

PEMANFAATAN K-MEANS UNTUK MENINGKATKAN PENJUALAN (STUDI KASUS: ALYA COLLECTION)

Windi Aulia Safitri¹, Herman Saputra^{2*}, Afdhal Syafnur¹

¹Sistem Informasi, Universitas Royal

²Sistem Komputer, Universitas Royal

*email: hermansaputra4@gmail.com

Abstract: *In a competitive digital era, business owners such as Alya Collection face challenges in determining the right fashion products to offer to consumers. Relying on intuition without data support often results in inaccurate and potentially detrimental decisions. This study aims to assist the sales decision-making process by applying a data mining method using the K-Means Clustering algorithm. This method was used to group sales data from February 2024 to January 2025 into three categories: best-selling, least-selling, and non-selling products. The study was conducted using a quantitative approach, through sales data collection and direct interviews with store owners. The system was built using the PHP programming language and MySQL database, and designed with tools such as UML, DFD, and ERD. The results of this clustering system are expected to provide strategic information useful for Alya Collection in planning stock, promotions, and product development, thereby increasing sales effectively and efficiently.*

Keywords: *data mining; k-means; sales; alya collection; information system*

Abstrak: Dalam era digital yang kompetitif, pemilik usaha seperti Alya Collection menghadapi tantangan dalam menentukan produk fashion yang tepat untuk ditawarkan kepada konsumen. Ketergantungan pada intuisi tanpa dukungan data seringkali mengakibatkan keputusan yang kurang akurat dan berisiko merugikan. Penelitian ini bertujuan untuk membantu proses pengambilan keputusan penjualan dengan menerapkan metode data mining menggunakan algoritma K-Means Clustering. Metode ini digunakan untuk mengelompokkan data penjualan dari Februari 2024 hingga Januari 2025 ke dalam tiga kategori: produk terlaris, kurang laris, dan tidak laris. Penelitian dilakukan dengan pendekatan kuantitatif, melalui pengumpulan data penjualan serta wawancara langsung dengan pemilik toko. Sistem yang dibangun menggunakan bahasa pemrograman PHP dan database MySQL, serta dirancang dengan alat bantu seperti UML, DFD, dan ERD. Hasil dari sistem klasterisasi ini diharapkan dapat memberikan informasi strategis yang berguna bagi Alya Collection dalam merencanakan stok, promosi, serta pengembangan produk, sehingga mampu meningkatkan penjualan secara efektif dan efisien.

Kata kunci: *data mining; k-means clustering; penjualan; alya collection; sistem informasi*

PENDAHULUAN

Era digital saat ini, industri fashion khususnya toko pakaian mengalami pertumbuhan yang pesat karena meningkatnya minat masyarakat terhadap produk yang memiliki keunikan dan kualitas tinggi. Kemajuan teknologi serta kemudahan dalam mengakses informasi membuat konsumen lebih sadar akan tren global dan berbagai pilihan yang ada di pasaran. Kondisi ini menuntut para pelaku usaha untuk tidak hanya

menawarkan produk yang menarik, tetapi juga memberikan nilai lebih pada setiap barang, baik dari sisi desain, harga, maupun kualitas bahan yang digunakan.

Dalam persaingan pasar yang semakin ketat, para pelaku usaha di bidang fashion dituntut untuk cepat menyesuaikan diri dengan perubahan tren serta selera konsumen. Mereka perlu memahami pergerakan pasar, mengetahui produk yang sedang diminati, dan memanfaatkan media sosial maupun platform digital untuk memperluas jangkauan pelanggan. Dengan strategi yang tepat, pelaku usaha dapat membangun hubungan yang solid dengan konsumen, meningkatkan loyalitas, dan meraih keberhasilan di industri yang terus berkembang ini[1]. Kemampuan beradaptasi dan berinovasi menjadi faktor utama dalam menghadapi berbagai tantangan sekaligus memanfaatkan peluang di pasar fashion dan aksesoris. Salah satu bisnis yang menonjol dalam bidang ini adalah Alya Collection, yang berfokus pada penyediaan pakaian trendi dan sesuai dengan kebutuhan pasar[2].

Alya Collection yang berlokasi di Jalan Budi Utomo, Kelurahan Siumbut Umbut, Kecamatan Kota Kisaran Timur, Kabupaten Asahan, merupakan toko pakaian yang menjadi tujuan favorit bagi para pecinta fashion. Toko ini menyediakan berbagai jenis busana, mulai dari pakaian kasual yang nyaman hingga busana formal yang elegan, mencakup koleksi untuk wanita dewasa, remaja, hingga anak-anak. Dengan desain modern dan pemilihan bahan berkualitas tinggi, setiap produk dirancang dengan ketelitian dan perhatian terhadap detail, sehingga pelanggan dapat menemukan pakaian yang sesuai dengan gaya serta selera mereka.

Beberapa penelitian terdahulu menunjukkan bahwa metode K-Means Clustering efektif dalam mengelompokkan data penjualan untuk mengetahui pola pembelian konsumen dan menentukan produk unggulan. Beberapa di antaranya adalah: Iwan Pii, Nana Suarna, & Nining Rahaningsih (2023) menerapkan K-Means Clustering pada data penjualan pakaian Dameyra Fashion untuk mengidentifikasi produk yang laku dan tidak laku berdasarkan cluster hasil analisis. Hasil penelitian ini menunjukkan bahwa clustering dapat membantu toko dalam mengetahui produk mana yang paling diminati konsumen[3]. Idham et al. (2025) melakukan implementasi K-Means untuk clustering data penjualan pakaian, mengelompokkan transaksi berdasarkan frekuensi pembelian, nilai transaksi, dan preferensi pelanggan. Penelitian ini berhasil mengelompokkan transaksi penjualan ke dalam beberapa cluster yang berbeda karakteristiknya, sehingga dapat memberi wawasan kepada pemilik usaha dalam strategi pemasaran dan segmentasi pelanggan[4]. Riwan Irosucipto Manarung et al. (2025) mengaplikasikan algoritma K-Means untuk segmentasi kebutuhan produk pada bisnis retail. Penelitian ini menunjukkan bahwa clustering dapat memetakan pola permintaan produk yang berbeda sehingga membantu dalam perencanaan stok barang[5].

Keterbaruan (novelty) dari penelitian ini terletak pada penerapan metode K-Means Clustering secara langsung pada data penjualan Alya Collection dengan mempertimbangkan karakteristik lokal dan kebutuhan spesifik toko skala menengah. Penelitian ini tidak hanya mengelompokkan produk berdasarkan tingkat penjualan, tetapi juga mengaitkan hasil clustering dengan rekomendasi strategi pengadaan stok dan pengambilan keputusan bisnis yang lebih tepat sasaran. Dengan demikian, penelitian ini memberikan pendekatan yang lebih aplikatif dan kontekstual dibandingkan penelitian sebelumnya.

Namun, Alya Collection masih menghadapi kesulitan dalam pengambilan keputusan, terutama dalam menentukan produk yang sedang diminati pasar dan memiliki peluang tinggi untuk terjual. Selama ini, keputusan tersebut lebih banyak didasarkan pada perkiraan atau intuisi tanpa dukungan data yang valid. Akibatnya, proses pemilihan produk sering kali tidak tepat sasaran dan berisiko besar. Selain itu, toko ini belum memiliki sistem teknologi informasi yang mampu mengelola data secara efisien dan efektif. Ketiadaan sistem tersebut menyebabkan proses pengambilan keputusan menjadi kurang terarah dan dapat berdampak negatif terhadap keberlangsungan usaha. Dampak yang mungkin muncul antara lain kerugian akibat stok produk baru yang tidak diminati konsumen serta pemesanan barang yang tidak diperlukan, yang justru menambah biaya operasional.

Alya Collection dapat menggunakan teknik data mining sebagai sarana yang sangat efisien. Data mining sendiri adalah proses untuk menggali informasi, pengetahuan, serta pola tersembunyi dari sekumpulan data berukuran besar.[6]. Untuk mengatasi permasalahan tersebut, penelitian ini mengajukan penerapan metode *K-Means Clustering* sebagai solusi dalam upaya meningkatkan penjualan pakaian di Toko Alya Collection. Metode ini dipilih karena memiliki keunggulan dalam mengelompokkan serta mengorganisir data secara terstruktur.

METODE

Metode *K-Means Clustering*

K-Means adalah salah satu metode pengelompokan data *non-hirarki* yang bertujuan untuk membagi sekumpulan data menjadi satu atau beberapa klaster. Metode ini mengelompokkan data sedemikian rupa sehingga data dengan karakteristik serupa berada dalam klaster yang sama, sedangkan data dengan karakteristik yang berbeda ditempatkan pada klaster yang berbeda [7]. *Clustering* merupakan salah satu metode data mining yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*) serta tidak memerlukan target output. Dalam data mining ada dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering*. *Hierarchical clustering* adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat [8].

Proses algoritma ini terdiri dari empat langkah sebagai berikut: Langkah pertama adalah menentukan jumlah *cluster* K yang diinginkan. Menentukan nilai *centroid cluster*. Penentuan nilai *centroid* pada tahap awal dilakukan secara *random*, sedangkan pada tahap iterasi digunakan rumus seperti dibawah ini:

Keterangan:

Vij = *Centroid* rata-rata *cluster* ke-*I* untuk variabel ke-*j*

N_j = Jumlah Barang *Cluster* ke-*j*

$j, k = \text{Indeks dari cluster}$

i = Indeks dari variable

X_{kj} = Nilai data ke-k variabel ke-j untuk *cluster* tersebut

Menghitung jarak antara titik *centroid* dengan titik tiap objek menggunakan metode *Euclidean Distance*. *Euclidean Distance* adalah metode pengukuran jarak garis lurus antara dua titik dalam ruang *Euclidean*. Rumus untuk menghitung *Euclidean Distance* adalah sebagai berikut:

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \dots \dots \dots (2)$$

Keterangan:

V_{ij} = *Centroid* rata-rata *cluster* ke-*I* untuk variabel ke-*j*

Ni = Jumlah Barang *Cluster* ke-i

i,k = Indeks dari cluster

j = Indeks dari variable

X_{kj} = Nilai data ke- k variabel ke- j untuk *cluster* tersebut

Kelompokkan objek berdasarkan jarak ke *Centroid* terdekat.

Ulangi langkah ke-2 hingga ke-4, lakukan iterasi hingga *centroid* bernilai optimal.

HASIL DAN PEMBAHASAN

Analisis Data

Data yang dijadikan bahan masukan dalam penelitian ini adalah data *history* penjualan pada Toko Alya Collection. Data yang digunakan adalah data dari periode Januari 2024 sampai dengan Desember 2024. Adapun datanya adalah sebagai berikut :

Tabel 1. Rekapitulasi Data Penjualan Pakaian di Toko Alya Collection Pada Bulan Januari 2024 – Desember 2024

No	Barang	Bulan											
		Jan	Feb	Mar	Apr	Mei	Jun	Jul	Ags	Sep	Okt	Nov	Des
1	Gamis	10	15	8	6	11	8	5	10	6	12	16	11
2	Tunik	8	11	5	8	10	4	8	7	9	6	14	8
3	One Set	5	7	10	5	8	9	4	10	5	8	12	10
4	Kemeja	9	13	15	9	11	14	10	14	11	15	20	13
...
9	Sweater	0	4	0	6	0	0	2	15	5	2	13	11
10	Hoodie	0	1	3	5	0	2	4	11	4	0	8	2

Modeling

Metode *K-Means Clustering* hanya bisa mengolah data dalam bentuk angka, maka untuk data yang berbentuk nominal harus diinisialisasikan terlebih dahulu dalam bentuk angka. Langkahnya adalah:

Urutkan data sesuai dengan tingkat frekuensi kemunculannya. Beri inisialisasi pada data tersebut, dimulai dari data dengan frekuensi tertinggi bernilai 1, diikuti oleh data berikutnya dengan nilai 2, 3, dan seterusnya.

Berikut data penjualan pakaian berdasarkan kategori dari beberapa *cluster*, kemudian jadikan data tersebut menjadi cluster.

Menetukan Pusat Awal *Cluster* “*Centroid*”

Sebagai langkah awal disimpulkan bahwa:

Diambil data Gamis sebagai pusat *Cluster* Ke-1: (10,14, 8, 6,11, 8, 5, 10, 5, 12, 16, 11).

Diambil data Tunik sebagai pusat *Cluster* Ke-2: (8, 11, 5, 8, 10, 4, 8, 7, 9 6, 14, 8).

Diambil data One Set sebagai pusat *Cluster* Ke-3: (5, 7, 10, 5, 8, 9, 4, 10, 5, 8, 12, 10).

Perhitungan Jarak Pusat *Cluster*

Perhitungan jarak dari data ke 1 terhadap pusat *Cluster* 1.

$$A1 = \sqrt{(10-10)^2 + (15-15)^2 + (8-8)^2 + (6-6)^2 + (11-11)^2 + (8-8)^2 + (5-5)^2 + (10-10)^2 + dst + (11-11)^2} = 0,00$$

$$A2 = \sqrt{(8-10)^2 + (11-15)^2 + (5-8)^2 + (8-6)^2 + (10-11)^2 + (4-8)^2 + (8-5)^2 + (7-10)^2 + dst + (8-11)^2} = 11,225,$$

dan seterusnya perhitungan jarak dari anggota ke 3 sampai 91 terhadap pusat *cluster*.

Perhitungan Jarak dari data ke 1 terhadap pusat *Cluster* 2.

$$A1 = \sqrt{(10-8)^2 + (15-11)^2 + (8-6)^2 + (10-5)^2 + (4-6)^2 + (8-10)^2 + (7-8)^2 + (10-10)^2 + dst + (11-8)^2} = 11,225$$

$$A2 = \sqrt{(8-8)^2 + (11-11)^2 + (8-8)^2 + (6-6)^2 + (11-11)^2 + (8-8)^2 + (5-5)^2 + (10-10)^2 + dst + (8-11)^2} = 0,00,$$

dan seterusnya perhitungan jarak dari anggota ke 3 sampai 91 terhadap pusat *cluster*.

Perhitungan Jarak pertama dengan pusat *Cluster* ketiga.

$$A1 = \sqrt{(10-5)^2 + (15-7)^2 + (8-10)^2 + (6-5)^2 + (11-8)^2 + (8-9)^2 + (5-4)^2 + (10-10)^2 + dst + (11-10)^2} = 11,78$$

$$A2 = \sqrt{(8-5)^2 + (11-7)^2 + (5-10)^2 + (8-5)^2 + (10-8)^2 + (4-9)^2 + (8-4)^2 + (7-10)^2 + dst + (8-10)^2} = 11,87 \text{ dan, seterusnya}$$

perhitungan jarak dari anggota ke 3 sampai 91 terhadap pusat *cluster*.

Sehingga hasil perhitungan jarak selengkapnya adalah:

Tabel 2. Hasil Iterasi Pertama

No	Pusat Cluster	C1	C2	C3	Keanggotaan	Min Jarak	Kuadrat
1	Gamis	0,000	11,225	11,790	C1	0	0
2	Tunik	11,225	0,000	11,874	C2	0	0
3	One Set	11,790	11,874	0,000	C3	0	0
4	Kemeja	13,928	20,149	18,358	C1	13,928	194
...
10	Hoodie	27,821	21,119	19,209	C3	19,209	369

Kolom pertama memperlihatkan jarak masing-masing data ke pusat klaster pertama, kolom kedua menampilkan jarak data ke pusat klaster kedua, sedangkan kolom ketiga menunjukkan jarak data ke pusat klaster ketiga.

Penetuan Pusat *Cluster* Baru

Karena klaster C1 memiliki 29 anggota, maka perhitungan pembentukan klaster baru dari bulan Januari hingga Desember menjadi sebagai berikut:

$$C1 = \left(\frac{10+9+13+10+9+7+12+16+8+3+4+10+13+19+4+7+6+12+3+dst+12+8}{29} \right) = 8,83$$

$$= \left(\frac{15+13+10+13+12+10+10+15+11+8+11+16+21+10+14+10+10+dst+20+11}{29} \right) = 12,41$$

Dan seterusnya perhitungan *cluster* sampai bulan Desember.

Karena C2 hanya mempunyai 16 anggota maka *cluster* baru dari Januari sampai Desember menjadi:

$$C2 = \left(\frac{8+3+4+0+8+12+10+1+0+6+8+2+2+0+3+3}{16} \right) = 4,40$$

$$= \left(\frac{11+5+9+7+11+8+8+5+4+12+12+5+9+7+3+5}{16} \right) = 7,56$$

Dan seterusnya perhitungan *cluster* sampai bulan Desember.

Karena C3 hanya mempunyai 46 anggota maka *cluster* baru dari Januari sampai Desember menjadi:

$$C3 = \left(\frac{5+0+1+0+0+6+1+4+0+5+6+0+5+0+1+2+0+0+0+5+0+8+0+dst+0+0+7}{46} \right) = 2,434$$

$$= \left(\frac{7+3+9+4+1+4+4+9+4+5+10+4+7+5+7+10+8+2+3+6+9+4+dst+0+0+0}{46} \right) = 4,72$$

Dan seterusnya perhitungan *cluster* sampai bulan Desember.

Perhitungan jarak dilakukan dari data ke-1 hingga data ke-91 terhadap pusat klaster, sehingga diperoleh hasil perhitungan jarak secara keseluruhan sebagai berikut:

Tabel 3. Hasil Iterasi Kedua

No	Pusat Cluster	C1	C2	C3	Keanggotaan	Min Jarak	Kuadrat
1	Gamis	13,721	13,689	17,543	C2	13,689	187,398
2	Tunik	17,988	7,828	12,736	C2	7,828	61,273
3	One Set	17,788	11,775	8,897	C3	8,897	79,155
4	Kemeja	7,696	20,344	24,350	C1	7,696	59,221
...
9	Sweater	29,437	17,390	12,437	C3	12,437	154,676
10	Hoodie	32,173	19,194	11,795	C3	11,795	139,111

Penetuan Cluster Baru

Penetuan klister baru didasarkan pada nilai jarak masing-masing data terhadap pusat klister pertama, kedua, dan ketiga, yang dilakukan melalui proses iterasi dari iterasi ke-1 hingga iterasi ke-5. Pada iterasi ke-5 dan ke-6, hasil yang diperoleh menunjukkan kesamaan, sehingga tidak terjadi perubahan lagi pada proses pengelompokkan.

Karena C1 memiliki 33 anggota maka perhitungan *cluster* baru dari Januari sampai Desember menjadi:

$$C1 = \left(\frac{10+9+13+10+9+7+12+16+8+3+4+10+13+19+4+7+6+12+3+dst+12+8}{33} \right) = 7,76$$

$$= \left(\frac{15+13+10+13+12+10+10+15+11+8+11+16+21+10+14+10+10+dst+20+11}{33} \right) = 11,48$$

Dan seterusnya perhitungan *cluster* sampai bulan Desember.

Karena C2 hanya mempunyai 35 anggota maka *cluster* baru dari Januari sampai Desember menjadi:

$$C2 = \left(\frac{8+3+4+0+8+12+10+1+0+6+8+2+2+0+3+3+dst+0+6+3+0+3+5}{35} \right) = 4,26$$

$$= \left(\frac{11+5+9+7+11+8+8+5+4+12+12+5+9+2+0+4+9+4+9+dst+7+3+5}{35} \right) = 7,6$$

Dan seterusnya perhitungan *cluster* sampai bulan Desember.

Karena C3 hanya mempunyai 23 anggota maka *cluster* baru dari Januari sampai Desember menjadi:

$$C3 = \left(\frac{5+0+1+0+0+6+1+4+0+5+6+0+5+0+1+2+0+0+0+5+0+8+0+dst+0+0+7}{23} \right) = 1,434$$

$$= \left(\frac{7+3+9+4+1+4+4+9+4+5+10+4+7+5+7+10+8+2+3+6+9+4+dst+0+0+0}{23} \right) = 2,30$$

Dan seterusnya perhitungan *cluster* sampai bulan Desember.

Perhitungan jarak dilakukan antara data ke-1 sampai data ke-91 terhadap pusat klaster, dan hasil lengkap dari perhitungan jarak tersebut adalah sebagai berikut:

Tabel 4. Hasil Iterasi Kelima dan Keenam

No	Pusat Cluster	C1	C2	C3	Keanggotaan	Min Jarak	Kuadrat
1	Gamis	14,498	12,559	25,282	C2	12,559	157,723
2	Tunik	17,857	9,883	19,141	C2	9,883	97,666
3	One Set	17,525	9,474	16,552	C2	9,474	89,752
4	Kemeja	7,739	19,203	33,150	C1	7,739	59,897
...
9	Sweater	29,373	15,824	13,816	C3	13,816	190,894
10	Hoodie	31,764	18,402	8,614	C3	8,614	74,198

Berdasarkan hasil dari iterasi pertama dan kedua, dapat disimpulkan bahwa hasil pengelompokan pada kedua iterasi tersebut menunjukkan kesamaan, sehingga proses iterasi atau pengulangan tidak perlu dilakukan lagi.

Adapun hasil akhir dari proses clustering adalah sebagai berikut:

Tabel 5. Pengelompokan Anggota *Cluster*

No	Nama Pakaian	Cluster	Sangat Laris (C1)	Laris (C2)	Kurang Laris (C3)
1	Gamis	C2		*	
2	Tunik	C2		*	
3	One Set	C2		*	
4	Kemeja	C1	*		
...
9	Sweater	C3			*
10	Hoodie	C3			*
Total			3	5	2

KESIMPULAN

Berdasarkan hasil implementasi dan pengujian sistem pengelompokan data penjualan pakaian menggunakan algoritma K-Means Clustering di Toko Alya Collection, dapat disimpulkan bahwa sistem ini mampu mempercepat dan mempermudah proses pengelompokan data penjualan dibandingkan cara manual. Melalui penerapan algoritma tersebut, diperoleh tiga kategori penjualan, yaitu produk sangat laris, laris, dan kurang laris, sehingga pemilik toko dapat lebih mudah mengidentifikasi produk dengan tingkat permintaan tinggi maupun rendah. Dengan demikian, sistem ini membantu dalam pengambilan keputusan terkait pengelolaan stok dan strategi penjualan yang lebih efektif dan tepat sasaran.

DAFTAR PUSTAKA

- [1] A. Riset, A. Indrawati, N. Maidah, T. Novitasari, and E. M. Hanesti, “CUSTOMER RELATIONSHIP MANAGEMENT,” vol. 1, no. 2, pp. 15–18, 2024.
- [2] N. S. Putri and A. Muanas, “Pengaruh Digital Marketing , Customer Relationship Management , dan Kualitas Pelayanan Terhadap Loyalitas Pelanggan di Pasar BTC Surakarta Abstrak Pendahuluan,” vol. 8, no. 3, pp. 764–774, 2025.
- [3] I. Pii, N. Suarna, and N. Rahaningsih, “PENERAPAN DATA MINING PADA PENJUALAN PRODUK PAKAIAN DAMEYRA FASHION MENGGUNAKAN METODE K-MEANS CLUSTERING,” vol. 7, no. 1, 2023.
- [4] H. Rosika *et al.*, “PAKAIAN MENGGUNAKAN METODE K-MEANS Pada era digital perkembangan teknologi semakin berkembang khusunya pengelolaan data penjualan menjadi semakin krusial bagi bisnis untuk memahami perilaku konsumen , mengidentifikasi beberapa kelompok atau cluster memili,” vol. 5, pp. 221–231, 2024.
- [5] R. I. Manarung, E. Widodo, and A. M. Rifai, “Sales Data Clustering Using the K-Means Algorithm to Determine Retail Product Needs,” vol. 5, no. April, pp. 226–234, 2025.
- [6] D. I. Lazada, “PEMANFAATAN ALGORITMA K-MEANS DALAM,” vol. 13, no. 2, pp. 1297–1306, 2025.
- [7] Yudi Agusta, “K-Means-Penerapan, Permasalahan dan Metode Terkait,” *J. Sist. dan Inform.*, vol. 3, no. Pebruari, pp. 47–60, 2007.
- [8] E. Safitri, “Implementasi Algoritma K-Means Clustering Dalam Menentukan Strategi Marketing Dalam Penjualan Ikan (Studi Kasus: Grosir Ikan Tani Mas Tanjung Morawa),” *JIKTEKS J. Ilmu Komput. dan Teknol. Inf.*, vol. 02, no. 02, pp. 23–33, 2024.