produk. Setelah tombol Hitung Ulang RFM Analysis ditekan, sistem menampilkan hasil segmentasi produk dalam tabel, lengkap dengan skor RFM dan kategori segmennya sebagai dasar analisis performa produk.

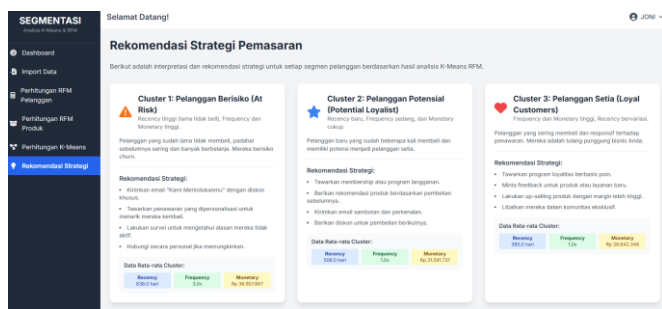
4. Halaman Perhitungan K-Means



Gambar 3.4 Halaman Perhitungan K-Means

Halaman Perhitungan *K-Means Clustering*. Pada bagian kiri terdapat grafik *Elbow Method* yang membantu menentukan jumlah *cluster* (*K*) yang paling optimal berdasarkan nilai *Within-Cluster Sum of Squares* (WCSS). Setelah memilih nilai *K*, pengguna memasukkannya ke kolom yang tersedia dan menekan tombol Jalankan *K-Means* untuk memulai proses *clustering*. Di sisi kanan halaman, sistem menampilkan hasil pengelompokan pelanggan dalam bentuk tabel, sehingga pengguna dapat melihat setiap pelanggan berada pada *cluster* yang mana untuk analisis segmentasi lebih lanjut.

5. Halaman Rekomendasi Strategi



Gambar 4.1 Halaman Rekomendasi Strategi

Halaman Rekomendasi Strategi Pemasaran yang memberikan interpretasi hasil segmentasi pelanggan berdasarkan analisis *K-Means* dan RFM. Setiap cluster pelanggan ditampilkan dalam kartu terpisah, lengkap dengan karakteristiknya serta rekomendasi strategi pemasaran yang sesuai.

1. *Cluster 1*: menggambarkan pelanggan berisiko yang sudah lama tidak bertransaksi, sehingga perlu pendekatan khusus untuk mengembalikan mereka.
2. *Cluster 2*: berisi pelanggan potensial yang mulai aktif membeli dan memiliki peluang menjadi pelanggan setia.
3. *Cluster 3*: merupakan pelanggan *loyal* yang sering bertransaksi dan sangat penting bagi kelangsungan bisnis.

Di bagian bawah setiap *cluster* juga ditampilkan data rata-rata *Recency*, *Frequency*, dan *Monetary* untuk membantu memahami profil masing-masing segmen secara lebih jelas.

IV. SIMPULAN

Penelitian ini sukses mengimplementasikan dan menguji sistem aplikasi berbasis web untuk segmentasi pelanggan. Berdasarkan pengujian *black box*, sistem ini terbukti valid dan mampu menjalankan seluruh alur kerja mulai dari impor data, kalkulasi RFM, proses *K-Means clustering*, hingga visualisasi hasil dan rekomendasi.

Penerapan algoritma *K-Means* pada data RFM pelanggan berhasil mengidentifikasi tiga segmen pelanggan utama. Pelanggan Baru (*New Customers*): Baru bertransaksi, frekuensi dan nilai belanja masih rendah. Pelanggan Pasif/Hilang (*Lost Customers*): Sudah lama tidak bertransaksi dan memiliki kontribusi rendah. Pelanggan Setia & Bernilai Tinggi (*Loyal & High-Value Customers*): Tulang punggung perusahaan dengan frekuensi dan total nilai belanja yang tinggi.

DAFTAR PUSTAKA

- Ahmad, A., & Dey, L. (2007). A *k-means clustering algorithm for mixed numeric and categorical data*. *Data & Knowledge Engineering*, 63(2), 503–527.SS
- Beni Tiya Kristanti, Achmad Junaidi, dan Eka Prakarsa Mandyartha. (2024). Implementasi *K-Means Clustering* dalam Segmentasi Pelanggan Berdasarkan Usia, Pendapatan, dan Model RFM (Studi Kasus: Lantikya Store Jombang).
- Chen, Y., & Liu, B. (2004). *Mining customer clusters for business analysis*. *Journal of Intelligent Information Systems*, 23(2), 133–150.
- Danang Satya Nugaha, Imam Thoib, Nafis Sururi, Fendy Bayu F, (2025). Segmentasi Pelanggan Berbasis Analisis RFM Menggunakan Algoritma *K-Means Clustering*.
- Dian Pratiwi, (2024). Analisis Segmentasi Pelanggan Ritel Online Menggunakan *K-Means Clustering* Berdasarkan Model *Recency*, *Frequency*, *Monetary* (RFM).
- Diotivano, Ikhwan Ruslianto, dan Dian Prawira, (2023). Sistem Segmentasi Pelanggan Potensial Menggunakan *Metode K-Means* dan Analisis RFM.
- Hadiwibowo. (2023). Evaluasi hasil klasterisasi menggunakan *silhouette coefficient*. *Jurnal Sains Data*, 8(4), 120–130.
- Hafidz, M., et al. (2024). Analisis segmentasi pelanggan menggunakan algoritma *K-Means* dan model RFM. *Jurnal Sains Komputer*, 12(1), 100–110.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann Publishers.
- Ilham. Sonata. (2019). Pengaruh Price Discount dan In-Store Display terhadap Impulse Buying Miniso. 46-53.
- Junaidi, & Madyartha. (2024). Optimalisasi algoritma *K-Means* untuk segmentasi pelanggan. *Jurnal Ilmu Komputer*, 10(1), 75–85.
- Karina Auliasari, & Mariza Kertaningtyas. (2019). Segmentasi pelanggan dengan algoritma *K-*

- Means clustering*. Jurnal Sistem Informasi Bisnis, 5(3), 60–70.
- Khan, S. S., & Ahmad, A. (2004). *Cluster center initialization algorithm for K-means clustering*. *Pattern Recognition Letters*, 25(11), 1293–1302.
- Kotler, P., & Keller, K. L. (2016). *Marketing Management (15th ed.)*. Pearson Education.s
- Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons.
- MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 281–297.
- Maryana, et al. (2018). Penerapan *data mining* untuk segmentasi pelanggan berbasis RFM. Jurnal Teknologi Informasi, 6(2), 45–52.
- Nurhayati, A., & Fajar, R. (2019). Analisis segmentasi pelanggan menggunakan metode RFM dan *K-Means clustering*. Jurnal Teknologi Informasi dan Komunikasi, 4(1), 25–33.
- Rousseeuw, P. J. (1987). *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Saragih, J. T. B., & Sinaga, A. (2021). Implementasi algoritma *K-Means clustering* untuk segmentasi pelanggan. Jurnal Informatika, 15(2), 87–95.
- Setiawan, B., & Rahmawati, D. (2020). Evaluasi hasil *clustering* menggunakan *Elbow* pada segmentasi pelanggan. Jurnal Data Sains, 2(3), 50–58.
- Singh, A., & Singh, P. (2020). *RFM analysis for customer segmentation: A case study*. *International Journal of Data Science and Analytics*, 7(3), 245–258.
- Triantara, (2025). Segmentasi Pelanggan Berbasis RFMT Menggunakan *K-Means* dan *Hierarchical Clustering*.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). Morgan Kaufmann.