

Segmentasi Produk Pakaian Menggunakan Algoritma K-Means Clustering dan Particle Swarm Optimization untuk Strategi Pemasaran

Rio Aji Hadyanta Putra^{*}, Putri Taqwa Prasetyaningrum

Fakultas Teknologi Informatika, Program Studi Sistem Informasi, Universitas Mercu Buana, Yogyakarta, Indonesia

Email: ^{1,*}191210085@student.mercubuana-yogya.ac.id, ²putri@mercubuana-yogya.ac.id

Email Penulis Korespondensi: 191210085@student.mercubuana-yogya.ac.id

Submitted: 16/05/2025; Accepted: 30/06/2025; Published: 30/06/2025

Abstrak—Penelitian ini bertujuan untuk menganalisis segmentasi produk pakaian menggunakan algoritma *K-Means Clustering* yang dioptimasi dengan *Particle Swarm Optimization* (PSO) untuk menghasilkan segmentasi produk yang akurat sehingga dapat mendukung strategi pemasaran yang lebih efektif pada sebuah perusahaan. Data yang akan digunakan untuk analisis berasal dari transaksi penjualan sebuah perusahaan produksi pakaian yang menyediakan berbagai kategori produk pakaian. Data transaksi penjualan produk pakaian yang digunakan dalam penelitian ini meliputi nomor transaksi, jenis produk, kuantitas yang terjual, dan total harga penjualan dengan total data sebanyak 333 baris data. Data tersebut diolah menggunakan bahasa pemrograman *Python* melalui *Visual Studio Code*. Proses segmentasi dilakukan dengan algoritma *K-Means* untuk mengelompokkan produk dan metode *Elbow* untuk menentukan kluster. Kluster yang dihasilkan metode *elbow* kemudian dioptimasi menggunakan PSO untuk mendapatkan jumlah kluster dan titik pusat kluster yang lebih optimal. Evaluasi kluster dilakukan dengan cara membandingkan nilai metrik *Davies Bouldin Index* (DBI), *Silhouette Score*, *Sum of Square Error* (SSE), dan *SSW/SSB*. Meskipun nilai DBI mengalami sedikit peningkatan dari 0.6690 menjadi 0.6878 yang menunjukkan peningkatan kemiripan antar kluster, namun peningkatan *Silhouette Score* dari 0.5513 menjadi 0.5569 mengindikasikan bahwa konsistensi kluster semakin membaik. Selain itu, penurunan nilai SSE dari 418.52 menjadi 313.25 menandakan bahwa penyebaran data dalam masing-masing kluster menjadi lebih rapat, sedangkan rasio *SSW/SSB* yang turun signifikan dari 0.4582 menjadi 0.3075 menunjukkan bahwa kluster yang terbentuk memiliki batas yang lebih tegas dan pemisahan yang lebih baik antar kluster. Hasil penelitian menghasilkan 4 kluster produk dengan karakteristik yang berbeda, sehingga perusahaan dapat menetapkan strategi pemasaran yang berbeda-beda.

Kata Kunci: Segmentasi Produk; K-Means Clustering; Particle Swarm Optimization; Davies Bouldin Index; Strategi Pemasaran

Abstract—This research aims to analyze product segmentation in the apparel industry using the K-Means Clustering algorithm optimized with Particle Swarm Optimization (PSO) to generate accurate product segmentation that can support more effective marketing strategies for a company. The data used in this analysis were obtained from sales transactions of a clothing manufacturing company that offers various categories of apparel products. The dataset consists of 333 rows and includes transaction numbers, product types, quantities sold, and total sales values. The data were processed using the Python programming language via Visual Studio Code. The segmentation process was initially performed using the K-Means algorithm to group products, and the Elbow method was applied to determine the optimal number of clusters. The number of clusters obtained from the Elbow method was then optimized using PSO to find more optimal cluster counts and centroids. Cluster evaluation was conducted by comparing the values of several metrics, including the Davies-Bouldin Index (DBI), Silhouette Score, Sum of Squared Error (SSE), and the SSW/SSB ratio. Although the DBI increased slightly from 0.6690 to 0.6878, indicating greater similarity between clusters, the improvement in the Silhouette Score from 0.5513 to 0.5569 suggests better internal consistency within the clusters. Furthermore, the reduction in SSE from 418.52 to 313.25 indicates a tighter distribution of data within clusters, while the significant decrease in the SSW/SSB ratio from 0.4582 to 0.3075 demonstrates more clearly defined cluster boundaries and improved separation. The results of the study produced four distinct product clusters, enabling the company to implement more targeted and differentiated marketing strategies.

Keywords: Product Segmentation; K-Means Clustering; Particle Swarm Optimization; Davies Bouldin Index; Marketing Strategy

1. PENDAHULUAN

Perkembangan teknologi di era digital telah mendorong kemajuan pesat dalam industri jasa produksi pakaian, terutama dalam cara perusahaan memahami preferensi konsumen dan menyusun strategi pemasaran. Hal tersebut terjadi seiring dengan meningkatnya permintaan akan produk pakaian yang beragam jenis. Untuk dapat bersaing secara efektif, perusahaan di sektor industri jasa pakaian perlu mengoptimalkan penawaran produk. Segmentasi produk menjadi salah satu solusi yang dapat digunakan perusahaan dalam mengelompokkan produk berdasarkan karakteristik spesifik dan merancang strategi pemasaran agar proses produksi lebih tepat sasaran. Ketidakpastian dalam transaksi penjualan sering membuat pemilik ragu dalam menentukan jumlah stok produksi yang tepat [1]. Dengan demikian, penerapan segmentasi produk yang tepat dapat meminimalkan risiko kegagalan produk di pasar sehingga perusahaan dapat menyusun portofolio produk yang relevan dengan kebutuhan target pasar.

Pengelompokan produk di industri pakaian tidaklah mudah, karena banyaknya variasi yang ada pada jenis pakaian yang dapat ditawarkan di pasar. Segmentasi produk pakaian memungkinkan perusahaan untuk dapat mengoptimalkan penawaran produk, menetapkan harga secara lebih strategis, dan merancang strategi promosi yang efektif. Dengan mengidentifikasi segmentasi produk yang tepat, perusahaan dapat mengembangkan produk yang mencerminkan kebutuhan dan keinginan pelanggan. *Clustering* memberikan dasar untuk menyesuaikan strategi pemasaran berdasarkan kebutuhan pasar, sehingga meningkatkan efisiensi operasional dan daya saing bisnis [2]. Selain itu, perusahaan dapat merancang produk serta strategi pemasaran yang lebih tepat dan efisien dengan memahami kebutuhan setiap segmen. Kemampuan untuk menyesuaikan produk, harga, dan promosi sesuai dengan

karakteristik setiap segmen produk bukan hanya menjadi keunggulan kompetitif, tetapi juga menjadi fondasi penting dalam membangun loyalitas pelanggan di tengah ketatnya persaingan industri pakaian.

Sejalan dengan pentingnya segmentasi tersebut, penelitian ini berfokus pada penerapan algoritma *K-Means Clustering*, sebuah teknik pengelompokan data non-hirarki yang memisahkan data ke kluster berdasarkan karakteristik yang serupa. Algoritma *K-Means* merupakan salah satu dari algoritma yang sering digunakan dalam pengelompokan data karena kesederhanaan dan efisiensi [3]. Teknik data mining dan optimasi algoritma berperan penting dalam membantu perusahaan mengidentifikasi pola penjualan dan preferensi konsumen sehingga dapat merumuskan paket produk yang tepat sasaran [4]. Selain itu, penerapan segmentasi produk yang tepat tidak hanya meningkatkan efisiensi produksi dan distribusi, tetapi juga berkontribusi pada peningkatan loyalitas pelanggan dengan menyesuaikan produk dan promosi berdasarkan kebutuhan spesifik tiap segmen [5]. Namun demikian, tantangan utama dalam penerapan *K-Means* terletak pada penentuan jumlah kluster (nilai k) yang tepat dan pemilihan pusat kluster awal yang optimal, karena kedua faktor tersebut sangat memengaruhi akurasi dan hasil akhir segmentasi yang dihasilkan. Algoritma *K-Means* sering terjebak pada solusi lokal optimal dan sangat sensitif terhadap inisialisasi titik centroid awal [6]. Dalam bidang pembelajaran mesin, masalah ketidakseimbangan kelas sering menjadi hambatan utama dalam mencapai performa klasifikasi yang optimal [7]. Untuk mengatasi tantangan tersebut, penelitian ini mengintegrasikan algoritma *Particle Swarm Optimization* (PSO) sebagai metode optimasi dalam proses *clustering*. PSO merupakan algoritma *metaheuristik* yang efektif dalam menyelesaikan permasalahan optimasi *non-linier* dan multidimensi. Dalam segmentasi produk, PSO digunakan untuk mencari solusi optimal terkait pemilihan pusat kluster awal dan jumlah kluster pada algoritma *K-Means*, sehingga dapat mengatasi kekurangan dari *K-Means* yang sangat bergantung pada inisialisasi awal. Integrasi antara metode pengelompokan dan optimasi menjadi solusi strategis perusahaan dalam menghadapi tantangan kompleksitas data besar dan dinamis di industri pakaian.

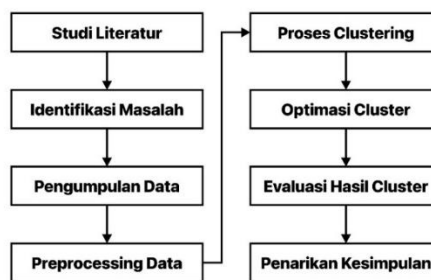
Penelitian yang dilakukan oleh Alganiu dkk. [8] melakukan perbandingan antara algoritma *K-Means* dan *K-Medoids* pada data transaksi minimarket menunjukkan bahwa hasil yang didapatkan oleh algoritma *K-Means* memiliki keterpisahan data yang buruk. Penelitian oleh Dinata dkk. [9] menerapkan *K-Means Clustering* pada data penjualan untuk menentukan stok optimal dengan memanfaatkan metode *elbow* dan validasi menggunakan *Davies Bouldin Index* (DBI) mampu mengidentifikasi kluster produk dalam pengambilan keputusan pengadaan barang. Penelitian yang dilakukan oleh Ngaeni dkk. (2024) menunjukkan bahwa kombinasi algoritma *K-Means* dan metode pengambilan keputusan seperti TOPSIS dapat digunakan secara efektif untuk mendukung strategi pemasaran dan mampu menghasilkan rekomendasi media promosi yang lebih sesuai untuk masing-masing kelompok sasaran [10]. Penelitian yang dilakukan oleh Ardi dkk. [11] menggunakan algoritma *K-Means Clustering* untuk segmentasi pelanggan pada bisnis *fashion* berhasil membentuk dua kluster pelanggan, namun penelitian ini belum mengintegrasikan metode optimasi sehingga hasil segmentasi masih bergantung pada inisialisasi *centroid* secara acak. Sementara itu, penelitian oleh Naufal & Suseno [12] menunjukkan algoritma klasifikasi dengan optimasi PSO secara signifikan meningkatkan akurasi prediksi kelayakan kredit pada koperasi hingga 87,23%. Penelitian oleh Dista dan Abdulloh [13] juga membuktikan bahwa algoritma *K-Means* dan PSO mampu meningkatkan efektivitas dalam proses segmentasi pengunjung mall, terlihat dari penurunan nilai DBI setelah dilakukan optimasi. Penelitian oleh Herdiana dkk. [14] membuktikan bahwa kombinasi *K-Means* dan PSO mampu meningkatkan akurasi segmentasi dalam industri kreatif yang lebih optimal dan membantu dalam penentuan strategi pemasaran berbasis data secara lebih presisi. Hal ini menunjukkan bahwa PSO memiliki kapabilitas tinggi dalam melakukan pencarian parameter optimal dalam ruang solusi yang kompleks.

Penelitian ini menganalisis data transaksi dari sebuah perusahaan yang bergerak di bidang jasa produksi pakaian yang menawarkan berbagai kategori produk pakaian. Pelanggan utama perusahaan ini mencakup berbagai instansi, baik dari sektor swasta maupun badan usaha milik negara yang memiliki preferensi dan kebutuhan berbeda-beda terhadap jenis dan spesifikasi produk pakaian. Penelitian ini bertujuan untuk menerapkan algoritma *K-Means Clustering* yang dioptimasi dengan PSO dalam segmentasi produk pakaian, sehingga perusahaan dapat mengembangkan strategi pemasaran yang lebih terarah, efisien, dan sesuai dengan kebutuhan pasar. Dengan menggabungkan algoritma *K-Means* dengan PSO, diharapkan dapat ditemukan jumlah kluster yang optimal dan menghasilkan segmentasi yang lebih representatif. Segmentasi yang lebih akurat tersebut dapat mendukung pengambilan keputusan yang lebih efektif dalam penyusunan strategi pemasaran.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dalam penelitian ini, diperlukan suatu kerangka kerja yang sistematis dan terstruktur untuk menyelesaikan masalah yang ada, khususnya dalam menerapkan algoritma *K-Means Clustering* yang dikombinasikan dengan *Particle Swarm Optimization* (PSO) untuk memperoleh hasil segmentasi produk yang optimal. Berikut gambar diagram alir yang menjelaskan tahapan penelitian ini:



Gambar 1. Tahapan Penelitian

Berdasarkan tahap penelitian yang ada pada Gambar 1 di atas, dapat diuraikan pembahasan masing-masing urutan tahapan penelitian yang dilakukan untuk menganalisis segmentasi produk. Penelitian dimulai dengan studi literatur yang bertujuan untuk memahami konsep dan teori yang relevan dalam tahap penelitian. Setelah itu, dilakukan identifikasi masalah untuk merumuskan tujuan dan ruang lingkup penelitian. Selanjutnya, pengumpulan data dilakukan dengan cara mengumpulkan data transaksi penjualan dari perusahaan. Setelah itu data yang telah diperoleh dilakukan *preprocessing data* agar siap digunakan dalam analisis. Tahapan *preprocessing data* sangat krusial dalam meningkatkan akurasi hasil *clustering* dengan menghapus data tidak relevan dan menormalisasi data [15]. Tahapan *preprocessing data* yang dilakukan dalam penelitian ini meliputi:

- Data cleaning, dalam proses ini data mentah yang telah dikumpulkan lalu disaring mana data yang tidak relevan, tidak lengkap, atau mengandung kesalahan. Tujuannya adalah untuk memastikan kualitas data yang akan dianalisis tetap akurat.
- Normalisasi data yang memiliki skala yang berbeda agar seluruh atribut numerik berada dalam skala yang sama. Normalisasi ini dilakukan untuk menghindari dominasi satu fitur terhadap fitur lainnya dalam proses clustering menggunakan *Label Encoding* dan *StandardScaler*.
- Data direduksi untuk mempercepat proses *clustering*, reduksi dimensi dapat membantu merangkum informasi dari beberapa fitur numerik menjadi dua komponen utama tanpa kehilangan informasi yang signifikan.

Tahap berikutnya adalah proses *clustering* di mana data dilakukan pengelompokan dengan algoritma *K-means*. Selanjutnya hasil pengelompokan dilakukan optimasi kluster (nilai k) menggunakan metode *Particle Swarm Optimization* (PSO) untuk mendapatkan jumlah kluster dan posisi *centroid* yang optimal. Setelah proses optimasi selesai, dilakukan evaluasi hasil kluster untuk menilai efektivitas segmentasi yang diperoleh dengan cara membandingkan beberapa nilai metrik seperti *Davies-Bouldin Index*, *Silhouette Score*, dan rasio dari SSW/SSB di mana SSW (*Sum of Square Error*) berfungsi untuk mengukur kerapatan data terhadap *centroid* kluster dan SSB (*Sum of Squares Between*) mengukur seberapa jauh jarak antar kluster dengan *centroid*. Akhirnya, tahapan penelitian diakhiri dengan penarikan kesimpulan berdasarkan hasil analisis dan interpretasi, yang selanjutnya dapat digunakan untuk mendukung strategi pemasaran perusahaan.

2.2 Metode Elbow

Metode *elbow* merupakan salah satu metode yang sering digunakan untuk menentukan jumlah kluster pada proses *clustering*. Tujuan dari metode ini yaitu untuk memaksimalkan atau menentukan jumlah kluster yang optimal [16]. Jumlah kluster dapat diketahui dengan cara melihat presentase setiap kluster yang membentuk siku pada suatu titik tertentu di grafik. Jumlah kluster dari pengujian dilakukan evaluasi dengan *Sum of Square Error* (SSE) untuk mengukur jarak terpisahnya satu kluster dengan kluster lainnya, semakin kecil nilai SSE maka semakin terpusat kluster yang terbentuk. Untuk menghitung sebuah SSE dapat menggunakan rumus sebagai berikut:

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|x_i - C_k\|^2 \quad (1)$$

Di mana, K menyatakan jumlah kluster, x_i adalah data ke- i dalam kluster, dan C_k merupakan titik pusat (*centroid*) dari kluster ke- k .

2.3 K-Means Clustering

Algoritma K-Means merupakan salah satu metode *unsupervised learning* yang digunakan untuk melakukan *clustering* atau pengelompokan data ke dalam beberapa kluster. Tujuan utama dari *clustering* adalah untuk mengidentifikasi struktur dalam data, yang dapat membantu dalam pemahaman lebih lanjut tentang kelompok atau kategori yang ada di dalamnya [17]. Metode ini termasuk dalam jenis *non-hierarchical clustering* yang bekerja dengan cara mempartisi objek yang ada ke dalam kluster atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu kluster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kluster yang lain. *K-Means Clustering* dalam proses segmentasi bekerja dengan cara membagi data ke dalam kluster berdasarkan kedekatan jarak antar data dan *centroid* secara iteratif. Algoritma *K-Means* menggunakan jarak *Euclidean* sebagai metrik kedekatan antar data dan *centroid*, yang dirumuskan sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \quad (2)$$

Di mana, $d(x, y)$ menyatakan jarak antara data uji dan *centroid*, yang dihitung menggunakan rumus Euclidean. Variabel x_i merupakan data training, sedangkan y_i adalah data testing. Indeks i menunjukkan variabel data dan n merupakan dimensi data.

2.4 Particle Swarm Optimization

Particle Swarm Optimization adalah metode optimasi yang dapat dikombinasikan dengan metode pengelompokan seperti *K-Means Clustering* untuk mengatasi keterbatasan dan kombinasi yang dihasilkan dapat berkualitas tinggi [18]. Inisialisasi dari sistem PSO ini dengan populasi secara acak dan mencari solusi optimal dengan cara memperbarui generasi [19]. *Particle Swarm Optimization* (PSO) merupakan algoritma untuk optimasi berbasis populasi yang bersifat stokastik, terinspirasi dari perilaku sosial suatu kelompok, di mana setiap partikel merepresentasikan satu kemungkinan solusi terhadap permasalahan optimasi [20]. Adapun rumus pembaruan kecepatan dan posisi partikel dalam PSO adalah sebagai berikut:

$$v_i(t + 1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot (p_i - x_i(t)) + c_2 \cdot r_2 \cdot (g - x_i(t)) \quad (3)$$

$$x_i(t + 1) = x_i(t) + v_i(t + 1) \quad (4)$$

Di mana, v_i adalah kecepatan dan x_i adalah posisi partikel ke- i , sedangkan p_i dan g masing-masing menunjukkan *personal best* partikel dan *global best* seluruh partikel. Nilai r_1 dan r_2 adalah bilangan acak antara 0 dan 1 yang memberikan elemen acak dalam pencarian. Parameter w adalah bobot *inersia* yang mengatur pengaruh kecepatan sebelumnya, sementara c_1 dan c_2 merupakan koefisien kognitif dan sosial yang masing-masing mengukur seberapa besar partikel dipengaruhi oleh posisi terbaiknya sendiri dan oleh posisi terbaik partikel lain dalam populasi. Setiap partikel memperbarui posisi dengan mempertimbangkan nilai terbaik yang pernah dicapai oleh partikel tersebut sendiri (*Personal best* atau *Pbest*) dan nilai terbaik yang dicapai oleh seluruh partikel dalam populasi (*Global best* atau *Gbest*).

2.5 Davies Bouldin Index

Evaluasi dengan menggunakan *Davies Bouldin Index* (DBI) merupakan bentuk evaluasi dalam proses *clustering*, di mana kualitas hasil pengelompokan dinilai berdasarkan nilai DBI yang paling rendah. DBI sendiri adalah salah satu teknik yang digunakan untuk menilai seberapa akurat suatu pembentukan kluster dalam algoritma *clustering*. Proses evaluasi dilakukan dengan menghitung *Sum of Squares Within* (SSW) untuk mengetahui kerapatan data terhadap kluster dan *Sum of Squares Between* (SSB) untuk mengetahui separasi antar kluster dari *centroid* awal, yang kemudian digunakan untuk memperoleh nilai DBI. Berikut persamaan yang digunakan untuk memperoleh nilai SSW dan SSB:

$$SSE = \sum_{i=1}^K \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (5)$$

Di mana, k menyatakan jumlah kluster, C_i adalah himpunan data pada kluster ke- i , x_j adalah titik data dalam kluster tersebut, dan μ_i adalah centroid kluster ke- i .

$$SSB = \sum_{i=1}^K n_i \cdot \|\mu_i - \mu\|^2 \quad (6)$$

Di mana, n_i menyatakan jumlah data pada kluster ke- i , dan μ adalah mean global dari seluruh data.

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Dataset yang digunakan dalam penelitian ini diperoleh dari data transaksi penjualan di sebuah perusahaan produksi pakaian selama periode tahun 2023 hingga 2024 yang berisi 333 baris data. Data tersebut mencakup informasi mengenai kategori produk, jenis produk, jumlah produk yang terjual, dan total harga penjualan. Tabel 1 menampilkan contoh data transaksi penjualan produk pakaian yang digunakan dalam penelitian.

Table 1. Data Transaksi

No	No Transaksi	Produk	Jenis Produk	Kuantitas	Total Harga
1	BP4001	Kemeja	Kemeja Tactical	26	Rp5.780.000
2	BP4002-1	Kaos	Kaos	24	Rp2.400.000
3	BP4003	Jaket	Jaket Bomber	161	Rp37.990.000
...
331	BP5327	Kaos	Kaos Polo	30	Rp2.750.000
332	BP5329	Kemeja	Kemeja Print	2	Rp400.000
333	BP5330	Kemeja	Kemeja PDH	24	Rp4.145.000

Data yang ada pada Tabel 1 dilakukan proses *Label Encoding* agar setiap data pada kolom Produk dan Jenis Produk dalam bentuk nilai unik yang mempresentasikan kategori tersebut. Proses ini bertujuan untuk memudahkan dalam mengolah data. Hasil dari *Label Encoding* disajikan pada Tabel 2 untuk kolom Produk dan pada Tabel 3 untuk kolom Jenis Produk.

Table 2. *Label Encoding* Kolom Produk

Produk	Label
Celana	0
Jaket	1
Jas	2
...	...
Polo	6
Rompi	7
Wearpack	8

Table 3. *Label Encoding* Kolom Jenis Produk

Jenis Produk	Label
Almamater	0
Celana Cargo	1
Celana Joger	2
...	...
Sweater	35
Sweater Zip	36
Wearpack	37

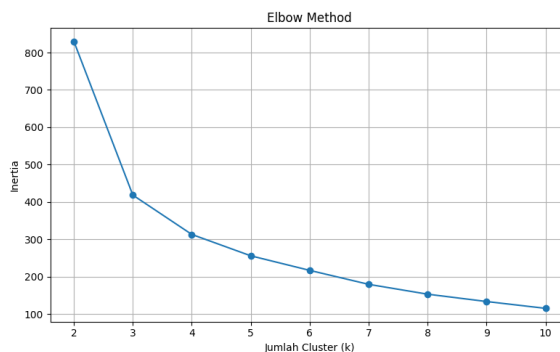
Agar semua data dalam tabel memiliki skala yang sama, maka dilakukan proses normalisasi data. Data yang disajikan pada Tabel 4 menunjukkan hasil normalisasi menggunakan *StandardScaler*. Metode ini mengubah setiap nilai menjadi skala standar dengan rata-rata 0 dan standar deviasi 1 yang berguna untuk menyamakan kontribusi antar fitur dalam pemodelan. Nilai positif dan negatif merepresentasikan deviasi terhadap rata-rata fitur. Proses ini penting untuk meningkatkan kinerja model pembelajaran mesin yang sensitif terhadap skala data, serta mengurangi bias antar fitur yang memiliki satuan berbeda.

Tabel 4. Hasil Normalisasi Data

No	Produk	Jenis Produk	Kuantitas	Total Harga
1	1.123485	0.483295	-0.366333	-0.291465
2	0.156015	0.031224	-0.383545	-0.446236
3	-1.391936	-0.872918	0.795466	1.183438
...
331	0.446256	0.031224	-0.331909	-0.430209
332	1.026738	0.483295	-0.572875	-0.537816
333	0.833244	0.483295	-0.383545	-0.366332

3.2 Proses Clustering

Clustering merupakan salah satu teknik dalam data mining yang digunakan untuk mengelompokkan data yang serupa berdasarkan karakteristik atau pola tertentu. *Clustering* memungkinkan perusahaan untuk dapat mengidentifikasi segmentasi dengan pola tertentu yang mendukung strategi pemasaran agar lebih efektif. Proses *clustering* pada penelitian ini dilakukan di *Visual Studio Code* menggunakan bahasa pemrograman *Python*. Proses *Clustering* diawali dengan mencari kluster (nilai *k*) menggunakan Metode *Elbow* dengan `random_state=42`.



Gambar 2. Grafik Hasil Metode *Elbow*

Grafik yang ada pada Gambar 2 menunjukkan hasil metode *Elbow* untuk menentukan jumlah kluster (nilai k) pada proses *clustering*. Berdasarkan grafik percobaan pada Metode *Elbow* di Gambar 2, terlihat bahwa penurunan nilai inerti mulai melambat signifikan dan hampir membentuk siku pada kluster 3, sehingga pada titik tersebut menjadi letak kluster (nilai k) awal yaitu 3 kluster.

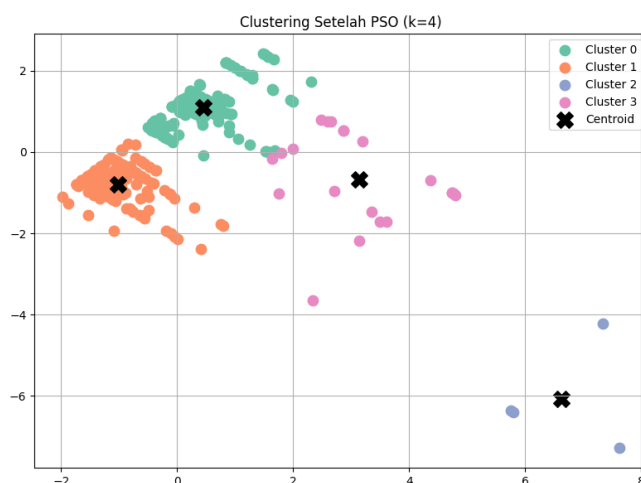


Gambar 3. Hasil *Clustering*

Visualisasi data yang ditampilkan pada Gambar 3 menampilkan visualisasi hasil *clustering* awal menggunakan algoritma *K-Means* sebelum penerapan *Particle Swarm Optimization* (PSO) dan menghasilkan 3 kluster ($k=3$). terlihat bahwa data terbagi menjadi tiga kluster yang memiliki tingkat kepadatan dan penyebaran yang berbeda. *Cluster 0* dan *Cluster 1* tampak cukup kompak dan terpusat di sisi kiri grafik, meskipun terdapat tumpang tindih di area batas kedua kluster tersebut. Sementara itu, *Cluster 2* menunjukkan persebaran yang cukup luas dan tidak merata, dengan beberapa titik data yang jauh dari pusat klasternya. Posisi *centroid* dari *Cluster 2* juga terlihat tidak representatif terhadap distribusi anggotanya yang mengindikasikan bahwa kluster ini memiliki variasi yang tinggi dan pemisahan yang kurang optimal, oleh karena itu diperlukan optimasi lebih lanjut untuk meningkatkan kualitas *clustering*.

3.3 Proses Optimasi Clustering

Proses optimasi *clustering* yang dilakukan dengan *Particle Swarm Optimization* (PSO) bertujuan untuk menentukan jumlah kluster (nilai k) yang lebih optimal. PSO tidak secara langsung memengaruhi nilai *inertia*, karena tujuan optimasi PSO difokuskan untuk mencari jumlah kluster terbaik berdasarkan nilai *Silhouette Score*, bukan berdasarkan meminimasi *inertia*.



Gambar 4. Hasil *Clustering* Setelah Optimasi

Visualisasi data yang ditampilkan pada Gambar 4 menggambarkan hasil *clustering* setelah penerapan optimasi dengan *Particle Swarm Optimization* (PSO). Jumlah kluster dari sebelumnya terdapat 3 kluster ($k=3$) menjadi 4 kluster ($k=4$) menunjukkan peningkatan struktur kluster yang lebih terdefinisi dibandingkan sebelumnya. Peningkatan yang paling terlihat pada pemisahan *Cluster 2* dan *Cluster 3* yang sebelumnya tergabung dalam satu kelompok dengan distribusi. Setelah penerapan PSO, kedua kluster tersebut berhasil dipisahkan menjadi kelompok yang lebih seragam dalam karakteristiknya dengan pusat kluster yang lebih tepat.

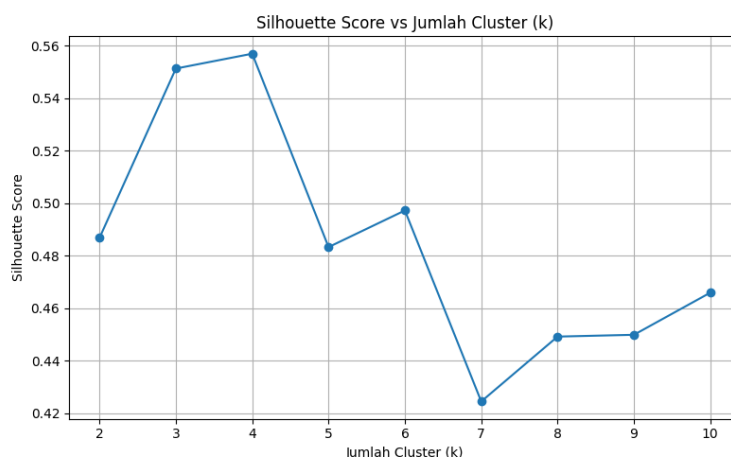
3.4 Evaluasi Cluster

Evaluasi performa *clustering* dilakukan dengan menggunakan metrik internal seperti *Silhouette Score* yang digunakan untuk menilai sejauh mana objek data cocok dengan kluster tempatnya berada dibandingkan dengan kluster terdekat lainnya, sehingga memberikan gambaran menyeluruh mengenai kekompakan dan keterpisahan antar kluster. Pada Tabel 5 disajikan data *Silhouette Score* untuk setiap jumlah kluster yang menjadi dasar dalam evaluasi nilai k .

Table 5. *Silhouette Score*

Jumlah Kluster (k)	<i>Silhouette Score</i>
2	0.4868121235195363
3	0.5512819399575227
4	0.5569341698401854
5	0.4832839793130053
6	0.49722628277572406
7	0.42447775459213494
8	0.4491906131049355
9	0.44987778031025033
10	0.4659164097595762

Berdasarkan data *Silhouette Score* pada Tabel 5 dapat disimpulkan bahwa kluster terbaik setelah optimasi PSO dicapai pada saat $k=4$ dengan *Silhouette Score* tertinggi sebesar 0.5569. Hal ini menunjukkan bahwa pemilihan jumlah kluster sebanyak 4 kluster memberikan hasil segmentasi yang paling optimal, dengan keseimbangan yang baik antara kekompakan data dalam kluster dan pemisahan antar kluster.



Gambar 5. Grafik *Silhouette Score* terhadap Jumlah Cluster (k)

Berdasarkan grafik yang disajikan pada Gambar 5 menunjukkan bahwa *Silhouette Score* terhadap jumlah kluster (k) yang dapat disimpulkan bahwa *Silhouette Score* tertinggi pada saat nilai $k=4$, yaitu mendekati angka 0.56. Hal ini menunjukkan bahwa pembentukan kluster paling optimal terjadi saat data dibagi menjadi 4 kluster, karena pada titik ini tercapai keseimbangan terbaik antara kekompakan data dalam kluster dan pemisahan antar kluster. Untuk mengetahui sejauh mana efektivitas algoritma PSO dalam meningkatkan kualitas *clustering*, dilakukan perbandingan hasil evaluasi kluster sebelum dan sesudah penerapan PSO yaitu $k=3$ dengan $k=4$ untuk menentukan mana yang menghasilkan segmentasi paling optimal.

Table 6. Evaluasi Perbandingan Jumlah Kluster

Jumlah Kluster (k)	DBI	<i>Silhouette Score</i>	SSE	SSW/SSB
3	0.6690	0.5513	418.52	0.4582
4	0.6878	0.5569	313.25	0.3075

Hasil evaluasi *clustering* yang disajikan pada Tabel 6 menunjukkan bahwa kluster dengan nilai $k=4$ yang diperoleh setelah optimasi PSO memberikan hasil yang lebih optimal dibandingkan nilai $k=3$ sebelum PSO. Meskipun nilai *Davies-Bouldin Index* (DBI) mengalami sedikit peningkatan dari 0.6690 menjadi 0.6878 yang menunjukkan peningkatan kemiripan antar kluster, namun peningkatan *Silhouette Score* dari 0.5513 menjadi 0.5569 mengindikasikan bahwa konsistensi kluster semakin membaik. Selain itu, penurunan nilai *Sum of Square Error* (SSE) dari 418.52 menjadi 313.25 menandakan bahwa penyebaran data dalam masing-masing kluster menjadi lebih rapat, sedangkan rasio SSW/SSB yang turun signifikan dari 0.4582 menjadi 0.3075 menunjukkan bahwa kluster yang terbentuk memiliki batas yang lebih tegas dan pemisahan yang lebih baik antar kluster.

3.5 Segmentasi Produk

Berdasarkan hasil segmentasi produk dengan *K-Means Clustering* yang dioptimasi menggunakan *Particle Swarm Optimization* (PSO), diperoleh 4 kluster utama. Berikut adalah rincian komposisi produk dalam masing-masing kluster beserta strategi pemasaran yang bisa diterapkan:

Kluster 1

- Jumlah Produk: 150
- Total Kuantitas: 6.277 unit (27.49%)
- Total Omset: Rp895.346.443
- Produk dengan Kontribusi Terbesar: Kaos Polo (21.27%), Kemeja PDL (16.31%), Kemeja PDH (16.11%), Rompi (12.38%)
- Target Pemasaran: Lembaga Pemerintah dan Perusahaan Korporasi
- Strategi Pemasaran: Penawaran dengan Katalog Produk

Kluster 2

- Jumlah Produk: 158
- Total Kuantitas: 6.472 unit (28.34%)
- Total Omset: Rp1.478.170.500
- Produk dengan Kontribusi Terbesar: Jaket Bomber (32.59%), Jaket Varsity (30.70%), Jaket Coach (7.52%)
- Target Pemasaran: Anak Muda, Mahasiswa, dan Komunitas
- Strategi Pemasaran: Marketing Digital dan Pemanfaatan Sosial Media

Kluster 3

- Jumlah Produk: 4
- Total Kuantitas: 2.900 unit (12.70%)
- Total Omset: Rp663.038.200
- Produk dengan Kontribusi Terbesar: Jaket Varsity (34.48%), Kaos Polo (34.48%), Jasket (31.03%)
- Target Pemasaran: Pasar Premium dan Brand Besar
- Strategi Pemasaran: Iklan dan *Endorsement*

Kluster 4

- Jumlah Produk: 21
- Total Kuantitas: 7.184 unit (31.46%)
- Total Omset: Rp1.007.801.500
- Produk dengan Kontribusi Terbesar: Rompi (61.11%), Kemeja PDL (15.95%), Kemeja PDH (8.39%)
- Target Pemasaran: Sektor Industri dan Proyek
- Strategi Pemasaran: Pendekatan Tender dan Pameran Industri

4. KESIMPULAN

Hasil *clustering* menunjukkan bahwa penggunaan algoritma *K-Means* yang dioptimasi dengan *Particle Swarm Optimization* (PSO) menghasilkan performa yang lebih optimal pada nilai $k=4$ dibandingkan $k=3$. Meskipun nilai *Davies-Bouldin Index* sedikit meningkat dari 0.6690 menjadi 0.6878, peningkatan *Silhouette Score* dari 0.5513 menjadi 0.5569 menunjukkan kluster yang lebih konsisten, dengan penurunan SSE dari 418.52 ke 313.25 serta rasio SSW/SSB yang turun tajam dari 0.4582 ke 0.3075, menandakan kluster yang lebih rapat dan terpisah jelas. Segmentasi ini menghasilkan 4 kluster produk dengan karakteristik dan potensi pemasaran yang berbeda. Kluster 1 berisi 150 produk seragam seperti Rompi, Kemeja PDH, dan Kemeja PDL, cocok untuk Lembaga Pemerintah dan Perusahaan Korporasi. Kluster 2 berisi 158 produk jaket seperti Jaket Bomber, Jaket Varsity dan Jaket Coach yang cocok dipasarkan pada segmen anak muda. Kluster 3 hanya berisi 4 produk dengan performa penjualan tinggi, ideal untuk pasar premium dan brand besar. Kluster 4 mencakup 21 produk seragam lapangan untuk sektor industri. Dengan segmentasi yang lebih terfokus ini, perusahaan dapat mengoptimalkan strategi pemasaran dan meningkatkan daya saing produk di pasar.

REFERENCES

- [1] K. Rahayu and P. T. Prasetyaningrum, "Sales Prediction on the Diamond Cell Counter Using Autoregressive Integrated Moving Average (ARIMA) Method," *Journal of Information Systems and Informatics*, vol. 5, no. 1, pp. 271–284, Mar. 2023, doi: 10.51519/journalisi.v5i1.450.
- [2] M. Almaripat, A. Faqih, and A. R. Rinaldy, "Sales Data Classterization Analysis Using K-Means Method for Marketing Strategy Development," *Journal of Artificial Intelligence and Engineering Applications*, vol. 4, no. 2, pp. 2808–4519, Feb. 2025, doi: <https://doi.org/10.59934/jaiea.v4i2.792>.
- [3] N. A. Maori, "Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means Clustering," *Jurnal SIMETRIS*, vol. 14, Nov. 2023, doi: <https://doi.org/10.24176/simet.v14i2.9630>.
- [4] W. Wahyuningsih and P. T. Prasetyaningrum, "Enhancing Sales Determination for Coffee Shop Packages through Associated Data Mining: Leveraging the FP-Growth Algorithm," *Journal of Information Systems and Informatics*, vol. 5, no. 2, pp. 758–770, May 2023, doi: 10.51519/journalisi.v5i2.500.



- [5] P. T. Prasetyaningrum, P. Purwanto, and A. F. Rochim, “Consumer Behavior Analysis in Gamified Mobile Banking: Clustering and Classifier Evaluation,” *Journal of System and Management Sciences*, vol. 15, no. 2, pp. 290–308, 2025, doi: 10.33168/JSMS.2025.0218.
- [6] I. Arfiani, H. Yuliansyah, and M. D. Suratin, “Implementasi Bee Colony Optimization Pada Pemilihan Centroid (Klaster Pusat) Dalam Algoritma K-Means,” *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 4, pp. 756–763, Mar. 2022, doi: 10.47065/bits.v3i4.1446.
- [7] P. T. Prasetyaningrum, P. Purwanto, and A. F. Rochim, “Enhancing Element Game Classification: Effective Techniques for Handling Imbalanced Classes,” *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 1, pp. 555–571, 2024, doi: 10.22266/ijies2024.0229.47.
- [8] A. S. Algani, A. R. Juwita, R. Rahmat, and S. Faisal, “Perbandingan Algoritma K-Means dan K-Medoids untuk Clustering Pada Transaksi Penjualan Minimarket,” *Journal of Computer System and Informatics (JoSYC)*, vol. 6, no. 1, pp. 14–24, Nov. 2024, doi: 10.47065/josyc.v6i1.5873.
- [9] F. A. Dinata, A. Nazir, M. Fikry, and I. Afrianty, “Implementasi Data Mining Untuk Prediksi Stok Penjualan Keramik dengan Metode K-Means,” *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 3, pp. 701–708, May 2024, doi: 10.47065/josyc.v5i3.5200.
- [10] N. S. Ngaeni, K. Kusri, and K. Kusnawi, “Analisis Kombinasi Algoritma K-Means Clustering dan TOPSIS Untuk Menentukan Pendekatan Strategi Marketing Berdasarkan Background Target Audiens,” *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 2, pp. 393–403, Feb. 2024, doi: 10.47065/josyc.v5i2.4948.
- [11] R. B. Ardi, F. Ely Nastiti, and S. Sumarlinda, “Algoritma K-Means Clustering Untuk Segmentasi Pelanggan (Studi Kasus : Fashion Viral Solo),” *INFOTECH journal*, vol. 9, no. 1, pp. 124–131, May 2023, doi: 10.31949/infotech.v9i1.5214.
- [12] A. R. Naufal and A. T. Suseno, “Penerapan Fitur Seleksi dan Particle Swarm Optimization pada Algoritma Support Vector Machine untuk Analisis Credit Scoring,” *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 1, pp. 184–195, Nov. 2023, doi: 10.47065/josyc.v5i1.4409.
- [13] T. M. Dista and F. F. Abdulloh, “Clustering Pengunjung Mall Menggunakan Metode K-Means dan Particle Swarm Optimization,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, p. 1339, Jul. 2022, doi: 10.30865/mib.v6i3.4172.
- [14] O. Herdiana, S. Maulani, E. A. Firdaus, and K. Kunci, “Strategi Pemasaran Produk Industri Kreatif Menggunakan Algoritma K-Means Clustering Berbasis Particle Swarm Optimization,” *Nuansa Informatika*, vol. 15, no. 2, 2021, doi: 10.25134/nuansa.v15i2.4394
- [15] L. Fernando and M. I. Fianty, “Optimizing Motorcycle Sales: Enhancing Customer Segmentation with K-Means Clustering and Data Mining Techniques,” *Journal of Information Systems and Informatics*, vol. 6, no. 3, pp. 1484–1498, Sep. 2024, doi: 10.51519/journalisi.v6i3.799.
- [16] P. Vania and B. Nurina Sari, “Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Klaster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means,” *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 21, pp. 547–558, 2023, doi: 10.5281/zenodo.10081332.
- [17] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa,” *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, vol. 3, no. 1, pp. 46–56, Mar. 2024, doi: 10.58602/jima-ilkom.v3i1.26.
- [18] Budiman, “Optimalisasi K-Means Berbasis Particle Swarm Optimization untuk Hasil Produksi Tanaman Sayuran di Indonesia,” *Nuansa Informatika*, vol. 17, pp. 2614–5405, 2023, [10.25134/fkom%20uniku.v17i1.6646](https://doi.org/10.25134/fkom%20uniku.v17i1.6646)
- [19] M. F. Wahyudi, S. Setiawidayat, and F. Hunaini, “Metode Particle Swarm Optimization Untuk Menentukan Daya Optimal Turbin Gas Pltgu Grati Berdasarkan Heat Rate,” *JASEE Journal of Application and Science on Electrical Engineering*, vol. 2, no. 01, pp. 37–46, Apr. 2021, doi: 10.31328/jasee.v2i01.161.
- [20] T. M. Dista and F. F. Abdulloh, “Clustering Pengunjung Mall Menggunakan Metode K-Means dan Particle Swarm Optimization,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, p. 1339, Jul. 2022, doi: 10.30865/mib.v6i3.4172.