

Segmentasi Produk Berdasarkan Penjualan Menggunakan Algoritma K-Means Clustering (Studi Kasus: Shopee)

¹Dennis Ma'rifatul Kurnia, ²Yulistiya Nur Hidayah, ³Erna Daniati

¹⁻³ Sistem Inofrmasi, Universitas Nusantara PGRI Kediri

E-mail: ¹denniskurnia33@gmail.com, ²yulistianurhidayah@gmail.com,
³ernadaniati@unpkediri.ac.id

Penulis Korespondensi : Erna Daniati

ABSTRAK ---Dalam lingkungan *e-commerce* seperti *Shopee*, terdapat ribuan produk dengan variasi harga dan performa penjualan yang berbeda. Untuk membantu penjual dan manajemen *marketplace* dalam pengambilan keputusan strategis, diperlukan pendekatan analisis data yang mampu mengelompokkan produk berdasarkan karakteristik relevan. Metode yang dinilai paling efektif dalam penelitian ini adalah analisis K-Means Kluster yaitu salah satu teknik *clustering* non-hierarkis yang bekerja dengan cara membagi sekumpulan objek atau data ke dalam beberapa kelompok (*cluster*) berdasarkan tingkat kemiripan karakteristik yang dimilikinya. Produk yang memiliki karakteristik yang sama akan tergolong ke dalam *cluster* yang sama, sementara produk dengan karakteristik yang berbeda akan dikelompokkan ke dalam *cluster* yang berbeda. Penelitian ini bertujuan untuk melakukan pengelompokan produk berdasarkan daya jual serta melihat keterkaitan antar variabel produk. Hasil segmentasi ini diharapkan memberikan informasi yang berguna untuk rekomendasi strategi penjualan, meningkatkan efisiensi promosi, dan memberikan *insight* bisnis yang lebih tajam bagi penjual maupun pihak pengelola *marketplace* dalam mengembangkan strategi pemasaran berbasis data.

Kata Kunci – K-Means Kluster, Produk, Shopee

ABSTRACT -- In an *e-commerce* environment like *Shopee*, there are thousands of products with different price variations and sales performance. To help sellers and marketplace management in making strategic decisions, a data analysis approach is needed that is able to group products based on relevant characteristics. The method that is considered the most effective in this study is the K-Means Cluster analysis, which is a non-hierarchical clustering technique that works by dividing a set of objects or data into several groups (*clusters*) based on the level of similarity of their characteristics. Products that have the same characteristics will be included in the same cluster, while products with different characteristics will be grouped into different clusters. This study aims to group products based on selling power and see the relationship between product variables. The results of this segmentation are anticipated to provide useful information for sales strategy recommendations, increase promotion efficiency, and provide sharper business insights for sellers and marketplace managers in developing data-based marketing strategies..

Keywords – K-Means Cluster, Products, Shopee

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Kemajuan teknologi informasi yang terus berkembang turut memengaruhi dinamika kehidupan manusia, tak terkecuali dalam sektor bisnis [1]. Dalam dunia bisnis, tingkat persaingan yang tinggi mendorong setiap perusahaan untuk mengoptimalkan seluruh potensi yang dimiliki guna mempertahankan eksistensi dan bersaing dengan kompetitor. Salah satu aspek penting yang perlu diperhatikan adalah kemampuan perusahaan dalam mengenali serta mengakomodasi karakteristik pelanggan [2][3]. *Shopee* adalah salah satu *marketplace* daring terkemuka yang beroperasi di kawasan Asia Tenggara, termasuk Indonesia. Jutaan penjual aktif dengan ribuan produk yang terus bertambah setiap harinya, Segmentasi produk membantu *Shopee* dalam mengelompokkan produk berdasarkan kategori, harga, kualitas, tingkat permintaan, hingga perilaku konsumen. *Online shop* mengalami kesulitan dalam menentukan jumlah stok minimum tiap produk karena kurangnya analisis terhadap minat dan permintaan konsumen [4]. Melimpahnya data penjualan dan produk pada *platform* seperti *Shopee* menimbulkan tantangan dalam memahami preferensi konsumen dan strategi pemasaran yang tepat. Oleh karena itu, dibutuhkan metode yang mampu mengelompokkan produk berdasarkan pola-pola tertentu dalam data, salah satunya adalah teknik *clustering* dengan pendekatan *K-Means*.

K-Means Clustering adalah salah satu metode analisis kluster non-hierarkis yang berusaha untuk membagi objek-objek ke dalam beberapa kluster berdasarkan kesamaan karakteristik. Dengan demikian, objek yang memiliki karakteristik serupa akan digabungkan dalam kluster yang sama, sementara objek dengan karakteristik berbeda akan ditempatkan dalam kluster yang terpisah [5][6]. Metode *K Means* memiliki pola formulasi dimana proses mencoba mencari jarak terpendek [7]. Bobot untuk setiap atribut membuat pembuat keputusan harus memikirkan beberapa pertimbangan. Ini dapat dibantu oleh proses pengelompokan yang menghasilkan bobot untuk setiap atribut [7]. Penelitian ini memiliki tujuan untuk melakukan segmentasi terhadap produk-produk yang dijual dan telah dibeli di *Shopee* berdasarkan karakteristik yang dimiliki. Melalui segmentasi ini, diharapkan dapat ditemukan pola-pola kesamaan antar produk sehingga memudahkan dalam pengelompokan dan kajian lebih rinci. Dengan menerapkan algoritma *K-Means Clustering*, penelitian ini bertujuan untuk mengelompokkan produk berdasarkan kesamaan tertentu yang dimilikinya. [8]. Hasil dari segmentasi ini dapat dimanfaatkan untuk mengidentifikasi produk-produk yang termasuk kategori *best seller*, menemukan peluang untuk melakukan strategi *cross up-selling*, serta menyusun strategi pemasaran yang tepat. Selain itu, segmentasi juga bertujuan untuk membantu perusahaan memahami kebutuhan pasar dengan lebih baik sehingga mampu meningkatkan profitabilitas yang lebih terarah pada setiap segmen produk yang terbentuk [9].

II. METODE

Penelitian ini menggunakan data sekunder yang merupakan ciri khas pendekatan kuantitatif. Dataset yang dianalisis diunduh dari platform *kaggle.com* dan telah dihimpun oleh pihak ketiga yang memiliki kredibilitas, meskipun identitas pihak tersebut tidak diungkapkan secara rinci untuk menjaga kerahasiaan. Pada penelitian ini, algoritma *K-Means* diterapkan untuk melakukan segmentasi produk berdasarkan pola penjualan yang teridentifikasi dalam data. Selanjutnya, proses evaluasi dilakukan untuk mengidentifikasi jumlah kluster yang paling tepat,

sehingga di peroleh segmentasi yang menggambarkan karakteristik produk secara lebih akurat. Penelitian ini menerapkan metode *K-Means Clustering* untuk mengelompokkan data kategorikal ke dalam beberapa klaster yang lebih terstruktur dan mudah untuk dianalisis. Pada metode ini, data yang memiliki kesamaan akan digabungkan dalam satu klaster, sedangkan data yang berbeda akan ditempatkan dalam klaster yang terpisah [1] [2]. Tujuannya adalah agar setiap klaster berisi data yang memiliki karakteristik yang serupa. Dataset yang kami gunakan adalah dataset shopee sample data yang memuat harga dan penjualan produk. Peneliti menggunakan metode KDD (*Knowledge Discovery in Databases*) digunakan untuk memahami masalah, memilih data yang relevan, dan melakukan *preprocessing* sebelum melakukan analisis data [12]. Pengetahuan yang diperoleh dapat digunakan sebagai dasar pengetahuan (*knowledge base*) yang berguna dalam proses pengambilan keputusan [13] [14].

Tahapan yang dilakukan dalam penelitian ini mencakup [15]

1. **Preprocessing data** terdiri dari Data dibersihkan dengan menghilangkan nilai yang hilang atau tidak valid, serta dilakukan normalisasi untuk memastikan semua fitur memiliki skala yang seragam dan reduksi dimensi jika diperlukan untuk mengurangi kompleksitas dan mempercepat proses klasterisasi.
2. **Klasterisasi** untuk membagi produk dalam beberapa segmen. Algoritma *K-Means Clustering* digunakan untuk mengelompokkan produk *Shopee* ke dalam beberapa segmen berdasarkan kesamaan karakteristik penjualannya, seperti jumlah penjualan dan total pendapatan. Metode ini bekerja dengan cara membagi data ke dalam sejumlah klaster yang ditentukan sebelumnya, di mana setiap data akan masuk ke klaster dengan *centroid* terdekat.
3. **Optimalisasi nilai k** untuk mengidentifikasi jumlah klaster yang paling optimal, diterapkan metode *Within-Cluster Sum of Squares* (WCSS). WCSS dihitung dengan menjumlahkan kuadrat selisih jarak antara setiap titik data dan pusat klaster yang bersangkutan. Persamaan perhitungannya dapat dijelaskan sebagai berikut:

$$CSS = \sum_{P_i \in \text{cluster}_1} \text{dist}(P_i, C_1)^2 + \sum_{P_i \in \text{cluster}_2} \text{dist}(P_i, C_2)^2 + \dots + \sum_{P_i \in \text{cluster}_n} \text{dist}(P_i, C_n)^2$$

[1]

Dimana:

x adalah titik data,

μ_i adalah pusat klaster ke- i ,

k adalah jumlah klaster, dan

C_i adalah himpunan anggota dari klaster ke- i .

Nilai k yang optimal ditentukan dengan menggunakan *Elbow Method*, yaitu metode yang memplotkan nilai WCSS terhadap berbagai jumlah klaster yang diuji. Titik optimal pada grafik ini ditandai dengan *elbow point*. Grafik ini menyajikan gambaran yang jelas tentang jumlah klaster yang paling tepat untuk data yang dianalisis.

III. HASIL DAN PEMBAHASAN

A. Proses *Preprocessing* Data

Sebelum memulai analisis dengan algoritma *K-Means Clustering*, langkah pertama yang krusial adalah mempersiapkan data melalui proses *preprocessing*. Proses ini memastikan bahwa dataset yang digunakan bersih dan siap untuk diolah lebih lanjut. Berikut

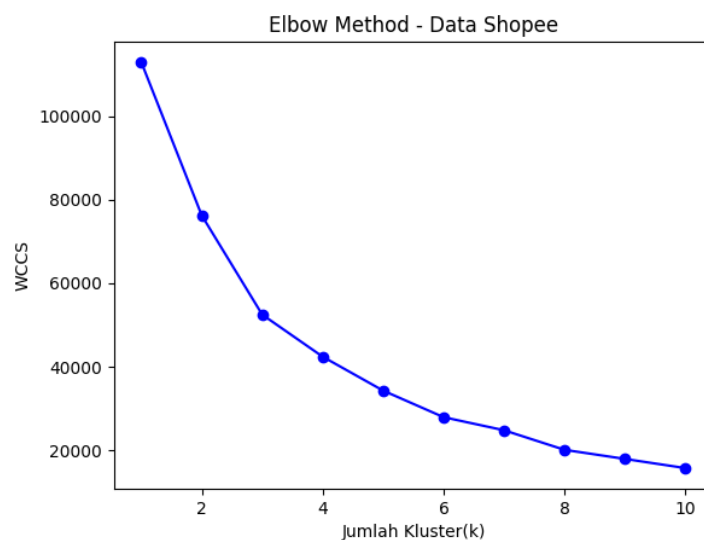
adalah gambaran singkat tentang kolom-kolom yang ada di dataset beserta karakteristiknya:

Tabel 1 : Hasil Data Preprocessing

NO	Tabel Dataset		
	Nama Column	Keterangan	Jumlah Unik
1	price_ori	Hasil asli produk (dalam RM)	2562
2	price_actual	Harga setelah diskon (dalam RM)	2946
3	total_sold	Jumlah produk terjual	1318
4	item_rating	Nilai rata-rata rating produk	21
5	total_rating	Total jumlah ulasan/rating	1318
6	favorite	Jumlah favorit/suka dari pengguna	1260

B. Analisis Klasterisasi

Setelah data penjualan *Shopee* diproses dan dibersihkan, langkah selanjutnya adalah menentukan jumlah kluster optimal yang akan digunakan dalam proses segmentasi produk. Untuk itu, digunakan metode *Elbow Method*.



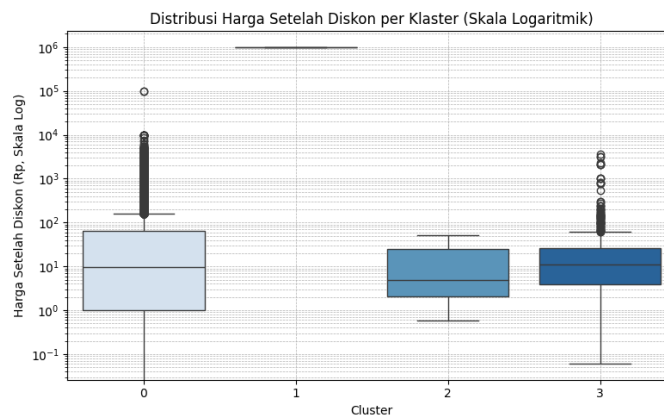
Gambar 1 : Elbow Method untuk Data Penjualan Shopee

Grafik *Elbow Method* menunjukkan penurunan tajam pada nilai *Within-Cluster Sum of Squares* (WCSS) dari $k=1$ hingga $k=4$. Pada titik $k=4$, terlihat adanya “elbow point”, yaitu saat penurunan WCSS mulai melambat secara signifikan. Ini menunjukkan bahwa penambahan kluster setelah titik tersebut tidak memberikan peningkatan berarti terhadap kualitas segmentasi. Oleh karena itu, dipilih $k=4$ sebagai jumlah kluster yang optimal untuk algoritma *K-Means*.

Setelah jumlah kluster ditetapkan, proses perhitungan *K-Means* dilakukan dengan menerapkan algoritma iteratif hingga posisi *centroid* stabil. Setiap produk dikelompokkan ke dalam salah satu dari empat kluster berdasarkan kedekatan terhadap *centroid* kluster yang

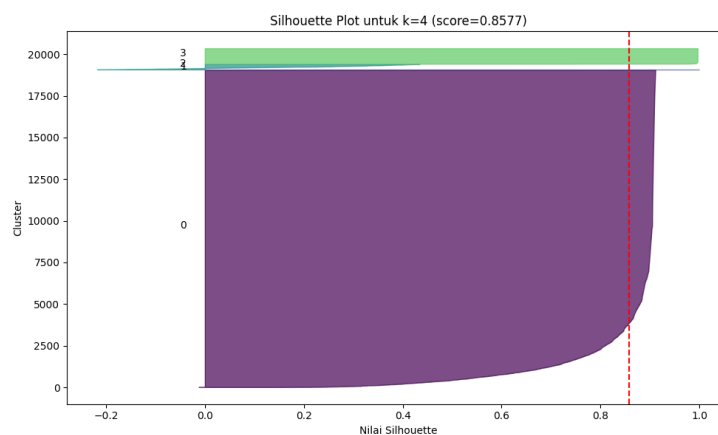
dihitung menggunakan jarak *Euclidean*.

Hasil klasterisasi divisualisasikan melalui distribusi harga setelah diskon di setiap klaster menggunakan *boxplot* dalam skala *logaritmik*.



Gambar 2 : Distribusi Harga Setelah Diskon per Klaster

Hasil dari klasterisasi divisualisasikan menggunakan *boxplot* yang menggambarkan distribusi harga setelah diskon pada masing-masing klaster, dalam skala *logaritmik*. Dari visualisasi tersebut, tampak perbedaan yang jelas antar klaster. Klaster 0 memiliki median harga sekitar 100 RM dan mencakup produk-produk premium, seperti gadget dan barang fesyen bermerek. Klaster 1 memiliki median harga sekitar 10 RM, dengan variasi harga yang cukup besar, mencerminkan produk-produk menengah seperti aksesoris atau pakaian kasual. Sementara itu, Klaster 2 dan 3 menunjukkan median harga di bawah 10 RM dengan rentang distribusi yang lebih sempit, mengindikasikan dominasi produk-produk ekonomis atau kebutuhan sehari-hari.



Gambar 3 : Silhouette Plot untuk k=4

Untuk menilai kualitas hasil klasterisasi, digunakan *Silhouette Score*. Nilai rata-rata yang diperoleh adalah **0,8577**, mendekati nilai maksimal 1, yang menandakan kualitas segmentasi sangat baik. Klaster 0 merupakan klaster terbesar dengan sekitar 10.000 produk, sebagian besar memiliki nilai *silhouette* di atas 0,8, yang menunjukkan pemisahan klaster yang kuat. Klaster 3 juga memiliki ukuran cukup besar, sekitar 5.000–7.000 produk, dengan kualitas klaster tetap tinggi. Sementara itu, Klaster 1 dan 2 berukuran lebih kecil namun tetap menunjukkan pemisahan yang cukup jelas.

C. Interpretasi Hasil

Dari hasil klasterisasi menggunakan *K-Means*, diperoleh 4 klaster produk yang berbeda berdasarkan harga dan performa penjualannya. Masing-masing klaster dapat diinterpretasikan sebagai berikut:

- a) Klaster 0: Produk dengan harga menengah dan penjualan stabil. Produk dalam klaster ini memiliki performa cukup baik dan cenderung diminati konsumen tanpa harus menawarkan harga sangat murah.
- b) Klaster 1: Produk dengan harga tinggi namun penjualan tinggi pula. Produk ini termasuk dalam kategori premium yang kemungkinan memiliki nilai *brand* atau kualitas tinggi.
- c) Klaster 2: Produk dengan harga rendah namun performa penjualan cukup bagus. Cocok untuk pasar massal dan strategi volume penjualan tinggi.
- d) Klaster 3: Produk dengan harga rendah dan penjualan rendah. Produk dalam klaster ini mungkin kurang diminati pasar atau tidak memiliki keunggulan kompetitif.

IV KESIMPULAN

Penelitian ini berhasil melakukan segmentasi produk *Shopee* menggunakan algoritma *K-Means Clustering* dengan $k=4$, menghasilkan nilai *silhouette* rata-rata 0.8577, menunjukkan kualitas klasterisasi yang sangat baik. Segmentasi berdasarkan harga setelah diskon (dalam RM) dan jumlah penjualan menghasilkan empat klaster: produk premium (Klaster 0), produk menengah (Klaster 1), produk murah stabil (Klaster 2), dan produk ekonomis (Klaster 3). Produk unggulan diidentifikasi pada Klaster 0 dan 2, sementara Klaster 3 menunjukkan produk kurang diminati. Hasil ini memungkinkan peningkatan akurasi rekomendasi produk dan strategi promosi yang lebih tepat sasaran, sehingga meningkatkan efisiensi pemasaran dan pengalaman pengguna di *platform Shopee*. Penelitian lebih lanjut dapat mempertimbangkan variabel tambahan seperti kategori produk atau tren musiman untuk analisis yang lebih mendalam.

UCAPAN TERIMAKASIH

Peneliti menyampaikan apresiasi kepada Program Studi Sistem Informasi Universitas Nusantara PGRI Kediri atas bantuan dan fasilitas yang telah diberikan. Penulis mengucapkan terima kasih kepada dosen pembimbing atas bimbingan yang telah diberikan selama penelitian ini, serta kepada para penulis referensi ilmiah yang telah menjadi sumber inspirasi. Semoga artikel ini bermanfaat dalam pengembangan ilmu di bidang sistem informasi dan analisis data.

DAFTAR PUSTAKA

- [1] L. Azzahra and Amru Yasir, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *J. Ilmu Komput. dan Sist. Inf.*, vol. 3, no. 1, pp. 1–10, 2024, doi: 10.70340/jirsi.v3i1.88.
- [2] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model Dan Teknik Clustering," *J. Terap. Teknol. Inf.*, vol. 2, no. 1, pp. 23–32, 2018, doi: 10.21460/jutei.2018.21.76.
- [3] E. F. L. Awalina and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail,"

- J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023, doi: 10.34010/jati.v13i2.10090.
- [4] E. Muningsih and S. Kiswati, “Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang,” *J. Bianglala Inform.*, vol. 3, no. 1, pp. 10–17, 2015.
- [5] L. H. Annisa and D. Rusvinasari, “Segmentasi Pembelian Produk Menggunakan Algoritma K-Means Berdasarkan Clusterisasi pada pemilihan menu yang ada diUMKM Kuliner,” vol. 9, no. 3, pp. 203–212, 2024, doi: 10.30591/jpit.v9i3.6556.
- [6] D. K. Utami, N. Irawati, and S. Sumantri, “Analysis of the k-Means Method in Clustering Acceptance of PKH Aid in Pulau Rakyat Tua Village,” *Sistemasi*, vol. 12, no. 3, p. 953, 2023, doi: 10.32520/stmsi.v12i3.3236.
- [7] E. Daniati and H. Utama, “Clustering K means for criteria weighting with improvement result of alternative decisions using SAW and TOPSIS,” *2019 4th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2019*, pp. 73–78, 2019, doi: 10.1109/ICITISEE48480.2019.9003858.
- [8] R. B. Ardi, F. Ely Nastiti, and S. Sumarlinda, “Algoritma K-Means Clustering Untuk Segmentasi Pelanggan (Studi Kasus : Fashion Viral Solo),” *INFOTECH J.*, vol. 9, no. 1, pp. 124–131, 2023, doi: 10.31949/infotech.v9i1.5214.
- [9] M. Harahap, Y. Lubis, and Z. Situmorang, “Analisis Pemasaran Bisnis dengan Data Science : Segmentasi Kepribadian Pelanggan berdasarkan Algoritma K-Means Clustering,” *Data Sci. Indones.*, vol. 1, no. 2, pp. 76–88, 2022, doi: 10.47709/dsi.v1i2.1348.
- [10] S. L. Achmad, A. Fauzi, R. Rahmat, and J. Indra, “SEGMENTASI PELANGGAN MENGGUNAKAN K-MEANS CLUSTERING DI TOKO RETAIL,” *J. Tek. Inf. dan Komput.*, vol. 7, no. 2, p. 736, Dec. 2024, doi: 10.37600/tekinkom.v7i2.1226.
- [11] . F., F. T. Kesuma, and S. P. Tamba, “Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering,” *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 67–72, 2020, doi: 10.34012/jusikom.v2i2.376.
- [12] E. Dwi *et al.*, “Penggunaan Data Mining untuk Prediksi tingkat Obesitas di Meksiko Menggunakan Metode Random Forest,” *Agustus*, vol. 8, pp. 2549–7952, 2024.
- [13] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [14] F. Gullo, “From patterns in data to knowledge discovery: What data mining can do,” *Phys. Procedia*, vol. 62, pp. 18–22, 2015, doi: 10.1016/j.phpro.2015.02.005.
- [15] B. H. Prakoso, E. Rachmawati, D. R. P. Mudiono, V. Vestine, and G. E. J. Suyoso, “Klasterisasi Puskesmas dengan K-Means Berdasarkan Data Kualitas Kesehatan Keluarga dan Gizi Masyarakat,” *J. Buana Inform.*, vol. 14, no. 01, pp. 60–68, 2023, doi: 10.24002/jbi.v14i01.7105.